Augmenting the power of blood tests for cancer p. 368

Invading ants lead to shift in lion prey pp. 370 & 433

Engineered enzymes for siloxane degradation p. 438

Science

$15
26 JANUARY 2024
science.org
AAAS

PUMP IT UP

Pushing water uphill to store green energy p. 358

EDITORIAL

Protein design meets biosecurity

The power and accuracy of computational protein design have been increasing rapidly with the incorporation of artificial intelligence (AI) approaches. This promises to transform biotechnology, enabling advances across sustainability and medicine. DNA synthesis plays a critical role in materializing designed proteins. However, as with all major revolutionary changes, this technology is vulnerable to misuse and the production of dangerous biological agents. To enable the full benefits of this revolution while mitigating risks that may emerge, all synthetic gene sequence and synthesis data should be collected and stored in repositories that are only queried in emergencies to ensure that protein design proceeds in a safe, secure, and trustworthy manner.

Nature's proteins elegantly address the challenges faced during the slow march of evolution, but today's problems, such as global pathogens, neurodegenerative diseases, and ecosystem degradation, require new solutions. AI-accelerated protein design can help tackle many of these issues. Machine learning–based methods enable the fast creation of biomolecules with diverse structures and functions that often have no detectable sequence homology to any known proteins. Concurrently, exponential improvements in DNA synthesis cost, quality, and speed have simplified encoding these proteins into synthetic genes. Last year, the first drug developed through computational protein design, the COVID-19 vaccine SKYCovione, was approved internationally. Many more such innovations are possible with this approach—and on short order. But as reflected in last year's global AI Safety Summit in the United Kingdom, the road to regulating AI is likely to be long and complicated. Progress in computational protein design could be hindered by overly restrictive AI regulations. The good news is that AI tools for protein design are highly specialized, and hence risk mitigation should be more straightforward.

Prior to the 2023 AI Safety Summit, a conference in Seattle, Washington, convened international representatives from academia, industry, philanthropy, and government agencies to discuss AI-enabled protein design, particularly for pandemic preparedness and drug development. The manufacture of synthetic DNA was recognized as a key biosecurity control point. Among the recommendations that emerged from the meeting was a policy of screening and logging all synthesized genetic sequences. This would present a practical barrier to the creation of harmful biomolecules, whether accidental or intentional.

Since 2004, the regulation of DNA synthesis, proposed and then voluntarily adopted by members of the International Gene Synthesis Consortium (IGSC), has been widely practiced in academia and the biotechnology and pharmaceutical industries. Currently, requests to academic, private, and government institutions for DNA sequences are screened by the IGSC for homology to pathogen components on a consensus list.

Going forward, these checks could be linked with the synthesis itself—whether chemical or enzymatic—such that each synthesis machine requires cryptographic short exact-match searches for each new genetic sequence. Screening sequences alone may not be sufficient because proteins generated through de novo design may have little or no sequence similarity to any natural proteins, complicating homology detection. Hence, there is a need to log synthesized sequences, using encryption as necessary to protect trade secrets. If a new biological threat emerges anywhere in the world, the associated DNA sequences could be traced to their origins. A "selective revelation" policy could ensure that such queries occur only under exceptional circumstances and on the basis of preestablished criteria. As biological complexity makes it highly unlikely that a dangerous agent could be created in one attempt, this capacity to trace nascent threats to their origins should be effective. Besides providing an audit trail, awareness that all synthesized sequences are being recorded may deter bad actors. Screening and logging practices should be standardized, practiced internationally, and extended to benchtop nucleic acid synthesizers.

> "Enhanced security need not threaten information sharing..."

This protein design security strategy depends on input from all relevant communities to support the required infrastructure and define the human, institutional, and governance requirements. Ideally, an international group such as IGSC should take the lead but work with governmental and nongovernmental organizations. Enhanced security need not threaten information sharing or transparent communication, the hallmarks of modern science; the use of biosecurity as an excuse to not share new methods and advances should be discouraged by science funders, publishers, and policy-makers. Rather, security in this fast-moving field should be framed as maximizing progress to address pressing societal concerns.

–David Baker and George Church

David Baker
is a professor in the Department of Biochemistry and director of the Institute for Protein Design at the University of Washington, Seattle, WA, USA.
dabaker@uw.edu

George Church
is a professor of genetics at Harvard Medical School, Boston, MA, USA.
gchurch@genetics.med.harvard.edu

10.1126/science.ado1671

NEWS

4.66 billion

Tons of coal produced by China in 2023, a 2.9% increase from 2022. (National Bureau of Statistics)

IN BRIEF

Edited by **Kelly Servick** *and* **Shraddha Chakradhar**

An infant receives the RTS,S malaria vaccine at a hospital in Cameroon on 22 January.

GLOBAL HEALTH

Long-awaited malaria shots rolled out

After a 60-year quest, the first-ever routine childhood malaria vaccinations—those given as part of the regular immunization schedule—were administered to infants and toddlers in Cameroon on 22 January. They received RTS,S or Mosquirix, made by GlaxoSmithKline and approved for general use in 2021 by the World Health Organization (WHO). The vaccine's efficacy wanes substantially over time, but a 4-year pilot rollout required by WHO in Ghana, Kenya, and Malawi showed it slashed illness and death in young children. Nineteen other African countries aim to begin administering RTS,S or another recently approved malaria vaccine, R21, routinely this year. Both are given as a series of four shots, normally beginning in the sixth month of life. Malaria kills about 470,000 children younger than age 5 in Africa annually. In Cameroon, malaria incidence grew by 49% between 2015 and 2022.

U.S. agency powers go on trial

REGULATORY POLICY | The U.S. Supreme Court last week appeared closer to overturning a 40-year legal doctrine, known as the Chevron deference, in which judges allow federal agencies to interpret ambiguous legislative mandates. To do so, agencies often rely on staff scientists. The two cases before the court involve a requirement that herring companies pay for onboard monitoring of their catch to prevent overfishing. The companies say the rule is an overreach of the Department of Commerce's authority under a 1976 law. In accepting the case, the high court agreed to reexamine a two-step process in which judges defer to "reasonable" actions by federal agencies. Associate Justice Elena Kagan said abandoning Chevron—based on a 1984 ruling involving the fossil fuel giant—would mean no longer allowing "people who actually know about" the topic to implement government policies, citing as examples the agency expertise needed to regulate artificial intelligence and ensure drug safety. Associate Justice Neil Gorsuch disagreed, saying deference to federal agencies "abdicates the court's responsibility" to interpret the law. Most legal experts expect Gorsuch to be on the winning side of the court's decision, due out this spring.

Japan makes first Moon landing

PLANETARY SCIENCE | Japan became the fifth country to successfully land a functioning craft on the Moon last week when its Smart Lander for Investigating Moon (SLIM) touched down on 20 January, early Japan time. Communications from the craft indicated all was normal except that the solar panels were not generating power. In a series of posts on the social media platform X (formerly Twitter) on 22 January, the Japan Aerospace Exploration Agency (JAXA) reported that after downloading data and images from SLIM, it had turned off the power to avoid draining the batteries. Telemetry data indicate the solar panels are not facing the Sun. The JAXA team hopes they will be able to generate power when the Moon's trajectory turns them toward sunlight. SLIM, which was developed to

OCEANOGRAPHY

Record-setting deep reef off U.S. coast fully mapped

The world's largest deep-sea coral reef sits just off the Atlantic coast of the United States, stretching from Florida to South Carolina, researchers reported last week in *Geomatics*. The reef, which was previously known but incompletely surveyed because of the high cost of seabed mapping, lies as deep as 1 kilometer below the ocean's surface and covers some 26,000 square kilometers—three times the size of Yellowstone National Park. Using data from 31 sonar mapping surveys, the team detected nearly 84,000 distinct coral mounds, including thickets of *Desmophyllum pertusum* (pictured here), which thrives in cold, sunless conditions. Little is known about these reefs beyond the important habitat they provide for sharks, starfish, shrimp, and other marine life. Given that only 25% of the ocean floor has been mapped, larger, unknown reefs likely exist.

PHOTO: NOAA OFFICE OF OCEAN EXPLORATION AND RESEARCH, WINDOWS TO THE DEEP 2019

demonstrate techniques for landing small, light spacecraft near lunar and planetary features of interest, touched down within the planned 100 meters of its target. Japan follows the United States, the Soviet Union, China, and India in reaching the Moon.

Face recognition outpaces laws

ARTIFICIAL INTELLIGENCE | Rapid progress in facial recognition technology has prompted the National Academies of Sciences, Engineering, and Medicine (NASEM) to call for more oversight from governments, courts, and the private sector. In a report released on 17 January, a NASEM committee warned of limitations and potential misuse of the technology, which already has various applications, from unlocking phones to screening entrants at concerts. The report highlights major concerns about violations of privacy and biases against people of color, women, and the elderly, and notes that law enforcement applications of facial recognition have already resulted in wrongful arrests. The committee recommends new federal regulations that would address potential civil rights violations from the use of this technology and maintain privacy when data are stored in government and private servers. The report also proposes requiring training for law enforcement and other personnel who handle sensitive data.

Cancer studies face retractions

RESEARCH INTEGRITY | Dozens of papers by the head of the Dana-Farber Cancer Institute and three senior DFCI researchers need to be retracted or corrected, the institute disclosed last week. The announcement came following a 2 January blog post by freelance data sleuth Sholto David alleging data and image manipulation in 57 papers on fundamental aspects of cancer biology authored between 1997 and 2017 by DFCI President and CEO Laurie Glimcher, Chief Operating Officer William Hahn, Senior Vice President Irene Ghobrial, and center Director Kenneth Anderson. DFCI has requested retractions for six papers and corrections in 31 others for which the authors "have primary responsibility for the potential data errors," says Barrett Rollins, the institute's research integrity officer. DFCI reportedly began its investigation a year ago, and Rollins says he is still investigating 16 other papers containing data from other DFCI labs that David flagged.

U.S. earthquake risk assessed

SEISMOLOGY | Nearly 75% of the U.S. population could experience a damaging earthquake and severe ground shaking in the next 100 years, according to a new assessment of the country's seismic hazard, published by the U.S. Geological Survey (USGS) this month. Incorporating nearly 500 hazardous faults newly characterized since the last survey in 2018, the report finds that the mid-Atlantic and Northeast are slightly more at risk from a quake than previously thought, as are seismically active California and Alaska. Recent volcanic unrest has also heightened risk in Hawaii, USGS concluded. The agency's hazard assessment, updated every 5 years, has wide-reaching influence, helping dictate building codes and insurance premiums.

Humans host odd RNA circles

BIOLOGY | Bacteria from the human mouth and gut contain previously unknown circles of RNA resembling the genomes of flu, Ebola, and other viruses, researchers reported this week. The particles, which the scientists are calling obelisks, contain as little as 3% as much genetic material as full-size viruses. They turn up in 7% of the human gut microbes and half the mouth microbes, the researchers reported on 21 January in a preprint posted to bioRxiv. Some obelisks, like some RNA viruses, seem to replicate not by hijacking host cells, but by carrying an enzyme to make copies of themselves. Researchers suspect these particles could affect the function of their hosts' genes, but their potential effects on human health remain unclear.

The 1.635-billion-year-old Chuanlinggou Formation in China (above) yielded microscopic, algalike fossils, including some with spores (below, top image).

PALEONTOLOGY

Tiny fossils upend timeline of multicellular life

Eukaryotes organized into multicellular forms 1.6 billion years ago

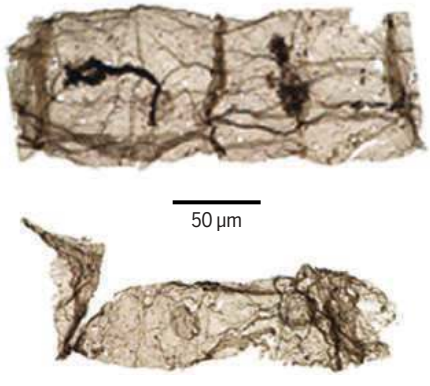
By **Elizabeth Pennisi**

A new study describing a microscopic, algalike fossil dating back more than 1.6 billion years supports the idea that one of the hallmarks of the complex life we see around us—multicellularity—is much older than previously thought. Together with other recent research, the fossil, reported this week in *Science Advances*, suggests the lineage known as eukaryotes—which features compartmentalized cells and includes everything from redwoods to jellies to people—became multicellular some 600 million years earlier than scientists once generally thought.

"It's a fantastic paper," says Michael Travisano, an evolutionary ecologist at the University of Minnesota who helped show that yeast can become multicellular in the lab. "This gives us a better idea of the grand vision of life."

Typically, biologists subdivide that grand vision into two categories: eukaryotes, with their DNA packaged into nuclei, and prokaryotes such as bacteria, which have free-floating DNA. Prokaryotes evolved first, up to 3.9 billion years ago; within a few hundred million years, some of them, the cyanobacteria, began to form chains of cells, considered an advance in life's complexity. About 2 billion years ago, much larger, single-cell eukaryotes bearing nuclei showed up. For decades, researchers thought eukaryotes didn't form simple multicellular structures until 1 billion years after they arose, and that once chain structures evolved, more elaborate body plans—animals with organs—appeared soon after. "There was this perception that multicellularity was hard [to evolve]," Travisano says.

Then in 1989, researchers described *Qingshania magnifica*, a microscopic fossil they suggested was a primitive green alga, a multicellular eukaryote. No one paid the discovery much mind, even though it came from the Chuanlinggou Formation in North China, which includes layers that are 1.6 billion years old. But since 2015, Maoyan Zhu and Lanyun Miao, paleobiologists at the Chinese Academy of Sciences's Nanjing Institute of Geology and Palaeontology, have collected rocks from the same area, dissolved them, and eventually uncovered 279 microscopic fossils, all but one of them specimens of *Q. magnifica*.



In this week's paper, they report that the fossils consist of strings of up to 20 cylindrical cells, with adjoining cell walls, like plants, visible under a microscope as dark rings. Several fossils had spores—with their own cell walls—suggesting the filaments had specialized reproductive structures.

"What's striking about these fossils is they are really rather enormous for that age, and they are multicellular," says Jochen Brocks, an organic geochemist at Australian National University. William Ratcliff, an evolutionary biologist at the Georgia Institute of Technology who also works on multicellular yeast, adds that he's impressed by the level of internal detail revealed in the ancient life. "I got a little dopamine hit seeing those internal sporelike compartments."

Miao performed chemical tests on the fossils and found the structures of their organic carbon compounds were different from those in cyanobacteria fossils in these rocks. Her team concluded the filaments were most likely green algae, similar to modern eukaryotes such as *Urospora wormskioldii*.

"The authors have done a commendable job of interpreting the fossils," says Stefan Bengtson, paleobiologist emeritus at the

Swedish Museum of Natural History. "The hypothesis that these are filamentous green algae is a good start."

The new findings build on work Bengtson and colleagues reported in 2017, when they proposed that 1.6-billion-year-old fossils found in India represented red algae. In 2021, another team described "walled microfossils," which they interpreted as a diverse set of eukaryotes, in deposits from Canada dating back 1.57 billion years. And just last month, Leigh Anne Riedman and Susannah Porter, paleontologists at the University of California, Santa Barbara, and colleagues described what they say are several eukaryotic fossils found in 1.642-billion-year-old rocks from Australia.

The sheer diversity of body plans found in these early forms of multicellular life is astounding, Riedman notes. Some were cylindrical with chambers. Others were spherical. One had a lid that appeared to open, possibly to get rid of the cell's contents. "Every indication suggests eukaryotes were much more diverse and complex by this time than previously appreciated," she says.

If simple but diverse multicellular forms appeared so early, then complex multicellularity took a lot longer to evolve than most researchers had thought; the first creatures with organs and cells that did not have direct access to the outside environment didn't appear until less than 1 billion years ago. Such a delayed timeline makes sense to Shuhai Xiao, a geobiologist and a paleobiologist at the Virginia Polytechnic Institute and State University. Truly complex eukaryotes "have multiple cells that stay together, communicate with each other, and have different sizes, shapes, and functions," he explains. "It takes time [to make such advances]."

If the recent findings hold up, they are "remarkable" and transformative, says László Nagy, an evolutionary biologist at the Hungarian Research Network's Biological Research Centre. But he's cautious about claiming similarities to living algae. "It is challenging to compare a 1.6-billion-year-old organism to extant ones," Nagy says. "This is such a long time that any resemblance to extant organisms may be due to chance." And Ratcliff says these organisms may not even be eukaryotes: "It's possible that [these fossils] are just superweird bacteria that don't resemble extant species."

But Harvard University paleontologist Andrew Knoll, a co-author on the *Science Advances* paper, says the data and the presence of cell walls—which prokaryotes lack—are proof enough. "If this were found in [400-million-year-old] Devonian rocks, people would describe it as algae and no one would bat an eyelash," he says. ■

With reporting by Dennis Normile.

SCIENCE POLICY

Strong medicine for Argentina's beleaguered science

Daniel Salamone, the country's new science head, faces skepticism as he touts new priorities and entrepreneurship

By **María de los Ángeles Orfila**

Protests are a common sight in Buenos Aires amid Argentina's prolonged economic crisis. But lately the presidential palace and the Ministry of Economy haven't been the only scenes of turmoil. In recent weeks angry crowds have also gathered at the National Council of Scientific and Technical Research of Argentina (CONICET), the country's main science agency. CONICET's own scientists and administrators are marching with drums, megaphones, and banners, hoping their voices reach the agency's new and controversial head, veterinarian Daniel Salamone.

Appointed by Argentina's newly elected libertarian president, Javier Milei, Salamone took office at the end of 2023, facing formidable challenges. He reports to a president who abolished the Ministry of Science and, during the electoral campaign, accused CONICET of being "unproductive." One of the top science agencies in South America, with 11,800 researchers, CONICET lacks an approved budget. Even if the government matched the $400 million the agency received in 2023, the country's 200% annual inflation has sharply eroded its value.

Some Argentine scientists also question Salamone's ability—or desire—to defend their interests. But he accuses his critics of "a visceral response" to Milei. "It could be a biased viewpoint on my part, but previous governments haven't shown particular adeptness in managing science," he said in a recent interview with *Science*.

Salamone's high-profile work at CONICET on cloning and his experience in the private sector helped him catch Milei's eye. The Argentine president, who owns five cloned mastiffs, calls Salamone the "national cloner" for achievements that include cloning cows engineered to secrete human growth hormone in their milk, which could provide a cheap source of the hormone for medicines. His team also produced the first cloned horse in South America and created pigs that could be used for skin grafts because they lack key molecules that trigger immune rejection.

"Six companies emerged from our laboratory," Salamone says. One was Kheiron Biotech, which has already seen the birth of more than 400 horse clones and is approaching a production rate of 150 clones per year. Gabriel Vichera, a former student of Salamone's who co-founded the company, recalls him as "a highly pragmatic person who always seeks to ensure that his research finds practical application."

Salamone says his record of founding companies appealed to Milei, a free-market partisan. "[It aligned] with the direction he

PHOTO: CONICET

Daniel Salamone's work on livestock won him a reputation as Argentina's "national cloner."

wanted to give to CONICET: advancing rigorous scientific endeavors while bolstering entrepreneurial initiatives."

CONICET needs to rethink how it spends its resources, Salamone says, for example by bolstering research that helps alleviate the poverty affecting 40% of Argentines. Engaging with the private sector could free up funds for such research, he says. A law could soon be passed to aid the transfer of scientific and technological knowledge to private entities. Proposed by Milei, it is currently under examination in the National Congress.

Critics worry the law would starve basic research of funding. One, Víctor Ramos, president of the National Academy of Exact, Physical and Natural Sciences of Argentina, says Salamone "holds very simplistic ideas and demonstrates a complete ignorance of the scientific system."

Ramos and others are also alarmed that CONICET has suspended scholarships and promotions until it has a 2024 budget. "Those who were accepted to enter the research career are left on the streets. The most promising talents selected are on the verge of seeking opportunities abroad," he warns. Salaries for Argentine scientists are already among the lowest in the region.

Since Salamone took office, the agency has dismissed 50 administrators. More than 260 leaders of CONICET's Scientific and Technological Centers across the country have protested the move, issuing a statement decrying "the dismantling of a portion of CONICET's organizational capabilities."

Salamone says he will not abandon research areas within CONICET that have an international reputation, such as paleontology, or other disciplines that "generate knowledge for humanity." That commitment will require a stable research budget, says Jorge Montanari, director of the National University of Hurlingham's Laboratory of Nanosystems for Biotechnological Application. Creating technology-based companies is worthwhile, he says. But that, he adds, will take "a critical mass of expanding research that becomes increasingly appealing for private investment. It's about never cutting back."

Salamone emphasizes that he wants to retain talented scientists in Argentina and even entice those abroad—like his daughter, a neuroscientist in Sweden—to return. "I always tried to show my daughter with my work that excellent science can be conducted here, but she remains unconvinced," he says. He's hoping Milei's presidency and his own initiatives will make a difference. "We have the potential to halt the decline of recent years." ■

María de los Ángeles Orfila is a journalist in Montevideo, Uruguay.

COVID-19

SARS-CoV-2 sneezing explained

Viral protein prods nerve cells, may be treatment target

By **Mitch Leslie**

SARS-CoV-2 has many ways of making people miserable, including by causing them to sneeze. Now, researchers have discovered the basis for this nose-tickling effect. One of the virus' proteins stimulates neurons in respiratory passages, triggering the sneeze reflex. The results could spawn novel treatments to ease COVID-19 symptoms and to reduce transmission of SARS-CoV-2.

They might also apply to other sneeze-inciting viruses. "Prior to this study, nothing was known about how viruses cause sneezing," says neuroimmunologist Isaac Chiu of Harvard Medical School, who wasn't connected to the research. The study is the first to show that a viral protein "can be directly sensed by neurons to cause sneezing."

Sneezes are protective, ejecting bothersome and potentially harmful substances from the body. They also help pathogens such as SARS-CoV-2 reach new hosts. A human sneeze can catapult 40,000 virus-laden droplets as far as 8 meters. But researchers assumed sneezes are an incidental byproduct of illness, as infected cells spill out molecules that irritate nasal passages.

Neurophysiologist Diana Bautista of the University of California, Berkeley and colleagues suspected SARS-CoV-2 might play a more direct role. Infected cells pump out large amounts of the viral protein PLpro, part of a family of enzymes called proteases that carve up other proteins. Previous research showed that other proteases made by plants, bacteria, and even humans stimulate sensory neurons, the cells that induce sneezing.

The researchers squirted PLpro into the noses of mice and found it stimulated a subgroup of sensory neurons called nociceptors that produce pain and itch sensations. The team then tested the protein's effect on sneezing. The rodents started sneezing about 14 seconds after PLpro exposure, versus 30 seconds after getting a control mixture. Mice dosed with PLpro sneezed almost four times more than controls in the first 2 minutes, the team reported in an 11 January preprint on bioRxiv.

"We were excited and horrified" by the results, Bautista says, because they show a powerful effect on sneezing that could promote virus transmission. By inserting blue dye into the animals' noses along with the test solutions and measuring spatter on the floors of their cages, the team showed sneezes expelled large quantities of nasal secretions.

The team couldn't test whether PLpro enhances coughing, another virus-spreading symptom, because researchers aren't sure whether mice actually cough, Bautista notes. But her group did implicate PLpro in face and mouth pain, also common in COVID-19. When they injected the protein into the rodents' cheeks, the animals wiped their faces with their front paws more often, a sign that they were hurting.

A mouse sneezes after exposure to PLpro.

The researchers tested two other coronaviruses and found that PLpro from one of them, the cause of severe acute respiratory syndrome, also stimulates sensory neurons. Other viruses, including some that cause colds, also carry the protein, Bautista and colleagues note, suggesting they, too, might actively trigger sneezes.

PLpro activates nociceptors by prompting protein channels to allow in calcium, but it doesn't act on the channels directly. The researchers think it targets a different receptor they have yet to identify.

"What they've found is very compelling," says neurobiologist Theodore Price of the University of Texas at Dallas. Because PLpro is necessary for SARS-CoV-2 to infect cells, researchers are already exploring it as a drug target. Dozens of compounds that might block the protein are in preclinical development. The new results suggest these candidates could also quell symptoms and hinder transmission. But neuro-immunologist Felipe Ribeiro of Washington University School of Medicine in St. Louis cautions that researchers must rule out the possibility that sneezing speeds recovery from COVID-19. "You have to show that blocking it is not harmful." ■

Over the past 40 years, the European wolf popula has grown to about 20,000 animals.

CONSERVATION

Europe divided on proposal to allow wolf culls

Farmers argue that downgrading wolf protection will spare livestock—but some scientists say culls could backfire

By **Gennaro Tomma**

Far from being confined to folk tales, wolves in Europe are startlingly plentiful today. Now, governments want to reduce the numbers to protect livestock, sparking debate—with scientists caught in the fray.

Late last month, the European Commission released a proposal to weaken protections for wolves living in the 27 nations of the European Union, drawing criticism from environmental groups. Just days later, environmentalists persuaded a court in Switzerland, which is not a member of the EU, to partially block a new government plan to kill up to 70% of the nation's wolf population.

After centuries of hunting, only small and scattered populations of wolves survived in Europe by the 1970s, but recent studies estimate some 20,000 animals now roam the continent. The rebound is largely due to protections provided to wolves and other large carnivores under the Berne Convention on the Conservation of European Wildlife and Natural Habitats, a 40-year-old conservation agreement.

Gennaro Tomma is a freelance science journalist from Italy.

As the number of wolves has increased, however, so has predation on domestic livestock. Every year wolves kill 65,000 farm animals, mainly sheep, according to the Commission. Although this amounts to just 0.07% of the continent's sheep, farm groups across Europe have lobbied officials to weaken rules against killing wolves.

On 20 December 2023, the Commission responded by releasing a proposal to downgrade the wolf's protection status from "strictly protected" to "protected." The change would allow EU nations to cull wolves at scale for the first time in 4 decades, although countries would still be obligated to ensure that wolves maintain a "favorable" conservation status. Each nation would decide its own culling quotas, time frames, and culling methods, which supporters of the plan say will make it easier to keep wolf populations at healthy but more manageable levels.

"The comeback of wolves is good news for biodiversity in Europe, but the concentration of wolf packs in some European regions has become a real danger, especially for livestock," said Ursula von der Leyen, president of the Commission, in a statement that accompanied the proposal.

Many environmental groups have criticized the plan. In an open letter, some 300 organizations including the World Wildlife Fund and Rewilding Europe accused the commission of soliciting anecdotal evidence on the impact on wolves during an "irregular" consultation process, rather than gathering reliable scientific data. "We are concerned that the discussion of this issue has so far been largely dominated and driven by farming industry and hunting interest representatives," they write, pointing to a survey that suggests most rural inhabitants believe wolves should continue to be strictly protected. "Unless there is substantial new science-based evidence gathered by the European Commission services, we believe the science and public opinion are clear: the modification of the protection status of the wolf … is not justified."

Some scientists agree, pointing to a lack of evidence that culling actually reduces predation on sheep. "Implementing selective culls would be expensive, and in most cases ineffective," says Gianluca Damiani, a wolf expert at Tuscia University. Killing wolves and breaking up packs could actually make the problem worse, he says, because domestic livestock make an easy meal for a wolf that is lost and alone. Damiani would prefer to see any funding earmarked for culling instead go toward providing protections for livestock, such as electric fences and dogs.

Further research is needed to understand how to effectively cull wolves, agrees Luigi Boitani, chair of the Large Carnivore Initiative for Europe and a leading wolf expert. For example, it's still unclear what percentage of wolves has to be removed in order to reduce livestock kills. Boitani says reducing protection for wolves makes sense "from a continental perspective," given that the European wolf population is in good health overall. He worries, however, about the EU's plan to allow individual counties to make decisions about culls. Many EU nations—such as Belgium and Slovenia—do not have self-sufficient numbers of wolves. To avoid decimating small, vulnerable wolf populations, wolf management should instead happen at the European level, he says. Before Switzerland's cull was partially suspended, environmentalists had expressed similar concerns about the survival of the country's relatively small wolf population.

The EU's proposal still needs to go through a protracted process before becoming law. Among other steps, any change to wolves' conservation status will need unanimous approval from all 27 EU member states. "The road is long," Boitani says. "Downlisting remains probable but not certain." ■

PHOTO: AXEL GOMILLE/NPL/MINDEN PICTURES

PUBLISHING

Retractions lag for wave of suspect papers

Years after whistleblowers questioned nearly 300 papers, journals are slow to respond

By **Jeffrey Brainard**

Over the past decade, a team of scientific sleuths uncovered one of the most extensive known bodies of faked research. They notified 78 journals about almost 300 papers by a pair of Japanese physicians that bore signs of fabrication and other ethical lapses. Nearly half have been retracted, putting the authors, Yoshihiro Sato and Jun Iwamoto, in fourth and sixth place, respectively, on Retraction Watch's list of authors with the most retractions. But when the investigators contacted editors to encourage reviews of the remaining papers, the response was mostly silence.

The critics' efforts to correct the record, which they detail in a paper published last month in *Accountability in Research*, offer a high-profile example of familiar problems in scientific publishing. Retractions come slowly—often years after complaints arise, if at all—in part because journals may defer to institutional investigations, which can be slow, unreliable, or absent. Journals' decisions also lack transparency. As such, efforts to track the fate of suspect papers are vital to "ensure that journal articles represent a robust and dependable body of evidence," says Ursula McHugh, an anesthesiologist at St James's Hospital in Dublin who has studied retractions.

In 2016, the investigator team—Andrew Grey, Mark Bolland, and Greg Gamble of the University of Auckland and Alison Avenell of the University of Aberdeen—published an analysis of 33 papers by Sato, Iwamoto, or both. It described implausible data and other suspicious aspects of the papers, about bone fractures and osteoporosis. Some, for example, claimed to have recruited thousands of research subjects with no obvious staff or funding. In time, the list of suspect papers grew. Before his death, Sato admitted fabricating results but absolved Iwamoto (*Science*, 17 August 2018, p. 636).

The investigators reported their findings to publishers and editors at journals where the papers had appeared and tracked their responses. In 2020, they sent a final wave of prompts and notifications, and in April 2023 they tallied the results.

Journals ultimately took editorial action for 136 papers, retracting 121, correcting three, and marking 12 with editorial expressions of concern. For 57 other papers, journals told the team they had conducted reviews and determined that 22 should be retracted. No further action was taken until last month, when industry giant Elsevier retracted seven of them, 3.5 years after the Committee on Publication Ethics (COPE) threatened sanctions, Retraction Watch reports. The team received no responses about an additional 107 papers in 41 journals across 21 publishers, including Elsevier and Springer Nature.

COPE recommends that journal editors wait for results of institutions' probes before acting, and lack of such an investigation stalled action on at least some papers. Editors at the *Journal of Bone and Mineral Metabolism* (*JBMM*) told the investigators that they recommended retracting 11 papers. Yet the publisher, Springer Nature, retracted just one, declining to retract others on grounds that an institutional investigation had not been completed. (Across its journals, Springer Nature has retracted 13 of the 45 Sato-Iwamoto papers it published.) Co–Editor-in-Chief Toshio Matsumoto of the University of Tokushima told *Science* that Keio University, where Iwamoto was a lecturer, has not responded to inquiries from the journal. He calls the impasse frustrating, but says, "We're stuck."

The new study casts doubt on whether publishers should in fact wait on these investigations: After institutions concluded that 84 of the papers were not problematic, publishers nevertheless retracted two-thirds of them. Many institutions are conflicted about criticizing their scholars, Grey says.

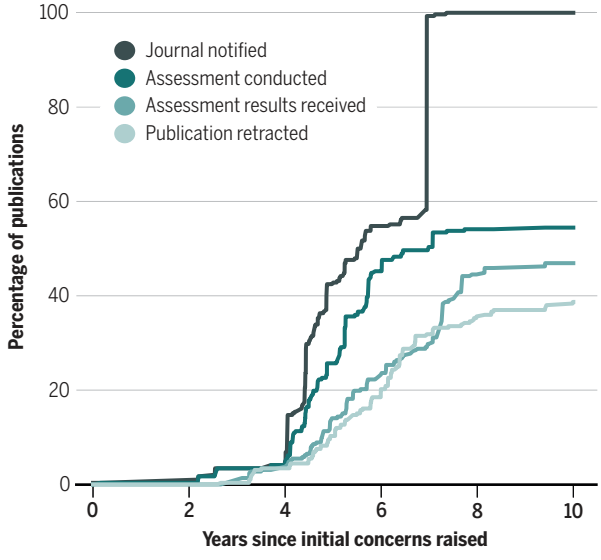
In written responses to questions from *Science*, Chris Graf, director of research integrity at Springer Nature since 2021, says the company is prepared to act on suspicious papers in the absence of institutional findings. Describing its decision not to retract the 10 papers in *JBMM*, Graf cited a different consideration, the company's need to prioritize scrutiny of "papers with well articulated concerns" that are "valid, specific, and actionable." The investigators' allegations of implausibly high research productivity by Sato and Iwamoto in those papers put them "at the lower end of the priority list." (Grey says the team's critique of those papers uncovered additional flaws.)

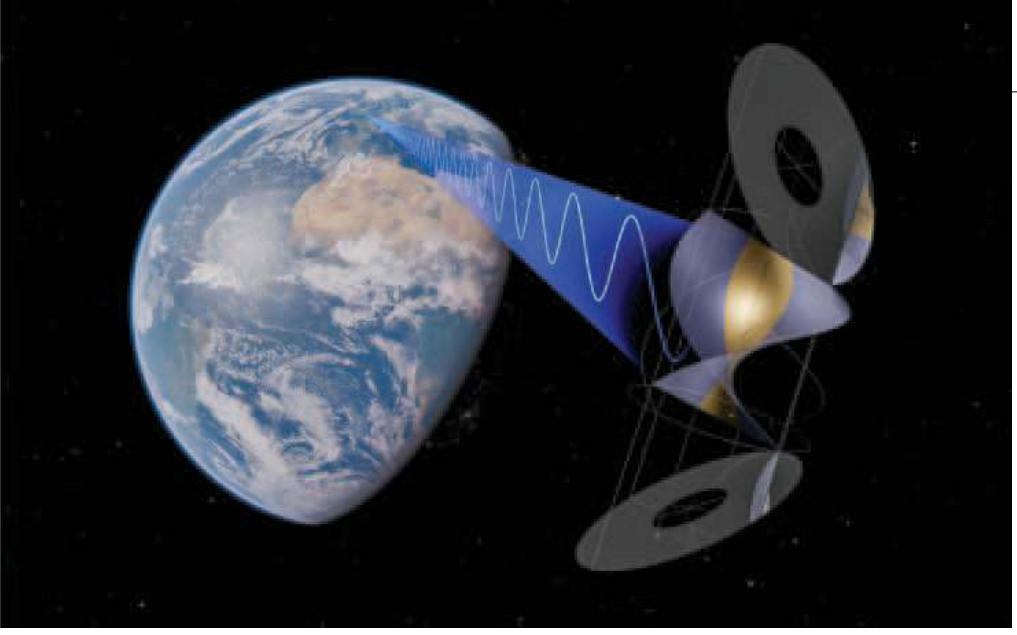
For the papers Springer Nature has retracted, Graf concedes, "We should have acted faster to assess and act, where appropriate, on the issues that have been identified." The age of some of the papers, several published more than 2 decades ago, complicates matters, he adds.

Grey wonders how many of Sato's and Iwamoto's papers would have been retracted if his group had not pestered the journals and their publishers. Almost never did a journal retract a paper or tell the team that it was investigating before the team made contact, he says. But as years have passed, responses dwindled. "There was a growing sense that they actually were a bit sick of us," he says. So, a decade after they began to prod journals and publishers about the Sato-Iwamoto papers, Grey says he and his colleagues "have had enough. We're not doing this anymore." ■

Ten years of prodding

Starting in 2013, watchdogs asked 78 journals to review nearly 300 papers authored by medical researchers Yoshihiro Sato and Jun Iwamoto. They recorded when journals disclosed that they had assessed a paper; when the watchdogs were informed what, if any, action each journal took or planned to take; and when papers were retracted.





NASA says it would cost $276 billion to build an orbiting power station.

ENERGY

Is NASA too pessimistic about space-based solar power?

Agency says orbiting power stations would be costly, but rocketry breakthroughs could make it economical

By **Daniel Clery**

This month, NASA cast a shadow on one of the most visionary prospects for freeing the world from fossil fuels: collecting solar energy in space and beaming it to Earth. An agency report found the scheme is feasible by 2050 but would cost between 12 and 80 times as much as ground-based renewable energy sources. Undaunted, other government agencies and companies are pushing ahead with demonstration plans. Some researchers say NASA's analysis is too pessimistic.

"There are assumptions that are just wrong and others that are incredibly conservative," says Martin Soltau, co-CEO of Space Solar, a company funded by the U.K. government and industry. "There's no imagination." He notes that NASA itself says slightly rosier assumptions—including a drop in launch costs that many think is within reach—would make the technology competitive with renewable sources on Earth.

Space-based solar power has many charms. For one, there are no clouds in space, and, in the right location, no night. In geostationary orbit, arrays of solar panels can track the Sun and gather energy 24/7, sending it to Earth in microwave beams gentle enough to avoid frying birds and airplanes. With free real estate, the orbiting structures can be made big enough to produce a few gigawatts (GW), rivaling the output of a nuclear or coal-fired power plant. Lifting thousands of tons of material into orbit is the main problem. NASA studied the idea in the 1970s but found that with space shuttle launches and astronaut assembly it was prohibitively expensive.

Advances in robotic assembly and sharp drops in the costs of solar panels and rocket launches are prompting space agencies to take another look. NASA, for instance, examined the life cycle cost of electricity for a 2-GW orbiting power station, in two configurations: one that uses steerable mirrors to concentrate light onto photovoltaic cells and converts the energy into microwaves for beaming to Earth, and another that uses "sandwich panels," with solar cells on one side and a microwave transmitter on the other. The more flexible mirror system can beam power 99% of the time, whereas the flat panels are limited to 60% by the need to face the Sun.

The report found the mirror configuration was more cost-effective. But even it would require lifting 5900 tons to orbit and more than 2300 rocket launches. Launch costs account for 71% of the total price tag of $276 billion.

NASA is counting on Starship, a fully reusable giant rocket under development by SpaceX that will be capable of lofting up to 150 tons at a time to low-Earth orbit. The company's partially reusable Falcon 9 rocket has already revolutionized the launch business since its debut in 2010, lowering launch costs from upward of $7000 per kilogram of payload to less than $3000 per kilogram. "Once Starship is operational, it's all going to change again," says Laura Forczyk of Astralytical, a space industry consultancy.

After talking with industry experts, NASA settled on $1000 per kilogram of payload carried by Starship, says the report's lead author, Erica Rodgers of NASA's Office of Technology, Policy, and Strategy. That figure looks on the high side to others. A SpaceX adviser told a conference last year the company would achieve a figure of $200 per kilogram. In its projections, the European Space Agency (ESA) has used launch cost estimates of $300 to $500 per kilogram, says Sanjay Vijendran, who heads the agency's Solaris program.

NASA also assumed that for every launch of hardware into low orbit another 12 would be needed to supply fuel for rockets to transfer the hardware to much higher geostationary orbits. Soltau says this has "a massive multiplying effect" on cost and that solar-powered space tugs could transfer the hardware much more cheaply—although more slowly. He also points out that NASA compared the cost of first-of-a-kind space hardware against mature wind and solar technologies on Earth. When NASA adopted rosier assumptions—$500 per kilogram launch costs, electric space tugs to boost orbits, and cheaper hardware—it found that space-based solar power was not only just as cheap as ground-based renewable energy, but also just as green, in terms of its life-cycle greenhouse gas emissions.

Given its pessimistic bottom line, the NASA report recommends proceeding cautiously. But others are pushing ahead. Last week, researchers at the California Institute of Technology announced the completion of a $100 million mission, funded by philanthropists Donald and Brigitte Bren, which tested transmitting power in space using a microwave beam.

In 2025, the U.S. Air Force Research Laboratory and Japan's space agency will each test beaming power from a spacecraft to the ground. ESA is seeking funding from its member states for technology development. Space Solar is asking the U.K. government to fund a $800 million, 1-megawatt orbiting demonstrator. "Space has a huge role to play in achieving net-zero emissions," Soltau says. "NASA should absolutely be at the forefront of this." ■

IMAGE: SPACE SOLAR GROUP

FEATURES

WATER BATTERIES

Pumped storage hydropower plants can bank energy for times when wind and solar power fall short

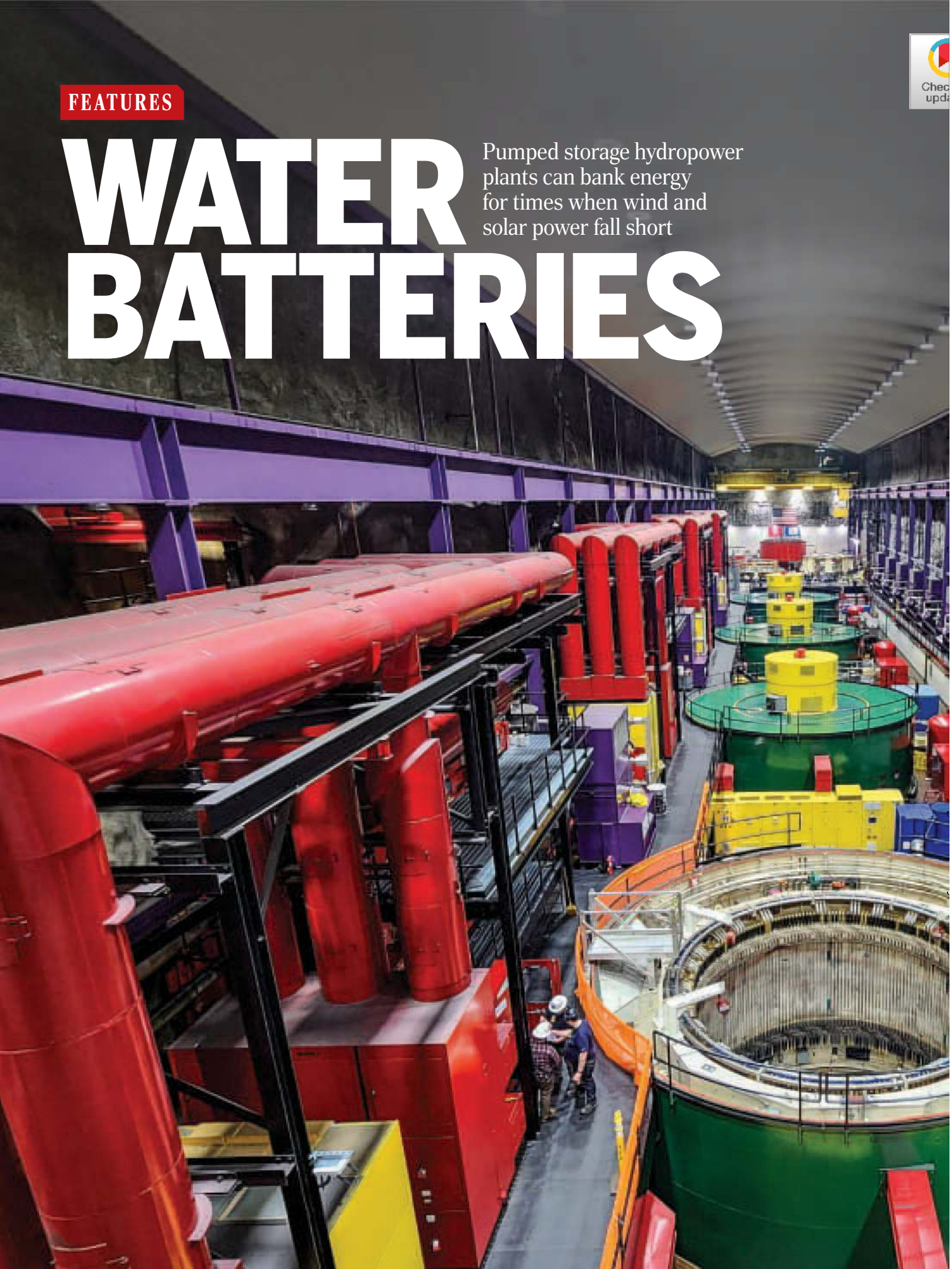


PHOTO: TENNESSEE VALLEY AUTHORITY

In an underground powerhouse, four reversible turbines (green cylinders) pump water to the top of Raccoon Mountain—and generate 1700 megawatts of electricity when it comes back down.

By **Robert Kunzig**

The machines that turn Tennessee's Raccoon Mountain into one of the world's largest energy storage devices—in effect, a battery that can power a medium-size city—are hidden in a cathedral-size cavern deep inside the mountain. But what enables the mountain to store all that energy is plain in an aerial photo. The summit plateau is occupied by a large lake that hangs high above the Tennessee River, so close it looks like it might fall in.

Almost half a century ago, the Tennessee Valley Authority (TVA), the region's federally owned electric utility, built the lake and blasted out the cavern as well as a 329-meter-tall shaft that links the two. "It was quite an effort to drill down into this mountain, because of the amount of rock that's here," senior manager Holli Hess says dryly. The cavern holds a candy-colored powerhouse, filled with cherry-red electrical ducts and vents and beams in a pale grape. Four giant cylinders, painted bright green and yellow, are the key machines: Each one houses a turbine that becomes a pump when it spins the other way, and a generator that is also an electric motor.

At night, when demand for electricity is low but TVA's nuclear reactors are still humming, TVA banks the excess, storing it as gravitational potential energy in the summit lake. The pumps draw water from the Tennessee and shoot it straight up the 10-meter-wide shaft at a rate that would fill an Olympic pool in less than 6 seconds. During the day, when demand for electricity peaks, water drains back down the shaft and spins the turbines, generating 1700 megawatts of electricity—the output of a large power plant, enough to power 1 million homes. The lake stores enough water and thus enough energy to do that for 20 hours.

Pumped storage hydropower, as this technology is called, is not new. Some 40 U.S. plants and hundreds around the world are in operation. Most, like Raccoon Mountain, have been pumping for decades.

But the climate crisis is sparking a fresh surge of interest. Shifting the electric grid away from coal and gas will require not only a lot more solar panels and wind turbines, but also a lot more capacity to store their intermittent output—to keep electricity reliable when the Sun doesn't shine and winds are calm. Giant versions of the lithium-ion batteries

in electric vehicles are also being deployed on the grid, but they're too expensive to do the job alone. Dozens of new technologies, including different battery designs, are at various points on the road from lab bench to commercialization.

Pumped storage, however, has already arrived; it supplies more than 90% of existing grid storage. China, the world leader in renewable energy, also leads in pumped storage, with 66 new plants under construction, according to Global Energy Monitor. When the giant Fengning plant near Beijing switches on its final two turbines this year, it will become the world's largest, both in terms of power, with 12 turbines that can generate 3600 megawatts, and energy storage, with nearly 40,000 megawatt-hours in its upper reservoir.

In the Alps, where pumped storage was invented in the late 19th century, Switzerland opened a plant in 2022 called Nant de Drance that can deliver 900 megawatts for as long as 20 hours. Austria, too, has ambitious plans. Down in Australia, one of two new plants already under construction will be the new record holder for energy, storing enough to supply 3 million people for 1 week. Called Snowy 2.0, it's scheduled to open by 2029.

"When people talk about batteries—these are little things," says Andrew Blakers of Australian National University, a solar-cell pioneer who has become an influential pumped storage evangelist. "And little Australia, where the population is smaller than California, has a single pumped-hydro system under construction that will be bigger than all the utility batteries in the whole world combined." It's not that Australia is particularly blessed by geography, Blakers says. From satellite data he and his team have compiled a global atlas showing about 1 million sites across all the continents that would be technically suitable for pumped storage.

Even in the United States, where no large pumped hydro facility has been constructed since the 1990s, the federal government is providing encouragement. A 2022 study by the National Renewable Energy Laboratory (NREL), a Department of Energy (DOE) lab, identified more than 14,000 potential sites for "closed-loop" plants, where both reservoirs are placed off-river to minimize environmental impacts. The 2022 Inflation Reduction Act has made generous tax credits available to pumped storage, as it does for renewables. TVA has begun what's likely to be a decadelong process to build another facility like Raccoon Mountain.

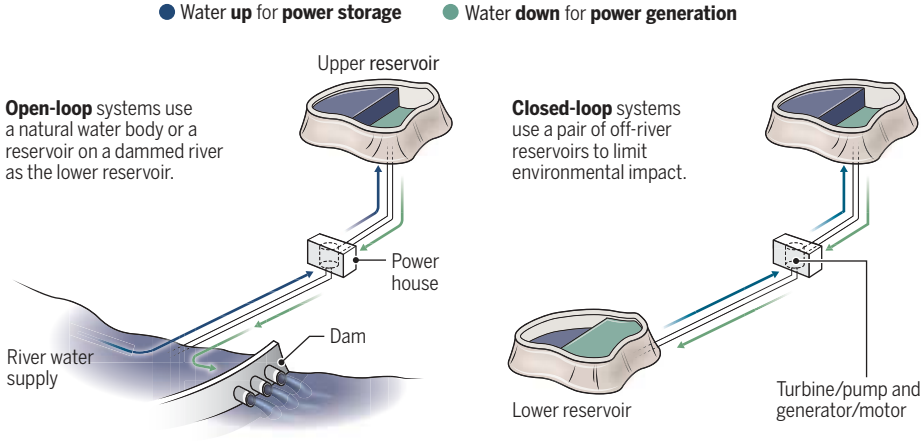
The Federal Energy Regulatory Commission (FERC) has issued dozens of preliminary permits, mostly in the mountainous West, to utilities and developers that want to stake claims to potential pumped storage sites. Three developers have completed the costly multiyear process to receive a FERC license, meaning their projects are shovel-ready. But none has begun construction, and it's far from clear the United States will share in the global boom.

The impact of these massive projects on the land and environment is one reason. But the bigger problem is that pumped storage is an enormous long-term investment—more than $2 billion for a large plant, according to a recent NREL estimate—and in the U.S. electricity market, the returns on that investment are uncertain. "Bankers and investors and utilities are thinking, 'I know there's a great value here, but can I quantify it?'" says Patrick Balducci, an economist at DOE's Argonne National Laboratory. "'Is this just going to reduce emissions and improve reliability and benefit everyone throughout the region—and I never get paid for it?'"

WHEN TVA BUILT Raccoon Mountain in the 1970s, the case for pumped storage was simpler. At the time the agency was also build-

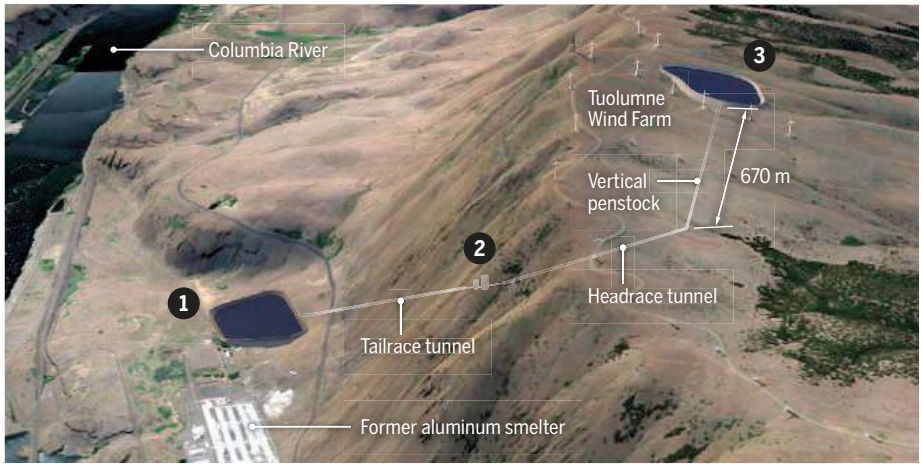
Reservoirs for green electricity

Electricity can be stored by using it to pump water from a low-lying reservoir into a higher one. When power is needed, the water flows back down and spins a turbine—often the pump, spinning in reverse. The flow rate and the elevation difference determine the power output, and the volume of the upper reservoir determines how much energy is stored—and thus how long the water battery lasts.



A technically perfect but contested site

With a 670-meter drop between the reservoirs, Rye Development's planned facility near Goldendale, Washington, could offer "12 hours of on-demand renewable electricity to every residence in Seattle," says Erik Steimle of Rye. Although on private property, it would partially occupy an area sacred to the Yakama Nation, which opposes the project.



1 Lower reservoir
On an old industrial site, it would be bounded by a 62-meter-high dam. Filled once from the Columbia River, it would be replenished as needed to make up for evaporation.

2 Underground powerhouse
A 137-meter-long cavern, joined by 9-meter-wide water tunnels to the reservoirs, would house three pump turbines with a total capacity of 1200 megawatts.

3 Upper reservoir
Surrounded by wind turbines and ranches, but also by Yakama food-gathering and heritage sites, it would be some 600 meters across and bounded by a 53-meter-high dam.

ing nuclear reactors, which are designed to run 24/7. Raccoon Mountain could pump at night when electricity was cheap and regenerate during the day when it was expensive. The economic benefit of such "energy arbitrage" was clear and drove the construction of many other pumped storage plants.

Today, with the growth of wind and solar power, the rationale has shifted. Grid operators increasingly need storage to meet their central challenge: balancing electricity supply against fluctuating demand every minute, day, and season. They do that now mostly by adjusting power generation at fossil fuel plants, which can be turned on and off as needed. Wind and solar aren't "dispatchable" that way; indeed their capricious ebbs and flows aggravate the balancing problem. But stored energy can help match renewable power to demand and allow coal and gas plants to be retired.

For now, lithium-ion batteries are filling the need. In places such as California they're starting to replace the gas "peaker" plants that utilities turn on to meet the demand peak that arrives in the late afternoon, just as solar power begins to dip. For that purpose—a few hundred megawatts of extra power for a few hours—a lithium battery plant is much cheaper, easier, and quicker to build than a pumped storage plant, says NREL senior research fellow Paul Denholm.

But a few hours of energy storage won't cut it on a fully decarbonized grid. Winter, especially, will tax renewable power, Denholm says. As people switch from gas heat to electric heat pumps, winter demand for electricity can begin to rival the summer peak caused by air conditioning. But whereas a summer peak usually subsides within a few hours as nightfall brings relief, a winter peak triggered by a cold snap can persist for much longer.

"In the end, the storage requirement is driven not by the summer afternoon air conditioning peak," Blakers says. "It's driven by a wet, windless week in winter. Try and do that with batteries." As you add more and more of them, each module as expensive as the last, the cost eventually becomes prohibitive.

Jeremy Twitchell and his colleagues at DOE's Pacific Northwest National Laboratory modeled how California would fare if it were to rely solely on expanding solar and wind power to meet its goal of a carbon-free grid by 2045. A nearly fivefold expansion would be enough to meet demand on an annual basis, they found, but it would lead to huge temporary excesses and shortfalls, including deficits as big as 30 gigawatts, the output of 15 Hoover Dams. The average shortfall would last nearly 15 hours.

"What that points to is that long-duration energy storage is an absolute necessity in a decarbonized grid," Twitchell says.

BLAKERS DID PIONEERING work on solar cells and helped accelerate the turn to renewables. But he felt countries wouldn't fully embrace green energy until they were convinced the grid will remain reliable. In 2015 he dropped his photovoltaic work to devote himself to the one technology he says is up to the task and available right now. "That's pumped hydro. Everything else is arm waving."

A massive penstock carries water between two reservoirs at Nant de Drance, a 900-megawatt plant in Switzerland.

PHOTO: FABRICE COFFRINI/AFP VIA GETTY IMAGES

His own country's leadership is convinced. Australia, the world's leading coal exporter and still dependent on the stuff itself, has committed to getting 82% of its electricity from renewables by 2030, more than doubling renewable capacity in just 7 years. To enable that expansion, the government is also investing heavily in pumped storage. More heavily than it had hoped, in fact: The gargantuan Snowy 2.0 project in New South Wales has been beset by delays and cost overruns.

The site, in a national park, already has two large hydroelectric reservoirs at different elevations that just needed to be connected by tunnels. But that connection is 27 kilometers long—which increases the risk of geologic surprises. Sure enough, one of Snowy's three tunnel-boring machines spent almost all of 2023 stuck in soft rock less than 200 meters from its starting point. In the summer, the government announced that the project's cost had ballooned to AU$12 billion.

Nevertheless, Snowy 2.0 will store 350,000 megawatt-hours—nine times Fengning's capacity—which means each kilowatt-hour it delivers will be far cheaper than batteries could provide, Blakers says. Yet his atlas shows that Australia has many sites more technically ideal than Snowy 2.0.

The ideal is a site that maximizes the vertical distance between the two reservoirs—the "head"—while minimizing the horizontal distance. "Everything just gets better as you go for larger head, because the pressure of water is bigger," Blakers says. Double the head and you can double the power capacity and the energy stored—or shrink the reservoirs, tunnels, and turbines.

In Queensland, Australia's largest coal-producing state, the government created a special organization, Queensland Hydro, to build pumped storage. Last year, it announced it would commit AU$14.2 billion to construct a 2000-megawatt, 24-hour plant above Lake Borumba, 1 hour north of Brisbane, and another AU$273 million to investigate Pioneer-Burdekin, a second site farther to the north that had emerged as a favorite from Blakers's atlas.

"It is an extraordinary site, it really is," says Chris Evans, the Queensland Hydro executive in charge of development. With nearly 700 meters of head and only 3.5 kilometers of horizontal distance between the intended reservoirs, Pioneer-Burdekin could generate 5000 megawatts for 24 hours, making it the world's most powerful. Together with Borumba, it could meet Queensland's typical demand on a rainy winter day and night. A decision on whether to proceed with the project is due later this year.

But the Queensland government, which operates 8000 megawatts of coal-fired power plants, is already committed to pumped storage as a cornerstone of its energy transition. The public ownership "is a real benefit about the electricity system, particularly in Queensland," Evans says. "It's enabling a smoother transition."

"MOST PUMPED STORAGE projects being built today are by these quasi-government setups," said Ushakhar Jha. Rye Development, the hydropower developer for which Jha is chief engineer, has been working for nearly a decade to get a project built privately. It holds one of the three outstanding FERC licenses, for a 400-megawatt project at Swan Lake in southern Oregon, and it's close to getting a license for a 1200-megawatt project near Goldendale, Washington, on the Columbia River Gorge. California, Oregon, and Washington state have all enacted grid-decarbonization deadlines. Rye smells a coming regional market.

In October 2023, I visited the Goldendale site with Jha and Michael Rooney, the firm's head of project development. On a blustery, overcast morning, we climbed up a gravel road through sagebrush steppe to Juniper Point, overlooking the Columbia River, to see where Rye plans to place an upper reservoir. Strong gusts drove the wind turbines high above us into a stately spin. All along this ridge and far across the river into the wheat fields of Oregon, the land was dotted with hundreds of white turbines. Far below us, the Bonneville Power Administration's John Day Dam interrupted the river.

Rooney and Jha explained why the site looked just about perfect to them. The landowner and local officials are eager to develop it. The lower reservoir, like the upper one about 600 meters across, would be built on the waste site of a derelict aluminum smelter. No new transmission towers would be required; a single 500-kilovolt line, attached to towers already built for the dam and the wind turbines, would connect the storage plant across the Columbia to the John Day substation, a gateway to utilities from Los Angeles to Seattle.

Finally, the project wouldn't require a single new road: The wind turbines and the smelter already have access roads. "This is a dream for hydro engineers like us, finding a site where you're only thinking about the specific core infrastructure," Jha said. The reservoirs would be barely 2 kilometers apart, with a head of 670 meters—close to ideal.

There's one major problem for the project: The original occupants of the land don't want it. The reservation of the Yakama Nation begins about 25 kilometers to the north, but Juniper Point, like most of central Washington, is on land the Native Americans were forced to cede to the U.S. in an 1855 treaty. The treaty reserved for them the right to continue fishing, hunting, and gathering food on the ceded land—and to the Yakamas, this part of the ridge above the gorge is sacred. Called Pushpum, it figures in their creation stories. Their ancestors gathered roots and shoots here, and some Yakamas still follow those traditions. Just last spring, Yakama fisheries biologist Elaine Harvey told me, her family celebrated her 8-year-old daughter's formal initiation to food gathering in a ceremony at the Rock Creek Longhouse. The little girl fed the foods she had gathered on Pushpum to the whole assembly.

Harvey and I were parked directly under a high-voltage transmission tower, on the north bank of the river, looking at the John Day Dam through a windshield wet with rain. A wooden fishing platform that her family still uses jutted into the river. This riverbank had been the site of her family village until the U.S. Army Corps of Engineers ordered it evacuated in 1957, when the Dalles Dam was completed 35 kilometers downstream. That dam drowned Celilo Falls, a fishing and trading hub that had been inhabited for 11,000 years. Roaring falls disappeared and were silenced under a lake.

To Harvey, the Goldendale pumped storage project is of a piece with that trauma. "They're going to build a 30-foot-diameter tunnel through the mountain, and that's our sacred mountain," she said. She and other tribal representatives stress they're not opposed to renewable energy—just to projects that damage their cultural heritage. "We're just trying to protect what we can, and people don't get it," she says.

FERC's draft environmental impact statement, released in March 2023, recommends licensing the Goldendale project. But it acknowledges that the plan would destroy five presettlement archaeological sites, interfere with Yakama food gathering, and change the visual feel of the place. It's not clear that those harms can be remedied. "We're not going to settle for mitigation," says Yakama Nation Tribal Council member Jeremy Takala. "We already know there is no way." The Columbia Riverkeeper, the Sierra Club, and other environmental groups are backing the tribe.

With its need for manhandling mountains, pumped storage inevitably risks exciting local opposition. But in general, that's not the biggest barrier to new facilities being built in the U.S. The market is.

Many utilities are interested in pumped storage, Balducci says, but the models they use to plan investments don't capture all the benefits it provides to the grid—let alone to the environment. He and his colleagues analyzed the Goldendale project and found that it would improve the overall stability of the Western grid and be "a key enabler" of the expansion of solar and wind energy needed to meet zero-carbon electricity targets. The problem is, although the grid will surely need more long-duration storage in coming decades, it doesn't need more yet, making utilities reluctant to commit.

"The market is incentivizing what the current grid needs," Denholm says. "Right now we need 4-hour storage. The market is not incentivizing what we might need 5 years from now." New pumped storage plants take longer than that to license and build, cost billions, and can last a century—a virtue, but also a commitment that takes nerve in a rapidly changing market.

IT'S POSSIBLE UTILITIES will be spared that choice by long-duration storage technologies that are still being developed. Pumped storage might be superseded by flow batteries, which use liquid electrolytes in large tanks, or by novel battery chemistries such as iron-air, or by thermal storage in molten salt or hot rocks. Some of these schemes may turn out to be cheaper and more flexible. A few even rely, as pumped storage does, on gravity.

The Yakama Nation favors one of those. The tribe is in conversation with a company called ARES, for "advanced rail energy storage," which this year plans to put its technology to a major test in an abandoned gravel quarry in Pahrump, Nevada. An electric motor-generator will haul a 340-ton concrete mass up a 50-meter-tall hill on a railcar; the energy released when the car rolls back down will generate 5 megawatts. The system doesn't require water or tunneling and so might be easier to site and have less permanent impact than pumped storage. It's "getting the advantages of pump storage without the disadvantages," says Russ Weed, chief development officer of ARES.

Power and energy could be increased in steps, by adding more rails, motor-generators, and cars. The Yakamas think an old landfill on their reservation could be a good site for a 500-megawatt system, and have applied for DOE grants to study it. "This isn't just a Yakama Nation solution, this is a state of Washington solution," says Ray Wiseman, head of Yakama Power, the tribe's utility.

Another gravity-based energy storage scheme does use water—but stands pumped storage on its head. Quidnet Energy has adapted oil and gas drilling techniques to create "modular geomechanical storage."

At Raccoon Mountain, west of Chattanooga, Tennessee, the upper reservoir is some 300 meters above the Tennessee River (left). The powerhouse is under the mountain.

PHOTO: TRACEY TRUMBULL

Energy is stored by pumping water from a surface pond under pressure into the pore spaces of underground rocks at depths of between 300 and 600 meters; electricity is generated by uncapping the well and letting the water gush to the surface and spin a turbine. The energy is stored not in the water itself, but in the elastic deformation of the rock the water is forced into.

Quidnet says it has conducted successful field tests in several states and has begun work on its first commercial effort: a 10-megawatt-hour storage module for the San Antonio, Texas, municipal utility. It should be online in 2025, CEO Joe Zhou says. Unlike pumped hydro, geomechanical storage doesn't carry the cost of tunneling, dam building, or getting a FERC license. And the technique exploits existing oil-and-gas technology. "We ourselves are repurposed oil and gas people," Zhou says.

IF ANYONE SHOULD be able to repurpose pumped storage for the era of renewables and get a new plant built, it's TVA. As a federal agency, it doesn't need a FERC permit. As a self-financing, vertically integrated utility responsible for delivering power to 10 million people in the Tennessee Valley, it can capture the benefits of pumped storage regardless of whether the market knows how to price them. But it does have to complete an environmental impact statement.

One morning last fall, at a site TVA is now considering in Pisgah, Alabama, project manager Scottie Lee Barrentine was studying black-and-white pictures of the construction of Raccoon Mountain. He was trying to learn more about how his predecessors had managed the challenge. "Nobody's around anymore," he says. Pisgah sits on top of a long ridge called Sand Mountain, about 80 kilometers downriver from Raccoon Mountain, and Barrentine's field headquarters was an empty wedding venue next to the potential location of an upper reservoir. The terrace offered an expansive view north across the Tennessee River. Like Raccoon Mountain, the Pisgah project would draw water from a TVA reservoir on the river itself.

TVA values Raccoon so much, a senior executive once told me, it might one day consider building two or three new pumped storage plants. Barrentine is hoping to deliver at least one, but it will take a decade if it happens at all. The decision won't be made until 2025, after the environmental impact statement. The plant would then take at least 8 years to design and build.

The environmental review is intended to reveal any reason not to build. Drill crews are looking for anything that might make tunneling hazardous. Biologists are combing the site for endangered species such as bats. Archaeologist Sarah Stephens and a team of 11 are digging shovel holes every 30 meters, 20,000 holes in all, looking for "anything from grandma's trinket to Native American arrowheads." There is no doubt, she says, that the Muscogee (Creek), Cherokee, and other Native Americans occupied this site at least occasionally for millennia. But they were mostly driven from the area in the 1830s, west to Oklahoma along the Trail of Tears.

Across the river from the wedding venue, the cooling towers of TVA's Bellefonte nuclear power plant rose on the far bank. No steam was billowing from them. TVA never quite finished the plant back in the past century; it had overestimated how fast demand for electricity would grow. It was a cautionary message for pumped storage hydropower: Projects that seem foresightful today may prove to be myopic—or too far ahead of their time.

TVA did, however, complete the high-voltage transmission line connecting the nuclear plant to a transmission artery south of the river. That line crosses the possible pumped storage site at Pisgah, and it may yet come in handy, Barrentine says. "I hope it will be energized one day." ■

Robert Kunzig is a journalist in Birmingham, Alabama.

INSIGHTS

POLICY FORUM

URBAN SCIENCE

Designing policy for Earth's urban future

Global impacts of cities must be better conveyed to multilateral organizations

By **Jessica Espey[1], Michael Keith[2], Susan Parnell[1], Tim Schwanen[3], Karen C. Seto[4]**

Although the importance of cities has been recognized through international agreements such as the 2030 Agenda for Sustainable Development, the worldwide impact of urban growth upon all Earth systems is not well recognized by the international policy community. Collectively, cities drive global change at an unprecedented scale, transforming land cover, hydrological systems, climate, biogeochemistry, and habitats. Cities are the nucleus from which humanities' impact on all Earth systems can be observed. One would thus expect urban dynamics and impacts to be at the top of global governance agendas. We argue that one key factor that contributes to this lack of recognition is the absence of a global-level urban science advisory system, which could support the United Nations (UN) and regional multilateral groups with international policy-making. Achieving such a system requires the acknowledgment of three things: aggregate or cumulative impacts of urbanization globally, urban blind spots in present international policy-making, and diversity and potential contributions of urban science.

The importance of recognizing cities and local governments in international policy-making has been highlighted before (*1*), but

People crowd a street in Tokyo's Harajuku district.

with little specific attention to the worldwide effects of an urban human species. Urban areas contribute 65 to 75% of global greenhouse gas emissions (*2*). Urban land areas will either double or triple between 2015 and 2050, and the building of new cities will require vast amounts of raw materials such as sand, metals, and wood, the acquisition of which will transform ecosystems all over the world. If humankind continues to build cities in the way that many have been designed and constructed over the past century—low density, energy and material intensive—more raw materials will be required than the planet can sustainably provide (*3*). And this is only to build tomorrow's cities, not power them.

Although there is an existing UN agency (UN Habitat) and there are various parallel initiatives underway to highlight urban challenges [such as a forthcoming special report by the Intergovernmental Panel on Climate Change (IPCC) on cities and climate change], these mechanisms are insufficient and tend to focus on single issues (e.g., climate change) or on cities' internal problems, which fails to convey the multifaceted ways in which cities are shaping the future of the planet. Further, many existing mechanisms [for example, highly organized city networks such as United Cities and Local Governments (UCLG)] tend to focus on city powers—that is, the power and authority of cities to affect change within nation-states—rather than on the aggregated power of cities to determine not only social and economic changes but also planetary-level environmental ones. Although these existing, parallel processes are important, they are failing to bring the seismic effects of urban change on the world to the attention of policy-makers.

A global-level urban science advisory system should look to change this, speaking to the influence of urban dynamics on all global systems and supporting policy-makers to design policy that is appropriate for a sustainable, urban planet. It need not be expensive and cumbersome, like the IPCC, and instead could mimic the likes of the long-standing and cost-effective Committee on Development Policy under the UN's Economic and Social Council (ECOSOC). Above all, it should provide a standing formal mechanism for the inputs of the ever-growing, highly diverse, urban science community to be synthesized and relayed to policy-makers in accessible, timely, and policy-relevant formats. To be clear, we are not suggesting that this be an advisory system concerned with local implementation of international policy commitments, for which there are countless, highly organized networks, nor that it be a scientific advisory group concerned with discrete city challenges, or indeed just a matter of giving cities "a seat at the top table," though this would undoubtedly help (*1*). This is about ensuring that world leaders and policy-makers have the information that they need at their fingertips to design a world that reflects and responds to humanity's urban future.

AGGREGATE EFFECTS OF GLOBAL URBANIZATION

The complexity and positive and negative consequences of urban development processes have been studied widely and constitute the central concerns of a proliferating literature that calls itself urban science (*4*, *5*). Although this literature is highly heterogeneous, it shares common attributes such as presenting cities as adaptive and open complex systems, or systems of systems (*5*), and presents useful conceptual approaches for integrating transdisciplinary analysis at the urban scale (*6*). Nonetheless, less has been written about the specific dynamics of the collective effects of urbanization and how the simultaneous mass expansion of urban environments affects all Earth systems (the lithosphere, hydrosphere, biosphere, and atmosphere), as well as the planet's social and economic cohesion, or lack thereof.

Perhaps best understood are the effects of cities on greenhouse gas emissions. According to the IPCC, urban areas collectively contribute about three-quarters of carbon dioxide equivalent emissions from final energy use (*2*). Less well known are the effects of urban agglomeration on ecosystem management. Worldwide, urban land expansion is one of the primary drivers of habitat and biodiversity loss (*7*). Biodiversity loss occurs not only because of the total land being reclaimed and occupied by cities but also because of the increasing fragmentation of the remaining nonurban land, which interrupts wildlife and ecological zones and increases risks from fire, pests, and diseases that may more easily spread across space (*8*). The effects of urban expansion on biodiversity loss are not only due to emissions, waste, and land use; emerging threats include those associated with the uptake of energy-efficient technologies such as light-emitting diode (LED) lighting and energy-efficient homes (*9*). Many of these effects will be long-lasting, if not irreversible. However, the direct effects of urbanization on land systems and biodiversity are only half of the equation. By some measures, 80% of the global gross domestic product is generated in cities. The long supply chains required to build, power, and feed cities means that even rural economies, agricultural systems, and livelihoods far distant from urban areas are affected by urbanization.

The processes of change are happening locally, under the remit of town planners and city councils, and it is at this level that preventative action will have to be taken; however, local actions need to be guided by collective global analysis, reflection, and response, which can identify global trends that are not discernible in discrete locations, such as regional or global biodiversity loss, global urban heat island effects, and much more. Such reflection and policy discussion are distinctively suited to deliberation in regional and global repre-

[1]School of Geographical Sciences, University of Bristol, Bristol, UK. [2]Centre on Migration, Policy and Society, University of Oxford, Oxford, UK. [3]School of Geography and the Environment, University of Oxford, Oxford, UK. [4]Yale School of the Environment, Yale University, New Haven, CT, USA. Email: jessica.espey@bristol.ac.uk

sentative forums, such as the UN General Assembly (UNGA). The UNGA is the preeminent site of international sustainable development policy-making and is the most representative intergovernmental forum in which worldwide phenomena can be discussed and addressed through coordinated global intervention. To be clear, the UNGA is not the vehicle to design detailed policy, but it is the place to highlight important scientific discoveries that have cross-country implications and to help national and subnational policy-makers coordinate their responses.

INVISIBILITY OF URBANIZATION IN INTERNATIONAL POLICY

The Sustainable Development Goals (SDGs), Paris Climate Agreement, and Habitat III were major coups for the global urban community, where, thanks to efforts by local government actors, their networks, and coalitions of concerned urban scientists, heads of state and government recognized cities and urban environments as epicenters of many, if not most, 21st-century sustainable development challenges. The SDGs also recognized the instrumental role of local governments and authorities in the implementation of sustainable development responses. But achieving these policy victories depended heavily on individuals exploiting their personal political connections and social capital to seek informal channels to communicate key evidence and ideas and influence negotiations. Such an ad hoc effort is not a sustainable approach for helping to inform global policy with urban concerns. A more systematic, institutionalized approach is needed.

Thus, despite the positive momentum around the SDGs, political attention to cities at the global level has waned since 2015. Although it was widely acknowledged that cities were on the frontline of COVID-19 and its response, the disproportionate burden they faced was not reflected in the outcome statement from the G20 summit on COVID-19 in 2020 (*10*). Likewise, urban governance challenges of migration, development, and health were ignored in the Declaration on the Commemoration of the 75th Anniversary of the United Nations by the UNGA in 2020, which specifically discussed the necessity of reinvigorating multilateralism to help deal with modern social, economic, and environmental crises (*11*). In August 2023, a secretary general's scientific advisory panel was announced at the 78th UNGA session, but it lacks a representative that can speak to urban science, which highlights this as an international policy blind spot. This omission risks a lack of attention not only to cities' multiple challenges but also, and importantly, to the worldwide effects of urban expansion. There are various parallel initiatives underway to highlight urban challenges in other forums, but these mechanisms tend to focus on single issues or city-specific challenges, which fails to convey the multifaceted ways in which urban expansion determines the future of planetary and global social systems.

The underrepresentation of cities in international policy processes partly reflects mandates, with many national governments arguing that these dialogues are the sole purview of national-level representatives. This Westphalian argument has some historic credence (among the Western nations who founded many of these institutions) but dismisses the huge political and economic power of cities in the 21st century. Better representation of local actors would undoubtedly go a long way to elevate urban concerns and demonstrate the potential of cities as sites for transformative change. Furthermore, it may help to draw the attention of international policy to seemingly local processes with potentially major implications when aggregated to the global scale, for example, cities affecting regional hydrological cycles and rerouting waterways or actors at the urban level removing vegetation that offers cooling and altering surfaces with concrete, asphalt, and other heat-trapping materials that collectively create urban heat islands.

IMPROVING THE POLICY RELEVANCE OF URBAN SCIENCE

Appropriate responses to macroscale urban change and its impacts on Earth systems requires not only listening to diverse urban stakeholders but also encouraging national leaders to heed the aggregated science on global urbanization. This demands more than increased representation; it requires a well-functioning urban science-policy interface. This interface would coordinate insights from various existing but fragmented urban initiatives—such as those under the IPCC, G7, and G20—as well as actionable insights from city networks and synthesize these with the latest science of planet-wide urbanization, helping to account for and monitor global phenomena and communicate the impacts of global urban change to heads of state and government and their appointed international deliberators. It would elevate and synthesize existing knowledge and make it readily accessible to all global policy-makers through one clear, empowered, and legitimate mechanism.

A key secondary benefit of such a platform is that it would help to coordinate and focus the inputs of the international urban science community on global-level urban challenges and align their inputs with policy opportunities and influencing windows. As many have argued, the broad and ever-growing urban science literature is disparate and fragmented (*3*, *5*). Urban science is not one coherent field or discipline but rather a loose collective of people working across very different sectors, from public health or energy to ecology and infrastructure, which in and of itself is a reflection of the far-reaching effects of urban change. Urban science's fragmentation is even more acute and problematic when planetary-level changes are considered. The broad nature of the research has prevented the emergence of very clear headline messages about the value and contribution of this global urban science for international decision-making. It has also hindered meaningful conversations across sectors about urban effects in areas like health and biodiversity.

Lessons from evidence-informed policy-making in other sectors suggest that it is hard to embed complex scientific concepts and ideas within high-level policy forums without coherent epistemic communities mobilizing around science-based advocacy messages and targeted, high-level engagement with policy officials. Without such communities, it is even harder to craft meaningful solutions because the way the problem is framed can determine how the solution is conceptualized. Although there are emerging efforts to coordinate this broad community, outside of formal governance structures (e.g., through the creation of dedicated urban science journals), they have mostly been driven by academics aiming to learn from each other's research, not by policy-minded actors aiming to use this research to help shape global policy dialogues on a much broader range of challenges. Put simply, most academics have been looking inward toward their communities when they need to also look outward at how they can engage in global policy processes and help to change the world.

As the negotiations on the SDGs demonstrated, having a clear political opportunity and the ear of policy-makers can help to mobilize and organize scientists to put aside their technical differences and focus on the most pressing and transformative challenges ahead, thereby transforming existing urban science from a collection of technical learning and expertise into practical guidance for international policy design in the urban century. Doing this through an institutional mechanism with a clear political mandate would ensure not only that it is sustainable but also that policy-makers

Morning smog shrouds Kolkata, one of the most air-polluted cities in India.

PHOTO: SUDIPTA DAS/NURPHOTO VIA GETTY IMAGES

are forced to recognize the intractability and complexity of the urban challenge and listen to scientists and experts to devise evidence-informed policy. Such a platform for global urban science may also encourage transdisciplinary cooperation between scientists and policy-makers across all scales and sectors to identify transformative solutions.

URBAN SCIENCE ADVISORY SYSTEM FOR THE PLANET

What could such an international advisory system look like? Although expensive and cumbersome, the IPCC and the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) are often showcased as models of effective science-policy interfacing. We note, however, that "attempting to transfer this model of knowledge production to other issues is problematic" [(*12*), p. 125]. We also caution that observers working on a wide range of policy challenges may envy the IPCC for its financial and political symbolic power, without necessarily taking heed of its challenges. We do not, therefore, propose the IPCC model as a format for an international urban science advisory system. Instead, we offer five key principles and considerations that would allow for such an entity to be established in a range of ways, as decided by UN member states.

Examples of what the advisory system might look like include an entity affiliated with the office of the UN secretary general or as part of the existing Multi-stakeholder Forum on Science, Technology and Innovation (STI Forum) for the SDGs (although it must be noted, with exasperation, that there was an almost total absence of urban discussion within the 2023 STI Forum). It could operate much like the long-running Committee for Development Policy, established in 1965, which is a subsidiary body of the UN ECOSOC that is composed of 24 experts nominated in their personal capacity by the secretary general and appointed for a period of 3 years. The committee has advised ECOSOC members and the secretary general's office on topics as wide-ranging as migration and aid effectiveness and provided a number of inputs on the post-2015 development agenda that resulted in the SDGs.

Several key principles and considerations should be recognized while establishing such a mechanism. Learning lessons from the challenges of existing science initiatives, the advisory system should be inclusive of broad science and knowledge from institutions outside of the Global North and Western academic literature [acknowledging the coloniality of many existing science systems (*13*)] and with a clear typology of evidence inputs that explains the utility of different forms of knowledge for different purposes (for example, the value of Indigenous knowledge for understanding changing local ecologies versus the value of peer-reviewed large-sample survey evidence for designing durable cross-country policy responses).

The composition of the body should be transdisciplinary. The roster should include urban scientists who can provide inputs across a wide array of sectors and spatial scales, from urban neighborhood to planet. Academics from nonurban domains, whose research speaks to processes of urban change and who can meaningfully contribute to tackling "wicked," interwoven problems should also take part. Such researchers may include those working on biodiversity or human health within and across urban and rural domains.

The body should have a clear political mandate and specified entry points into UN deliberative processes (e.g., an invitation to submit findings to member states at the start of each UNGA cycle). It should also be tasked to engage with parallel multilateral dialogues such as the G20, G7, and regional economic commissions, which are often the sites of prenegotiation and compromise before issues are discussed among the full UNGA membership.

There should be an imperative for the body to produce one key output per UNGA cycle that relates to the annual core themes such as global urbanization and inequality, or global urbanization and food systems. This will allow the advisory system to be (seen as) policy responsive and dynamic.

Finally, the body should prepare outputs that carefully synthesize the breadth of existing knowledge and streamline it to convey essential information with relative brevity and accessibility for nonspecialist policy audiences.

Whatever mechanism is used, it is well past time for evidence-based dialogue on the planet-wide effects of urbanization at the highest levels of international governance. Our planet's future is an urban future, and our systems of international administration must reflect that. ■

REFERENCES AND NOTES

1. M. Acuto, *Nature* **537**, 611 (2016).
2. S. Lwasa *et al.*, "Urban systems and other settlements" in *Climate Change 2022: Mitigation of Climate Change. Working Group III Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2022).
3. M. Swilling *et al.*, "The weight of cities: Resource requirements of future urbanization" (International Resource Panel, UN Environment Programme, 2018).
4. L. M. A. Bettencourt, *Introduction to Urban Science: Evidence and Theory of Cities as Complex Systems* (MIT Press, 2021).
5. M. Keith *et al.*, *Cities* **105**, 102820 (2020).
6. W. Zhou *et al.*, *npj Urban Sustain.* **1**, 1 (2021).
7. R. D. Simkin, K. C. Seto, R. I. McDonald, W. Jetz, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117297119 (2022).
8. D. Corcos *et al.*, *PLOS ONE* **14**, e0214068 (2019).
9. M. C. Stanley *et al.*, *Front. Ecol. Environ.* **13**, 553 (2015).
10. G20 Research Group, "Extraordinary G20 Leaders' Summit: Statement on COVID-19" (2020); http://www.g20.utoronto.ca/2020/2020-g20-statement-0326.html.
11. UN President of the General Assembly, "Declaration on the Commemoration of the Seventy-fifth Anniversary of the United Nations" (2020); https://www.un.org/pga/74/wp-content/uploads/sites/99/2020/06/letter-on-UN@75-Declaration-silence-17-June-2020.pdf.
12. W. Pearce *et al.*, *Environ. Sci. Policy* **80**, 125 (2018).
13. W. San Martín, *Glob. Environ.* **14**, 423 (2021).

ACKNOWLEDGMENTS

This piece emerged from a workshop sponsored by the UK Research and Innovation (UKRI) PEAK Urban program, grant reference ES/P011055/1. Funding for certain authors' time and input was provided by the UKRI PEAK Urban program, grant reference ES/P011055/1.

10.1126/science.adi6636

PERSPECTIVES

CANCER

Deploying blood-based cancer screening

AI-based risk assessment may enable personalized blood-based multicancer screening

By **Douglas S. Micalizzi[1,2], Lecia V. Sequist[2], Daniel A. Haber[1,2,3]**

The past 20 years have witnessed transformative advances in molecularly targeted and immunological treatments for advanced cancer, providing many patients with prolonged survival and quality of life. However, the main determinant of cure across diverse cancers remains the stage at diagnosis. Finding an invasive cancer while it is still localized and without clinically detectable metastatic spread provides the best chance at eradicating the primary tumor through surgery and/or radiation and killing any disseminated microscopic cells through therapeutic drugs. The recent development of blood-based multicancer detection (MCD) assays, together with advances in imaging and artificial intelligence (AI) algorithms, have the potential to transform early cancer detection. But these innovations are not without health and financial risk, and their increasing availability raises both opportunities and challenges, which are evident as clinics dedicated to early cancer detection are launched.

Take the case of a 55-year-old woman who has recently lost a close relative to cancer and is concerned about her own risk. Her family history does not fit a known cancer genetic susceptibility syndrome. She has a history of tobacco use, exercises routinely, maintains a normal body mass index, and drinks alcohol in moderation. She is up to date on current recommendations for cancer screening, including Pap smear, mammography, and colonoscopy. She decides to pay for a multicancer detection test (Galleri), which is currently available for purchase in the US but without US Food and Drug Administration (FDA) approval or insurance reimbursement. The test indicates a "cancer signal detected" with ovary as a top predicted tissue of origin, yet clinical work-up, including high-resolution imaging and the ovarian cancer antigen 125 (CA-125) blood marker, is negative. How should a patient who appears healthy but has a positive cancer signal on a blood test be counseled, and how common is such a scenario likely to be as MCD screening becomes increasingly available?

[1]Krantz Family Center for Cancer Research, Massachusetts General Hospital Cancer Center, Harvard Medical School, Charlestown, MA, USA. [2]Department of Medicine and Massachusetts General Hospital Cancer Center, Harvard Medical, School, Boston, MA, USA. [3]Howard Hughes Medical Institute, Chevy Chase, MD, USA. Email: dmicalizzi@mgb.org; lvsequist@mgb.org; dhaber@mgh.harvard.edu

Several MCD assays are at various stages of development, with the Galleri test from GRAIL being the most advanced in clinical studies, and in negotiations for approval by US and UK regulators (*1*). Galleri uses 40 ml of blood to extract free DNA in the plasma, a fraction of which may be derived from tumor cells if cancer is present (i.e., circulating tumor DNA, ctDNA). Given the large number of DNA methylation changes at CpG dinucleotides throughout the cancer genome, the test applies bisulfite sequencing to annotate over 100,000 genomic loci, using algorithms to identify a potential cancer signal and a likely tissue of origin, admixed with normal tissue-derived DNA in the blood. Other emerging blood-based cancer assays rely on the altered size distribution of cancer-derived ctDNA (DELFI) (*2*) or the presence of recurrent mutations and abnormal protein markers (CancerSEEK) (*3*). Beyond these and other ctDNA-derived assays, cancer-associated blood analytes include high-throughput proteomics, circulating tumor cells, exosomes, platelet-associated RNA, and circulating free RNA.

The argument for developing a single blood-based test to screen for multiple cancers, rather than a tumor type–specific test, is that shared molecular features of all cancers can be leveraged in this way, providing a "one test for all" clinical paradigm that could be readily implemented across asymptomatic populations. The caveat is that test performance and predictive power depend on the prevalence of the cancer under screening, and different cancers have distinct risk populations, as well as variable patterns in the time to progress from a single cell to an invasive cancer shedding ctDNA into the blood. A major unanswered question is whether the most lethal cancers that currently lack screening tests (e.g., pancreatic and ovarian cancers) exhibit a sufficient window of opportunity between plasma detectability and tumor metastasis to deploy curative surgery.

How effective are MCD screening tests at uncovering early-stage, potentially curable cancers? Initial studies (*1*) compared patients known to have different types of cancer with healthy individuals, reporting an overall sensitivity (correctly identifying a patient with cancer) for Galleri of 16.8% for stage I and 40.4% for stage II cancer, when the assay parameters were set at a threshold of 99.5% specificity (correctly identifying a patient without cancer). In the PATHFINDER trial, a population-based study of 6621 apparently healthy individuals over age 50, 1.4% had a positive cancer signal on Galleri testing; of these, cancer (of any stage) was ultimately confirmed in 38%, whereas 62% appeared to be false-positives. Such false signals may require costly imaging and invasive tests to rule out the presence of cancer and can cause unnecessary anxiety (*4*). Previously unsuspected stage I or stage II cancer was present in 14 of the 36 cases that were correctly identified by Galleri as having cancer, i.e., 0.2% of the initially screened population was discovered to have a potentially curable early-stage cancer. A major population-based trial is ongoing through the National Health Service (NHS) in the UK, involving randomization of 140,000 asymptomatic individuals between ages 50 and 75 to either standard clinical cancer screening protocols plus annual Galleri testing for 3 years, versus clinical screening alone. The primary end point for this study is earlier stage at cancer diagnosis within the MCD-tested cohort, rather than a reduction in overall cancer-related survival. This end point will deliver a more expedient trial readout but lacks the ability to assess for important confounders such as lead-time bias, when cancers are discovered earlier in their course owing to study intervention but not early enough to alter their curability.

Perhaps the most critical question regarding the implementation of MCD screening tests is whether they are best applied to all persons above a certain age, or whether advances in AI will enable more individualized risk-based screening strategies, thereby raising the baseline prevalence and hence predictive value of testing. Cancer risk increases by age, and in Western countries, the annual incidence is estimated to be 0.5% at age 50 and 1.5% at age 65. The positive predictive value (PPV) of a screening test, meaning the chance that a positive test result corre-

sponds to a true cancer case, combines assay-inherent specificity and sensitivity with the cancer prevalence in the tested population. Thus, a hypothetical test with 99% sensitivity at 99% specificity, when applied to a population with only a 1% cancer prevalence, will produce one false-positive result for every true-positive result (i.e., PPV 50%). However, if the cancer prevalence in the population rises to 5%, the PPV for the same test jumps to 84% (i.e., fewer false-positives).

Current-generation cancer risk calculators typically focus on a single cancer type [e.g., Tyrer-Cuzick for breast cancer; the prostate, lung, colorectal, and ovarian cancer screening trial (PLCO); and the colorectal cancer risk assessment tool (CCRAT)], and they use a limited number of static risk factors as input, generating validated risk predictions that can be used to select at-risk patients for classic cancer screening tests [e.g., mammogram, low-dose chest computed tomography (CT), and colonoscopy]. Similarly, there are well-established algorithms for cancer screening in individuals carrying highly penetrant inherited genetic mutations that confer susceptibility to melanoma, breast, ovarian, colon, renal, and endocrine cancers. There are, however, multiple risk modifiers that may only be accessible through complex algorithms. AI-driven approaches to cancer risk assessment may integrate traditional risk factors with new or harder-to-assess factors, including lower-penetrance genetic variants, diverse environmental exposures, and other health indicators. This approach was illustrated in a recent retrospective study using machine learning–based analysis of clinical records to predict risk at specific time intervals for pancreatic cancer (*5*), a tumor for which a validated risk calculator is not currently available. Additionally, radiology images of noncancerous tissue may now be analyzed to help predict an individual's future risk of breast or lung cancer, by using AI-powered techniques that are distinct from traditional clinical radiology assessments of current lesions (*6*, *7*). Thus, the evolution of individualized cancer risk assessment may enable more effective targeting of blood-based MCD screening to populations with an increased cancer prevalence, which would in turn improve PPV.

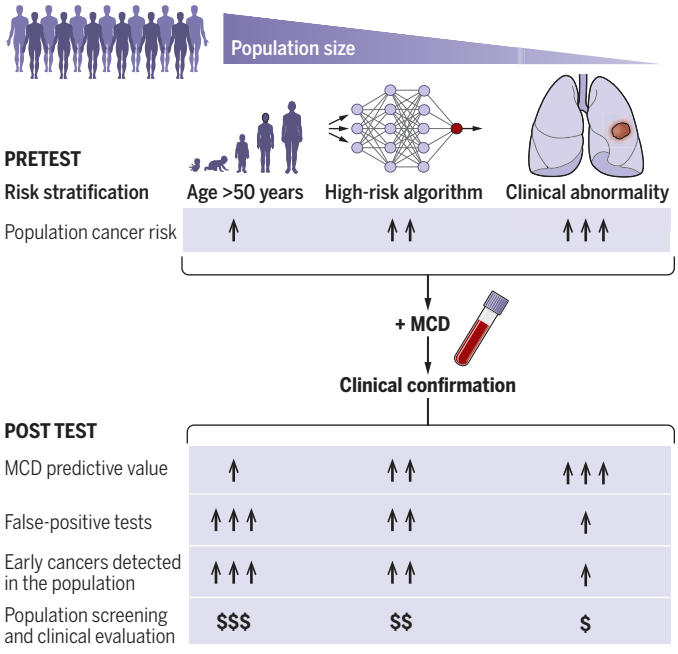
Beyond selection criteria for MCD testing, the clinical evaluation of patients with a blood-based cancer signal is critical to their deployment. For the Galleri test, DNA methylation patterns give an initial clue about the tissue of origin, providing a formula to begin clinical workup, but if this is unrevealing, the subsequent evaluation is unclear. Whole-body imaging [e.g., positron emission tomography (PET) scan and whole-body magnetic resonance imaging (MRI)] is a consideration, but it is fraught with poor sensitivity, incidental findings, and high cost. Notably, in the PATHFINDER study, 44 of the 90 patients with a positive Galleri test underwent an invasive diagnostic procedure to determine the presence or absence of cancer. GRAIL currently offers free repeat Galleri testing in 3 to 6 months if no cancer diagnosis is made after an initial positive test. It is also possible that routine application of orthogonal blood-based validation assays may play a role in reducing the fraction of false-positive results at initial screening. Such second-line assays could include high-sensitivity detection of cancer-associated DNA mutations or circulating tumor cells in the blood (*8*), or molecular probes coupled with high-sensitivity imaging analyses. However, without clinical confirmation, a positive cancer signal from a blood test represents either a false-positive result or a yet-undetectable malignancy that warrants ongoing vigilance, an unresolved dilemma that may be the source of profound anxiety.

There are additional clinical scenarios in which MCD testing may contribute to early cancer detection. Clinical medicine is replete with sophisticated imaging for diverse indications, increasingly yielding radiographic lesions of unknown significance. Examples include indeterminate lung nodules identified in 18% of individuals undergoing chest CT scoring of coronary calcium deposits for cardiac risk assessment (*9*) and incidentally discovered premalignant intraductal papillary mucinous neoplastic cysts in the pancreas of 10% of individuals over age 70 (*10*). Blood-based MCD tests might play a role in the evaluation of such incidental lesions, helping to assess the need for invasive biopsy or surgery. Additionally, MCD testing could be useful in individuals presenting with signs or symptoms that are consistent with but not diagnostic for cancer. Indeed, in such a high-risk population, a UK study of 5461 patients reported a PPV of 75% for the Galleri test among patients suspected of having cancer before a definitive clinical diagnosis (*11*) (see the figure).

As MCD-based cancer screening evolves, the history of cancer screening for prostate and lung cancers offers some distinct lessons about implementation. Age-based screening for prostate cancer using the blood protein marker prostate-specific antigen (PSA) is no longer routinely recommended after overambitious implementation in the 1990s highlighted its low PPV for invasive disease, to the extent that for every life saved by population-based PSA testing, another was lost through a biopsy or surgery-related complication (*12*). PSA testing for men aged 55 to 69 is currently left to the discretion of individual patients and their physicians (*13*). By contrast, in lung cancer, randomized controlled trials clearly demonstrated a 20% reduction in cancer mortality after low-dose chest CT screening among heavy smokers (*14*). Yet fewer than 10% of eligible patients undergo lung screening, owing to the lack of comprehensive implementation strategies, nihilism about lung cancer outcomes, and stigma about smoking (*15*). Furthermore, for both prostate and lung cancer screening, associated risk factors and availability

Multicancer screening tests according to risk

Population-based screening using age as the sole risk factor may have greatest benefit for the total number of early cancers detected in the population, but with a considerable number of false positives given the low disease prevalence, and at high cost. Risk stratification, potentially using AI-based risk calculators, may increase population prevalence, thereby improving positive predictive value (PPV) of the test. Applying multicancer detection (MCD) testing for evaluation of radiographic lesions of uncertain significance may be another relevant clinical application with high PPV.



GRAPHIC: K. HOLOSKI/*SCIENCE*

of sophisticated diagnostics are unequal across diverse communities in the US. Compared with white people, Black people suffer higher rates and worse outcomes for both prostate and lung cancers and, despite efforts to improve access, remain less likely to qualify for lung cancer screening (*15*). Just, equitable, and affordable deployment of cancer screening is a major concern that should be actively addressed in MCD test deployment. In this regard, cost-effectiveness analysis of MCD testing should be evaluated at all stages of implementation, including the downstream costs of clinical confirmation and their combination with standard screening approaches.

Most importantly, individual perception of personal risk for cancer is often difficult to quantify, but it underlies many patient preferences and decisions. MCD screening is not dissimilar from existing cancer screening tests in having imperfect sensitivity and a high false-positive rate. It differs perhaps in the public perception that molecular tests have a diagnostic level of certainty, whereas radiographic abnormalities tend to be understood as being preliminary until confirmed by definitive biopsy. Furthermore, organ-based cancer screening is more amenable to clinical confirmation than a multicancer signal in the blood, whose origin may elude immediate validation.

The role of MCD screening as a new tool within the spectrum of clinical care thus presents both an unprecedented opportunity and a major challenge. Coupled with such potent cancer detection technologies, the enhanced ability to objectively assess personalized cancer risk is probably the most important element in a rational cancer screening strategy, maximizing predictive power while minimizing unnecessary anxiety and medical workups. ■

REFERENCES AND NOTES

1. E. A. Klein *et al.*, *Ann. Oncol.* **32**, 1167 (2021).
2. S. Cristiano *et al.*, *Nature* **570**, 385 (2019).
3. A. M. Lennon *et al.*, *Science* **369**, eabb9601 (2020).
4. D. M. Schrag *et al.*, *Lancet* **402**, 1251 (2023).
5. D. Placido *et al.*, *Nat. Med.* **29**, 1113 (2023).
6. P. G. Mikhael *et al.*, *J. Clin. Oncol.* **41**, 2191 (2023).
7. A. Yala *et al.*, *J. Clin. Oncol.* **40**, 1732 (2022).
8. H. Guo *et al.*, *Cell* **186**, 2765 (2023).
9. C. Iribarren *et al.*, *Am. J. Med.* **121**, 989 (2008).
10. J. J. Farrell, *Gut Liver* **9**, 571 (2015).
11. B. D. Nicholson *et al.*, *Lancet Oncol.* **24**, 733 (2023).
12. J. J. Fenton *et al.*, *JAMA* **319**, 1914 (2018).
13. S. P. Basourakos *et al.*, *NEJM Evid.* **1**, 6 (2022).
14. National Lung Screening Trial Research Team *et al.*, *N. Engl. J. Med.* **365**, 395 (2011).
15. US Preventive Services Task Force *et al.*, *JAMA* **325**, 962 (2021).

ACKNOWLEDGMENTS

D.S.M. and L.V.S. are involved in a clinical demonstration project funded by GRAIL. L.V.S. receives institutional research funding from AstraZeneca, Novartis, and Delfi Diagnostics. D.A.H. is a cofounder of a circulating tumor cell (CTC) diagnostics company (TellBio) and holds patents relating to CTC diagnostics.

10.1126/science.adk1213

An ant–plant relationship is vital to a food web that includes the predation of zebra by lions in a Kenyan savanna.

ECOLOGY

A big-headed problem drives an ecological chain reaction

Disruption of key species interactions reverberates across an African savanna

By **Kaitlyn M. Gaynor**

Human activity is driving the rapid loss of global biodiversity, through declines in individual species and the wholesale destruction of ecosystems (*1*). This loss can arise from myriad forms of anthropogenic disturbance that include land conversion, hunting, pollution, resource extraction, and climate change (*2*). Although it is often straightforward to document the direct effects of disturbance on species and habitats, these impacts can ripple throughout food webs by altering interactions among species. These indirect effects may have far-reaching consequences that are not immediately apparent, but could fundamentally alter ecosystems. On page 433 of this issue, Kamaru *et al.* (*3*) describe how one disturbance—the introduction of an invasive species—disrupted an interaction between trees and ants, and traced its consequences through an African savanna landscape.

Species interactions are essential to the functioning of healthy ecosystems. Regardless of whether they benefit both species (mutualism), one species (predation), or neither (competition), species interactions can stabilize the composition of communities and the state of an ecosystem. Some interactions play a particularly outsized role in maintaining ecological dynamics by shaping the physical environment, cycling nutrients or energy, or limiting the populations of other species. These interactions may involve numerically abundant species (foundational interactions) or rare but important species (keystone interactions) (*4*). Given their central role, the disruption of such interactions by human disturbance can have reverberating and transformative ecological effects.

Humans have been characterized as a higher-order hyperkeystone species, given that human activities can radically alter interaction chains (*5*). However, it is often difficult to disentangle the pathways linking the fate of one species to another as disturbance cascades throughout complex ecosystems, even if these pathways involve foundational or keystone interactions. When an ecosystem is confronted with multiple anthropogenic pressures that have differential effects across species, it can be nearly impossible to attribute an observed system-wide change to a particular link in the chain. Studies often

Departments of Zoology and Botany, The University of British Columbia (Musqueam) Territory, Vancouver, BC, V6T 1Z4, Canada. Email: kaitlyn.gaynor@ubc.ca

rely on natural experiments involving the loss or reintroduction of a single species, but in many cases, the attribution of causality remains elusive. The reintroduction of wolves (*Canis lupis*) in the Greater Yellowstone Ecosystem has been associated with pronounced changes in tree communities and stream hydrology, but even this seemingly straightforward story has been complicated by subsequent studies that presented alternative explanations for observed changes (*6*). Given the challenges of studying complex and dynamic natural ecosystems, scientists still have a limited understanding of the extent to which humans have modified interaction networks.

Weaving together observations from a natural experiment and a controlled herbivore exclosure experiment, Kamaru *et al.* meticulously pieced together the causes and consequences of the disruption of a foundational interaction between acacia ants (*Crematogaster* spp.) and whistling-thorn trees (*Vachellia drepanolobium*) in central Kenya. Whistling-thorn trees provide food and shelter to these ants. In turn, the ants protect the trees from browsing elephants (*Loxodonta africana*) with their irritating bite. Thus, the ant–tree mutualism has long played a critical role in maintaining tree cover in the savanna landscape. That is, until this foundational interaction was disrupted by the big-headed ant (*Pheidole megacephala*), a pernicious invasive species. Although the geographic origin of big-headed ants is unknown, the movement of people and goods has enabled their global spread, with catastrophic consequences for native insects. In central Kenya, the big-headed ants have been overtaking the acacia ants over the past two decades and erasing their mutualism with the whistling-thorn trees.

Kamaru *et al.* documented each step of a chain reaction that started with big-headed ants and made its way to lions (*Panthera leo*). By comparing areas that had been invaded by big-headed ants to those beyond the invasion frontier, the authors found that by disrupting the ant–tree mutualism, invasion rendered trees more vulnerable to elephant damage. A subsequent reduction in tree cover then led to greater visibility for large mammals, with implications for predator–prey interactions. Zebra (*Equus quagga*) are on the lookout for lions, an ambush predator that prefers to maintain the element of surprise. The increased visibility in invaded sites proved beneficial to zebras, decreasing the occurrence of zebra kills by almost threefold compared with uninvaded sites. It remains unclear if and how lions will compensate for this decrease in zebra "catchability." Lions may experience population decline, concentrate hunting in low-visibility areas, or switch to alternative prey. There is some evidence for the latter, as zebra have recently made up a smaller proportion of lion diets.

The study of Kamaru *et al.* highlights the importance of looking beyond top-down effects of apex consumer loss and the bottom-up effects of habitat loss when evaluating the ecological impacts of disturbance. Much emphasis has been placed on trophic cascades following the loss of charismatic predators, in which changes in the density or behavior of prey alter food webs (*7*). However, many important species interactions appear elsewhere in an interaction network, and the foundational and keystone roles of smaller species at lower trophic levels should not be underestimated (*8*). Furthermore, the ecological importance of species interactions may not be mediated by trophic links among species that consume one another, but rather by nontrophic interactions. In the Kenyan savanna, the disruption of an ant–plant mutualism modified a predator–prey interaction by restructuring the physical habitat. In a sense, the elimination of an ant–plant interaction released another keystone species, the elephant, which then reengineered the environment.

Kamaru *et al.* harnessed central themes from community ecology—insect–plant mutualism, predator–prey interactions, foundational and keystone species—to piece together the puzzle of how an invasive species ultimately reshaped an ecosystem. In an era characterized by rapid environmental change, such applied ecological research is critical to understand how disturbance alters ecosystem structure and function. Although the disruption of foundational or keystone interactions can amplify the effects of disturbance on ecosystems, the maintenance of such interactions may also buffer against the effects of disturbance by increasing ecological resilience and resistance (*9*). Ultimately, the conservation of healthy ecosystems requires not only the prevention of species extinction but also the identification and preservation of the most important interactions among species. ■

REFERENCES AND NOTES

1. G. Ceballos *et al.*, *Sci. Adv.* **1**, e1400253 (2015).
2. D. Tilman *et al.*, *Nature* **546**, 73 (2017).
3. D. N. Kamaru *et al.*, *Science* **383**, 433 (2024).
4. A. M. Ellison, *iScience* **13**, 254 (2019).
5. B. Worm, R. T. Paine, *Trends Ecol. Evol.* **31**, 600 (2016).
6. B. L. Allen *et al.*, *Food Webs* **12**, 64 (2017).
7. J. A. Estes *et al.*, *Science* **333**, 301 (2011).
8. I. Shukla, K. M. Gaynor, B. Worm, C. T. Darimont, *Ecol. Evol.* **13**, e10561 (2023).
9. A. Vitali, D. P. Vázquez, M. F. Miguel, Y. Sasal, M. A. Rodríguez-Cabal, *J. Anim. Ecol.* **91**, 74 (2022).

10.1126/science.adn3484

ASTROPHYSICS

A very energetic Galactic particle accelerator

The most powerful plasma jets in the Milky Way emit very-high-energy gamma rays

By **Valentí Bosch-Ramon**

Since the 1912 discovery of cosmic rays (*1*), the origin of these extremely energetic particles has remained a mystery. Remnants of a star explosion (supernova) have been considered dominant sources of cosmic rays (*2*), at least for those originating in the Milky Way, but this may not be true at energies approaching peta–electron volts (1 PeV = 10^{15} eV) (*3*, *4*). Microquasars may also generate cosmic rays (*5*, *6*), but evidence has been scarce. These systems consist of a star and either a black hole or neutron star. Ionized matter (plasma) is emitted as jets flowing in opposite directions from the microquasar. On page 402 of this issue, the High Energy Stereoscopic System (H.E.S.S.) Collaboration (*7*) reports the detection of extremely energetic gamma rays produced in the large-scale jets of SS 433, the most powerful microquasar in the Milky Way. Thus, microquasars may indeed contribute to the most energetic Galactic cosmic rays.

SS 433 was the first Galactic object presenting mildly relativistic jets (particles traveling at a fraction of the speed of light) (*8*) and can be considered the first example of a microquasar (*9*). The kinetic luminosity (energy per second) of these jets is enormous (at least 10^{39} erg/s) and is mostly carried by atomic nuclei (protons and heavier) (*10*). The system sits at the center of a large supernova remnant called W50, which is shaped as an elongated nebula filled with hot plasma from the powerful jets. The interaction of microquasar jets and their environment can cause the acceleration of particles that produce

Departament de Física Quàntica i Astrofísica, Institut de Ciències del Cosmos Universitat de Barcelona (ICCUB), Universitat de Barcelona (IEEC-UB), E08028 Barcelona, Catalonia, Spain. Email: vbosch@fqa.ub.edu

gamma rays, photons that are more energetic than x-rays (*5*, *6*, *11*). The large-scale jets of SS 433 are the best indication that such outflows produce particles that emit photons with energies well above the tera–electron volt (1 TeV = 10^{12} eV) (*7*, *12*). Highly energetic electrons originating from efficient particle acceleration in these jets had been inferred earlier from the detection of x-rays (*13*).

The H.E.S.S. Collaboration's discovery marks the precise locations where gamma ray–emitting particles are accelerated—locations that overlap with the source of x-rays. This suggests a common origin. Both the x-rays and the gamma rays from the large-scale jets of SS 433 begin to arise at around 30 pc from the central source, at sites where the jets acquire further collimation (or narrowing alignment). This additional collimation (or recollimation) is likely a result of the complex interaction between the jets and their surrounding medium. The gamma rays come from slightly different regions within the jets, depending on their energy. More energetic gamma rays originate closer to the SS 433 binary system. This is best explained by considering relativistic electrons as the particles emitting the tera–electron volt gamma rays through the scattering of infrared photons (the infrared photons turn into gamma rays). These electrons meanwhile propagate away from the jet recollimation sites. They lose energy mostly by interacting with local magnetic fields, a process that generates the observed x-rays. The relativistic electrons are embedded in a flow that moves at a velocity substantially lower than the initial jet velocity. This change in velocity is consistent with the presence of a sudden drop in the velocity of the jet flow, known as a shock, at the point where the jets become recollimated—30 pc from the binary system. This suggests the presence of two shocks in which the relativistic electrons are accelerated. The acceleration mechanism would be similar to that in a supernova remnant, although the shocks in SS 433 jets are faster than supernova remnant shocks and can accelerate particles to higher energies.

In addition to electrons, atomic nuclei can also be accelerated in the large-scale jets of SS 433. Because relativistic nuclei are affected by weaker energy losses than those by electrons, the former could reach much higher energies when accelerated. In addition, these nuclei can carry a much higher total energy than electrons can, as expected for the sources of Galactic cosmic rays (*4*). The nuclei would not be detectable in the jets because gamma rays produced by these particles through collisions with other nuclei would be faint because of the jets' low densities. Nevertheless, the relativistic nuclei can eventually reach regions outside the jets and radiate more efficiently if denser material is present. Giga–electron volt (10^9 eV) gamma rays have been detected outside (near) the jets, apparently in a region of high ambient density, so they could be tracing the presence of relativistic nuclei. Properly estimating the total energy in these particles requires a detailed model, but these giga–electron volt observations suggest that a substantial fraction of the jet kinetic luminosity may be in the form of accelerated nuclei (*14*). All of this indicates that SS 433 is a good candidate to produce not only electrons with hundreds of tera–electron volts (*7*) but also large quantities of nuclei with peta–electron volt energies, and higher.

As pointed out by the H.E.S.S. Collaboration, SS 433 cannot be the source of the very energetic (peta–electron volt) cosmic-ray protons detected on Earth because the source is too young for its particles to reach Earth once they have escaped the source. However, closer and/or longer-lived microquasars, even if weaker (and individually harder to detect), could be contributing non-negligibly to local peta–electron volt cosmic rays. Presently, most known sources of very energetic photons seem to be leptonic (e^-/e^+) in nature (*15*), and the origin of peta–electron volt cosmic-ray nuclei is still an open question. However, the very energetic photons detected from the large-scale jets of SS 433 are an indirect indicator that these kinds of objects should not be neglected when seeking to explain the most energetic nuclei in Galactic cosmic rays. ■

REFERENCES AND NOTES

1. V. F. Hess, *Phys. Z.* **13**, 1084 (1912).
2. W. Baade, F. Zwicky, *Proc. Natl. Acad. Sci. U.S.A.* **20**, 259 (1934).
3. F. Aharonian, R. Yang, E. de Oña Wilhelmi, *Nat. Astron.* **3**, 561 (2019).
4. S. Gabici *et al.*, *Int. J. Mod. Phys. D* **28**, 1930022 (2019).
5. S. Heinz, R. Sunyaev, *Astron. Astrophys.* **390**, 751 (2002).
6. V. Bosch-Ramon, F. A. Aharonian, J. M. Paredes, *Astron. Astrophys.* **432**, 609 (2005).
7. H.E.S.S. Collaboration, *Science* **383**, 402 (2024).
8. A. Fabian, M. J. Rees, *Mon. Not. R. Astron. Soc.* **187**, 13 (1979).
9. I. F. Mirabel, L. F. Rodríguez, *Annu. Rev. Astron. Astrophys.* **37**, 409 (1999).
10. H. L. Marshall, C. R. Canizares, N. S. Schulz, *Astrophys. J.* **564**, 941 (2002).
11. P. Bordas *et al.*, *Astron. Astrophys.* **497**, 325 (2009).
12. A. U. Abeysekara *et al.*, *Nature* **562**, 82 (2018).
13. S. Yamauchi, N. Kawai, T. Aoki, *Pub. Astron. Assoc. Japan* **46**, L109 (1994).
14. J. Li *et al.*, *Nat. Astron.* **4**, 1117 (2020).
15. Z. Cao *et al.*, arXiv:2305.17030 [astro-ph.HE] (2023).

ACKNOWLEDGMENTS

V.B.-R. received financial support from the Spanish Ministry of Science and Innovation through the grants PID2022-136828NB-C41/AEI/10.13039/501100011033/ and European Regional Development Fund, European Union, CEX2019-000918-M (ICCUB María de Maeztu 2020–2023), and from the Generalitat de Catalunya through the grant 2021SGR00679.

10.1126/science.adn3487

SYNTHETIC BIOLOGY

Accelerated evolution of chosen genes

Orthogonal replication enables rapid continuous biomolecular evolution in *Escherichia coli*

By **Rory L. Williams** *and* **Chang C. Liu**

Directed evolution is a powerful strategy for engineering biomolecules. However, classical approaches to directed evolution usually rely on a repeated sequence of labor-intensive steps where genes encoding biomolecules of interest are diversified in vitro, then transformed into cells, expressed, and subjected to selection or screening to achieve desired activities. As a result, the evolutionary searches that these manually staged processes constitute are limited in scale and depth (*1*). Thus, there has been a growing effort to develop synthetic genetic systems that autonomously diversify user-defined genes in vivo so that biomolecules of interest quickly and autonomously evolve as cells are grown under selection for the gene's function (*2*). On page 421 of this issue, Tian *et al.* (*3*) report the establishment of an orthogonal DNA replication system in *Escherichia coli* (EcORep) that will allow in vivo rapid continuous directed evolution of RNAs, enzymes, proteins, and genetic circuits toward new activities.

The first orthogonal DNA replication system, OrthoRep, was reported in 2014 as a durable architecture for in vivo continuous evolution of user-defined genes (*4*, *5*). In OrthoRep, a linear DNA plasmid is replicated by a specific (orthogonal) DNA polymerase (DNAP) through protein-primed replication (*6*). Engineered error-prone variants of the DNAP selectively replicate and mutate the genes encoded on the linear plasmid while sparing the host genome, which cannot withstand high mutation rates. When cells with OrthoRep are grown under selection for desired functions, the encoded genes evolve rapidly, at scale, and at depth (*1*). OrthoRep was developed in the yeast *Saccharomyces cerevisiae*, and a second orthogonal DNA replication system was more recently established in the bacterium *Bacillus thuringien-*

sis (*7*). However, the most-used workhorse microbe in molecular and synthetic biology is neither *S. cerevisiae* nor *B. thuringiensis* but *E. coli*. Tian *et al.* now bring the orthogonal DNA replication strategy to *E. coli*, thereby broadening its scope.

To establish EcORep, Tian *et al.* domesticated replication machinery from the lytic bacteriophage PRD1, which is distinct among *E. coli* phages in its use of protein-primed replication of a linear phage genome (*6*, *8*). Previous work on the related Phi29 bacteriophage protein-primed replication system showed that four components—the terminal protein, DNAP, and two DNA binding proteins—were sufficient to reconstitute replication of a Phi29-based linear plasmid in vitro (*9*). Tian *et al.* hypothesized that a homologous set of four genes from the PRD1 phage would constitute a minimal replication system that could copy a PRD1-based linear plasmid in *E. coli*. Indeed, they demonstrated that when *E. coli* expressing the four genes were electroporated with a synthetic linear DNA plasmid encoding an antibiotic resistance marker flanked by the native PRD1 inverted terminal repeats (ITRs), the linear plasmid was successfully replicated (see the figure). Although the replication establishment efficiency was initially low, modifications involving the use of helper plasmids encoding the lambda phage Gam protein and/or an extra dose of the two PRD1 DNA binding proteins resulted in reliable replication and sustained plasmid maintenance over 100 generations of culturing under antibiotic selection.

Having established a functional EcORep system, Tian *et al.* engineered DNAP variants with elevated error rates to drive the introduction of mutations and thereby the evolution of genes encoded on the linear plasmid. They identified several error-prone DNAP variants with mutation rates ranging from 2.6×10^{-8} to 8.7×10^{-6} substitutions per base pair (s.p.b.). The most highly error-prone DNAP will require further engineering to support stable replication, but two DNAPs with intermediate mutation rates around 2.0×10^{-7} s.p.b. did support stable replication. Critically, in measuring these elevated error rates, Tian *et al.* demonstrated that the genomic mutation rate was unaffected, which validates orthogonality. Tian *et al.* conducted multiple directed evolution campaigns on the tetracycline efflux ABC transporter (*TetA*) for resistance to the tetracycline antibiotic tigecycline as well as on a variant of superfolder green fluorescent protein (*sfGFP*) for greater fluorescence, achieving substantial increases in cellular resistance to tigecycline and fluorescence, respectively.

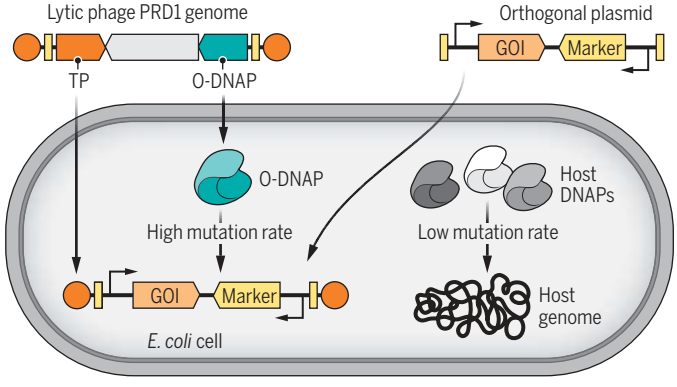
A next step for EcORep will be to increase its error rate by 2 to 3 orders of magnitude to reach those that maximize the rate of gene evolution, especially when evolution is not guided by strong positive selection that is capable of sequentially fixing beneficial mutations even when mutation rates are relatively low. At the current mutation rate of ~2.0×10^{-7} s.p.b., only ~0.2% of an arbitrary 1-kb gene will incur a single new mutation after ~10 generations, which could be increased. Further DNAP engineering, following similar strategies to those that brought OrthoRep's mutation rate to 10^{-5} s.p.b. (*5*), and, more recently, a preliminary report of 10^{-4} s.p.b. (*10*), should prove invaluable.

A practical benefit of EcORep is that it is easy to set up—encoding user-defined genes onto EcORep requires the straightforward transformation of a naked linear DNA construct containing short ITRs. By contrast, OrthoRep relies on recombination of new genes into yeast strains already containing an OrthoRep landing pad. Another benefit is the general accessibility of molecular biology and genetic techniques to manipulate *E. coli*, which should facilitate wide adoption. Additionally, the shorter generation time of *E. coli* (20 to 30 min) and the higher cell density (10^9 to 10^{10} cells/ml) compared with those of yeast (1.5 to 2 hours; 10^7 to 10^8 cells/ml) should afford faster evolution. Furthermore, the orthogonal replicon copy number control demonstrated by Tian *et al.* should prove useful in manipulating selection stringency and purifying selection in evolution experiments.

Another important characteristic of EcORep is that expression of genes from EcORep uses both the native transcriptional and translational machinery of the host cell, in contrast to the expression of genes from OrthoRep, which uses an orthogonal transcription system (*4*). This means that EcORep can enable the rapid evolution of native regulatory sequences and genetic circuit behavior both to study endogenous *E. coli* regulation and to evolve synthetic counterparts that are compatible with normal *E. coli* strains that do not contain EcORep. In addition, there are promising areas of synthetic biology, such as orthogonal tethered ribosomes (*11*) and recoded genomes (*12*–*14*), that are most advanced in *E. coli* among model organisms. EcORep is well-positioned to evolve specialized ribosomes or to test how proteins evolve under synthetic genetic codes that have noncanonical correspondences between codons and amino acids. As well, EcORep's replication machinery was assembled from the bottom up, which gives exceptional control over its components and affords opportunities that include transfer to other strains and host organisms or possibly the creation of self-replicating protein-DNA conjugates where user-selected proteins are fused to the terminal protein of EcORep. ■

An orthogonal replication system in *Escherichia coli*

The genes encoding the terminal protein (TP), an orthogonal DNA polymerase (O-DNAP), and two DNA binding proteins are taken from the PRD1 phage genome and encoded in *Escherichia coli*. The expression of these genes allows protein-primed mutagenic DNA replication (after the O-DNAP is engineered to be error prone) of an orthogonal plasmid containing the gene of interest (GOI). The resulting system, called EcORep, enables the specific and rapid continuous evolution of GOIs on the plasmid while sparing the host genome.



GRAPHIC: K. HOLOSKI/*SCIENCE*

Department of Biomedical Engineering and Center for Synthetic Biology, University of California, Irvine, CA, USA. Email: ccl@uci.edu

REFERENCES AND NOTES

1. G. Rix, C. C. Liu, *Curr. Opin. Chem. Biol.* **64**, 20 (2021).
2. R. S. Molina *et al.*, *Nat. Rev. Methods Primers* **2**, 36 (2022).
3. R. Tian *et al.*, *Science* **383**, 421 (2024).
4. A. Ravikumar, A. Arrieta, C. C. Liu, *Nat. Chem. Biol.* **10**, 175 (2014).
5. A. Ravikumar, G. A. Arzumanyan, M. K. A. Obadi, A. A. Javanpour, C. C. Liu, *Cell* **175**, 1946 (2018).
6. M. Salas, *Annu. Rev. Biochem.* **60**, 39 (1991).
7. R. Tian *et al.*, *Nat. Chem. Biol.* **19**, 1504 (2023).
8. H. Savilahti, D. H. Bamford, *J. Virol.* **67**, 4696 (1993).
9. M. Mencía, P. Gella, A. Camacho, M. de Vega, M. Salas, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18655 (2011).
10. G. Rix *et al.*, bioRxiv 10.1101/2023.11.13.566922 (2023).
11. C. Orelle *et al.*, *Nature* **524**, 119 (2015).
12. J. Fredens *et al.*, *Nature* **569**, 514 (2019).
13. N. Ostrov *et al.*, *Science* **353**, 819 (2016).
14. F. J. Isaacs *et al.*, *Science* **333**, 348 (2011).

ACKNOWLEDGMENTS

C.C.L. is a cofounder of K2 Biotechnologies, Inc., which uses OrthoRep for protein engineering.

10.1126/science.adn3434

BOOKS *et al.*

Tiehm's buckwheat, an endangered species, grow atop a significant lithium deposit in Nevada.

NATURAL RESOURCES

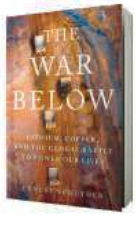
Material conflicts

A journalist probes tensions surrounding two minerals that are key to green technologies

By **Saleem H. Ali**

The mineral anatomy of technology has become a fascination for scholars and journalists alike in recent years. Popular writings on the topic have also gained traction because of a rise in "resource nationalism" surrounding mining practices for metals critical for both defense purposes and green technologies. Adding to this canon, journalist Ernest Scheyder's *The War Below* presents a fine-grained account of the environmental and social conflicts that permeate the landscape where two key minerals for the green energy transition—copper and lithium—are found. [For those interested, journalist Henry Sanderson's recent book *Volt Rush* includes two additional minerals—cobalt and nickel—in its coverage (*1*).]

Scheyder's choice of copper and lithium for his deep-dive analysis is partially determined by the field ethnography that he aims to provide of his travels to mining projects within the United States. These materials are at the forefront of critical mineral conflicts in the US. Although he also includes coverage of international projects, such as the Uyuni lithium fields of Bolivia, Scheyder's storyline most acutely reveals the fault lines and contradictions of American critical minerals policy.



The War Below
Ernest Scheyder
One Signal Publishers/ Atria, 2024. 384 pp.

The book begins in the 1980s, with the discovery of a distinctive plant species in the Nevada wilderness—*Eriogonum tiehmii*, commonly known as Tiehm's buckwheat. He interviews the discoverer of the species, botanist Jerry Tiehm, for the prologue of the book to understand the salience of such emblems of biodiversity. Four decades later, the species' habitat is the battleground for the development of one of the United States' most lucrative lithium deposits, and Tiehm's buckwheat has become a saber for environmentalists and Native Americans in the fight against mining development.

In many ways, the past 40 years have been the most consequential period in the development of complex mineral supply chains for the transition to green energy technologies. We have come to understand the urgency of climate change during this period and to appreciate the need to find new means of energizing human civilization. Yet, as biologist Barry Commoner warned in one of his "laws of ecology," "there is no such thing as a free lunch" in the Universe (*2*). *The War Below* successfully depicts why this aphorism is so apt for thinking about mineral resources. Even with recycling and circular economies, we must have enough stocks of metals to recycle. With lithium, which is flammable, there is also the challenge of transporting concentrated used material over long distances.

Scheyder offers readers an inside story of the travails that entrepreneurs face in mining, recycling, and material invention. The power wielded by individual tech tycoons and by public-sector officials is portrayed with attention to detail and a plethora of citations and interviews. In many cases, Scheyder personally visits with these key players. His interview questions are insightful, and he allows his subjects to tell their stories in their own words, with minimal editorializing.

This book is not meant to explicate the science behind these innovations in any detail. Indeed, Scheyder glides over some important details along the way. The Nobel Prize–winning history of the lithium-ion battery is mentioned, for example, with a greater emphasis on its curious connection to Exxon—where one of its inventors, Stanley Whittingham, worked many years ago—than on the contributions of the other two scientists who shared the prize, John Goodenough and Akira Yoshino, even though their work is what made the battery commercially viable.

In the book's opening epigraph, Scheyder quotes Eleanor Roosevelt, writing: "It takes as much energy to wish as it does to plan." This is an important reminder for policy-makers and scientists alike. Although Scheyder does not propose a plan for a green transition, the stories he shares are stark reminders of why we need to have a systems view of material supply, from mines to markets. ■

The reviewer is at the Department of Geography and Spatial Sciences, University of Delaware, Newark, DE 19716, USA, and the author of *Soil to Foil: Aluminum and the Quest for Industrial Sustainability* (Columbia Univ. Press, 2023). Email: saleem@udel.edu

REFERENCES AND NOTES

1. H. Sanderson, *Volt Rush: The Winners and Losers in the Race to Go Green* (Oneworld Publications, 2022).
2. B. Commoner, *The Closing Circle: Nature, Man, and Technology* (Random House, ed. 1, 1971).

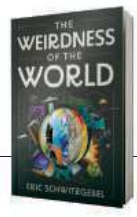
10.1126/science.adk9387

PHILOSOPHY

Entertaining audacious ideas

Unbound by empiricism, a philosopher's provocative musings inspire delight and vexation

The Weirdness of the World
Eric Schwitzgebel
Princeton University Press, 2024. 376 pp.



By **Edouard Machery**

There are two kinds of philosophers: swallows and moles. Swallows love to soar and to entertain philosophical hypotheses at best loosely connected with empirical knowledge. Plato and Gottfried Leibniz are paradigmatic swallows. Moles, on the contrary, rummage through mundane facts about our world and aim at better understanding it. Aristotle, William James, and Hans Reichenbach are paradigmatic moles.

Eric Schwitzgebel is unabashedly a swallow. In his delightful and beautifully written new book, *The Weirdness of the World*, he attempts to convince the reader of a number of provocative ideas. These include the notion that the United States of America might be conscious; that objects might not be in space and generally might be very different from what we take them to be; and that our present actions influence events in future worlds. The book is composed of 12 short chapters, each easily read. It sometimes covers material that will be familiar from introductions to philosophy (the mind-body problem and skepticism) or from recent popular philosophy (the simulation hypothesis or panpsychism), but even then, the discussion always takes a clever and original turn.

I, however, am a mole. The goal of my last book, *Philosophy Within Its Proper Bounds*, was to curtail the flights of fancy with which contemporary philosophers are enamored. Unsurprisingly then, I regularly balked at Schwitzgebel's arguments. (In all fairness, balking at other philosophers' arguments is part of the job description.)

In chapter 4, for instance, Schwitzgebel aims at convincing the reader to assign some small degree of credence—between 0.1 and 1%—to the proposition that some skeptical scenario is in fact actual. These scenarios include the idea that the reader is really a character in a simulation; that she is a brain haphazardly created out of cosmic dust (a Boltzmann brain); or that some even weirder scenario is actual (maybe all our experiences are just the feverish dreams of a madman). But why assign a probability to the possibility of a skeptical scenario? Search me! Schwitzgebel merely says that this assignment is "reasonable."

In addition, Schwitzgebel's arguments often rely on a commonsensical, but questionable, understanding of key ideas. In chapter 7—perhaps my favorite one—he argues that a few plausible physical assumptions, such as the infinity of the Universe, entail that an individual's actions will influence what will happen to her duplicate in a duplicate world emerging randomly at some point in the infinite future of the Universe. Schwitzgebel's argument appeals to probabilities: All possible events must happen in an infinite sequence of chance events. But stop and ask: Can we reasonably assume a probability measure defined over the events in an infinitely inflating cosmos? Probably not, as articulated by "the measure problem" in cosmology and philosophy of physics.

When Schwitzgebel examines whether the United States of America (chapter 3) or a snail (chapter 10) is conscious, he says little about the neurobiology of consciousness or efforts to identify markers of consciousness in a comparative manner. And he dismisses such efforts in chapter 10, because scholars with different assumptions about consciousness would reject these markers. But stop and ask: While his point is correct, is it a deep problem? As work in the history and philosophy of thermometry has shown, this kind of problem routinely gets solved with empirical developments of measuring tools and theoretical progress in understanding what is measured.

Schwitzgebel prides himself on painting awe-inducing philosophical possibilities that go against common sense and that we do not have compelling reasons to believe; they are, in his terminology, "wild." But he never asks why we should care about what common sense finds plausible or bizarre. Common sense is a poor guide to truth, and it should play little role in serious philosophizing.

Schwitzgebel ponders whether our actions could affect other versions of ourselves in future worlds.

The book concludes with a soaring defense of maintaining a childlike attitude toward philosophical and scientific mysteries. I find this defense extremely appealing, but I ultimately prefer Darwin's attitude, celebrating in the last paragraph of *On the Origin of Species* the "grandeur in this view of life" that comes from understanding the world as it is.

Should you read *The Weirdness of the World*? If you have read and enjoyed the work of Nick Bostrom or Philip Goff, then this book is definitely for you. It is brilliant, thought-provoking, and very enjoyable. But card-carrying moles who want to see philosophy add its stone to the ever-growing edifice of knowledge might be left dissatisfied. ■

10.1126/science.adn0629

The reviewer is at the Center for Philosophy of Science, University of Pittsburgh, Pittsburgh, PA 15260, USA. Email: machery@pitt.edu

ILLUSTRATION: GETTY IMAGES

LETTERS

Restoring degraded ecosystems effectively will require seeds from a wide variety of species.

Edited by **Jennifer Sills**

Aim for heterogeneous biodiversity restoration

Commitments to restore about 1 billion hectares by 2030 have emerged in the past decade (*1*), providing hope for tackling the global environmental crisis (*2*). However, restoration initiatives often use limited sets of species, with little regard for the regional diversity found in reference landscapes (*3*–*5*). In diverse tropical ecosystems, where restoring biodiversity is challenging [e.g., (*6*, *7*)], such practices can lead to homogeneous biological communities and habitats that do not fulfill the purpose of restoration.

Restoration science is still evolving (*2*), especially in the tropics (*8*). The Arc of Restoration program in the Amazon, launched during the 2023 United Nations Climate Change Conference (COP28), promises to restore 6,000,000 hectares of land by 2030, about 0.9% of the total Amazon area (*9*). The Atlantic Forest Treaty, launched in October 2023, pledges to restore 54,000 hectares by 2026, about 0.05% of Atlantic forests (*10*). These large restoration programs require complex supply chains for seed and plant material (*11*) and solid scientific knowledge about the species that comprise each ecosystem (*7*), both still limited in the tropics (*7*, *8*, *11*). To succeed, large-scale restoration requires the development of national and regional policies that promote the supply of considerable species sets (*11*), accounting for local and regional diversity.

Ideally, the heterogeneity found in reference ecosystems should guide goal setting and species selection (*7*, *12*). Therefore, protecting natural remnants that serve as propagule sources and references for restoration is essential (*7*). Remnant ecosystems also contribute to ecological connectivity, enabling and accelerating natural colonization processes.

Restoration is expected to yield outcomes that mitigate and facilitate adaptation to climate change. Despite vast efforts and investment, if restoration practices do not recreate the diversity found in reference ecosystems, restoration will not achieve those goals. Moreover, if remnant ecosystems are lost and can no longer serve as references, restoration efforts will be compromised. Conservation of remn ecosystems must be prioritized, and resto- ration projects must aim for fully restored ecosystems, which will benefit the social, ecological, and economic sectors.

Tiago S. P. Toma[1,2]*, Hernani F. M. Oliveira[1,3], Gerhard E. Overbeck[1,4], Carlos E. V. Grelle[1,5], Fabio O. Roque[1,6], Daniel Negreiros[1,2], Domingos J. Rodrigues[1,7], Aretha F. Guimaraes[1,8], Helena Streit[1,4], Michele S. Dechoum[1,9], Nathan C. Fonsêca[1,10], Tainá C. Rocha[1], Cássio C. Pereira[1,2], Adrian A. Garda[1,11], Helena G. Bergallo[1,12], Fabricius M. C. B. Domingos[3], Geraldo W. Fernandes[1,2]

[1]Knowledge Center for Biodiversity, Belo Horizonte, MG, 31270–901, Brazil. [2]Ecologia Evolutiva & Biodiversidade, Departamento de Genética, Ecologia e Evolução, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 31270–901, Brazil. [3]Laboratório de Evolução e Diversidade Zoológica, Departamento de Zoologia, Universidade Federal do Paraná, Curitiba, PR, 81531–990, Brazil. [4]Laboratório de Estudos em Vegetação Campestre, Departamento de Botânica, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 91501–970, Brazil. [5]Laboratório de Vertebrados, Departamento de Ecologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 21941–902, Brazil. [6]Instituto de Biociências, Universidade Federal de Mato Grosso do Sul, Campo Grande, MS, 79070–900, Brazil. [7]Instituto de Ciências Naturais, Humanas e Sociais, Universidade Federal de Mato Grosso, Sinop, MT, 78550–728, Brazil. [8]Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia, Manaus, AM, 69067–375, Brazil. [9]Laboratório de Ecologia de Invasões Biológicas, Manejo e Conservação, Departamento de Ecologia e Zoologia, Universidade Federal de Santa Catarina, Florianópolis, SC, 88040–900, Brazil. [10]Laboratório de Ecologia Vegetal, Departamento de Biologia, Universidade Federal Rural de Pernambuco, Recife, PE, 52171–900, Brazil. [11]Departamento de Botânica e Zoologia, Centro de Biociências, Universidade Federal do Rio Grande do Norte, Natal, RN, 59078–900, Brazil. [12]Departamento de Ecologia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, RJ, 20550–900, Brazil. *Corresponding author. Email: tiagoshizen@gmail.com

REFERENCES AND NOTES

1. A. Sewell, S. van der Esch, H. Löwenhardt, "Goals and commitments for the Restoration Decade: A global overview of countries' restoration commitments under the Rio Conventions and other pledges" (PBL Netherlands Environmental Assessment Agency, 2020); https://www.pbl.nl/en/publications/goals-and-commitments-for-the-restoration-decade.
2. S. J. Cooke, J. R. Bennett, H. P. Jones, *Conserv. Sci. Pract.* **1**, e129 (2019).
3. K. D. Holl, J. C. Luong, P. H. Brancalion, *Trends Ecol. Evol.* **37**, 777 (2022).
4. J. C. Luong, D. M. Press, K. D. Holl, *Biol. Conserv.* **280**, 109956 (2023).
5. C. Almeida, J. L. Reid, R. A. F. Lima, L. F. G. Pinto, R. A. G. Viani, *For. Ecol. Manag.* 10.1016/j.foreco.2023.121628 (2024).
6. C. Jakovac *et al.*, *Sci. Tot. Environ.* **906**, 167558 (2024).
7. T. S. P. Toma, G. E. Overbeck, M. de S. Mendonça, G. W. Fernandes, *Perspect. Ecol. Conserv.* **21**, 25 (2023).
8. S. Ballari, C. Roulier, E. Nielsen, J. C. Pizarro, C. B. Anderson, *Conserv. Societ.* **18**, 298 (2020).
9. S. Grattan, J. Spring, "Brazil launches $204 million drive to restore Amazon rainforest," *Reuters* (2023).
10. M. Euler, "Atlantic Forest: 100M native species seedlings to be planted by 2026," *Agência Brasil* (2023).
11. D. I. Urzedo, R. Fisher, F. C. M. Piña-Rodrigues, J. M. Freire, R. G. P. Junqueira, *Restor. Ecol.* **27**, 768 (2019).
12. L. Balaguer, A. Escudero, J. F. Martín-Duque, I. Mola, J. Aronson, *Biol. Conserv.* **176**, 12 (2014).

10.1126/science.adn3767

China's plan to control methane emissions

Methane is a powerful greenhouse gas, and its potential effects on global warming far exceed those of carbon dioxide (*1*). China, which has lagged behind developed countries in controlling methane emissions and has not signed the global methane pledge (*2*), has become the largest anthropogenic methane emitter of the world (*3*). However, in November 2023, the Ministry of Ecology and Environment and 10 other departments jointly issued China's first Methane Emission Control Action Plan (*4*). The plan is a positive step, but China will need to take additional action to effectively address methane emissions.

In 2022, China emitted 55,676 kilotons of methane, accounting for about 15.6% of the total global methane emissions (*5*). China's main anthropogenic methane sources include the coal and gas industry, rice cultivation, livestock, and waste (*3*). By playing a part in global warming, methane emissions increase the risk of food insecurity, disease transmission, and natural disasters in China and beyond (*6*, *7*). Methane also increases premature mortality in humans by contributing to the formation of hazardous air pollutants such as ground-level ozone (smog) (*7*).

The 2023 plan, an update to China's 2007 National Program on Climate Change, focuses on monitoring, quantifying, reporting, and supervising the methane emitted by the energy, agriculture, garbage, and sewage treatment sectors. The plan strengthens global methane governance, ensures that China cooperates with global efforts, and requires that China consider the relationship between methane emission control, energy security, and food security. National administration authorities are responsible for strictly supervising the execution of the plan to ensure that methane emission control targets are achieved. By implementing these measures, China will begin to fulfill its international obligations to control methane emissions.

To ensure the success of methane mitigation efforts, the Chinese government should commit to full transparency and embrace new technology. Methane data collection should use the most advanced international methane data collection methods, in which accurate empirical measurements replace generic emission factor estimations (*8*), and China should share all data with the international community. Enterprise-led efforts to mitigate carbon emissions, such as those in the technology industry (*9*) and in animal husbandry (*10*) will also substantially aid in controlling methane. China should invest in scientific and technological advances and accelerate their commercialization. These steps can maximize the effectiveness of China's methane control plan.

Heyuan You
School of Public Administration, Zhejiang University of Finance and Economics, Hangzhou 310018, Zhejiang, China and Department of City and Regional Planning, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. Email: youheyuan@gmail.com

REFERENCES AND NOTES

1. R. W. Howarth *et al*., *Clim. Change* **106**, 679 (2011).
2. US Department of State, "United States, European Union, and partners formally launch Global Methane Pledge to keep 1.5C within reach" (2023); https://www.state.gov/united-states-european-union-and-partners-formally-launch-global-methane-pledge-to-keep-1-5c-within-reach/.
3. Y. Zhang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2202742119 (2022).
4. Ministry of Ecology and Environment of People's Republic of China *et al*., "Notice of 11 departments including the Ministry of Ecology and Environment on the issuance of the 'Methane Emission Control Action Plan'" (2023); https://www.mee.gov.cn/xxgk2018/xxgk/xxgk03/202311/t20231107_1055437.html [in Chinese].
5. Methane Tracker (2023); https://www.iea.org/data-and-statistics/data-tools/methane-tracker. In the Country dropdown, select China.
6. D. T. Shindell *et al*., *Science* **335**, 183 (2012).
7. UN Environment Programme, "Global methane assessment: Benefits and costs of mitigating methane emissions" (2021).
8. UN Environment Programme, "An eye on methane: The road to radical transparency" (2023).
9. K. He *et al*., *Science* **377**, 1498 (2022).
10. R. Wang *et al*., *The Innovation* **3**, 100220 (2022).

10.1126/science.adn4233

Philippines must commit to carbon mitigation

In 2011, the Philippines' Climate Change Commission launched the National Climate Change Action Plan, which classified climate change adaptation as an "anchor strategy" and downplayed the importance of climate change mitigation (*1*). In 2023, the country unveiled the Philippine Development Plan, which includes an update to its climate change agenda. Although the new plan includes mitigation, it is still seen as merely a benefit of adaptation (*2*). The Philippines must reassess its priorities and commit to net-zero carbon emissions.

Although this large, archipelagic, lower-middle-income country (*3*) is a signatory to the Paris Agreement, only 2.71% of its Nationally Determined Contribution is unconditional (*4*), with the rest being contingent on international aid. Unlike most of its neighbors in the Association of Southeast Asian Nations, the Philippines has yet to make a net-zero emissions pledge (*5*). The lack of political will to commit to climate change mitigation has resulted in policies that undermine global goals, such as the steady upward trend in the carbon intensity of the Philippines' electricity grid (*6*) and the continued importation of fossil fuel (*7*).

Integrated assessment models show that emissions need to peak by 2025 and reach net zero by 2050 to achieve the long-term climate goals set by the Paris Agreement (*8*). The Philippines urgently needs to make a carbon neutrality pledge in solidarity with the global community. Such a commitment will pave the way for the development of a decarbonization portfolio that incorporates drastically reduced fossil energy use as well as engineered and nature-based CO_2 removal (*9*).

A more balanced climate policy will also allow the Philippines' emerging economy to better capitalize on synergies between mitigation and adaptation (*10*). For example, coastal ecosystems can be managed to sequester carbon (*11*) and to buffer against extreme weather. To ensure that future decarbonization measures remain viable in the face of climate change (*12*), the Philippines must replace its piecemeal approach to climate policy with an integrated systems outlook.

Maria Victoria Migo-Sumagang[1]*, Kathleen B. Aviso[2], Raymond R. Tan[2]
[1]Department of Chemical Engineering, University of the Philippines Los Baños, Laguna 4031, Philippines. [2]Department of Chemical Engineering, De La Salle University, Manila 0922, Philippines. *Corresponding author. Email: mpmigo@up.edu.ph

REFERENCES AND NOTES

1. "National Climate Change Action Plan 2011–2028" (Republic of the Philippines Environmental Management Bureau, 2011); https://climate.emb.gov.ph/wp-content/uploads/2016/06/NCCAP-1.pdf.
2. "Philippine Development Plan 2023–2028" (National Economic and Development Authority, 2023); https://pdp.neda.gov.ph/philippine-development-plan-2023-2028.
3. "Philippines data" (The World Bank, 2023); http://data.worldbank.org/country/PH.
4. "Philippine nationally determined contribution (NDC)" (Republic of the Philippines, 2021); https://niccdies.climate.gov.ph/files/documents/Philippines%20NDC%20Quick%20Facts.pdf.
5. K. E. Fahim, L. C. De Silva, F. Hussain, S. A. Shezan, H. Yassin, *Sustainability* **15**, 6961 (2023).
6. "Philippines energy profile" (International Renewable Energy Agency, 2023); https://www.irena.org/IRENADocuments/Statistical_Profiles/Asia/Philippines_Asia_RE_SP.pdf.
7. "2021 Key energy statistics" (Department of Energy, 2021); https://www.doe.gov.ph/energy-statistics/2021-key-energy-statistics-kes.
8. Intergovernmental Panel on Climate Change, "Climate change 2022: Mitigation of climate change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change" (Cambridge University Press, 2022); https://www.ipcc.ch/report/ar6/wg3/downloads/report/IPCC_AR6_WGIII_FullReport.pdf.
9. R. R. Tan *et al*., *Nat. Comput. Sci*. **2**, 465 (2022).
10. L. Chen *et al*., *Environ. Chem. Lett*. **20**, 2277 (2022).
11. A. Thorhaug *et al*., *Mar. Pollut. Bull.* **160**, 111168 (2020).
12. D. C. Marvin, B. M. Sleeter, D. R. Cameron, E. Nelson, A. J. Plantinga, *Sci. Rep*. **13**, 19008 (2023).

10.1126/science.adn5441

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

PLANT SYMBIOSIS

Signaling for fungal nutrition

Most vascular plants form symbioses in their roots with arbuscular mycorrhizal (AM) fungi. The fungi provide nutrients such as phosphate in return for lipids provided by the plant. The transcription factor RAM1 is required for AM symbioses and regulates lipid provisioning. Ivanov and Harrison found another mechanism by which plants control lipid production and transport to fungi. This system involves two cyclin-dependent kinase–like proteins that operate both in parallel and in conjunction with the established RAM1 pathway. This work gives insight into a regulatory process fundamental to plant nutrition and growth. —MRS

Science p. 443 10.1126/science.ade1124

Arbuscular mycorrhizal fungi and plant root cells, pictured here in a scanning electron microscopy image, exchange nutrients in a symbiotic relationship.

PHOTO: EYE OF SCIENCE/SCIENCE SOURCE

BIOCATALYSIS

Enzyme tackles carbon–silicon compounds

Methylsiloxanes are organosilicon compounds produced by humans for use in a wide range of consumer products. Because they are not naturally found in nature, they are not readily degraded by organisms and some also have the potential to bioaccumulate. Sarai *et al*. identified a cytochrome P450 enzyme that can perform a hydroxylation on the methyl groups of linear methylsiloxanes. They then expanded this activity using directed evolution, creating variants that were more efficient and also functioned on cyclic methylsiloxanes. Mechanistic experiments suggested that a second oxidation and an enzyme-facilitated rearrangement can lead to cleavage of the carbon–silicon bond and release of formaldehyde. —MAF

Science p. 438 10.1126/science.adi5554

QUANTUM OPTICS

Controlling interactions

A three-dimensional optical lattice filled with cold fermionic atoms is a powerful implementation of an optical atomic clock. Studying interactions in such systems can lead to both improved clock precision and insights into many-body physics. Hutson *et al*. investigated strontium-87 atoms placed in a cubic optical lattice and measured the effects of resonant dipole-dipole interactions. They found that the interactions caused a tiny clock shift, the magnitude of which could be controlled by varying the relative orientation of the probe light and the atomic dipoles. —JS

Science p. 384 10.1126/science.adh4477

CANCER IMMUNOLOGY

Moonlighting as tumor killers

$CD4^+$ T cells have been primarily valued for their helper functions in antitumor immune responses, but the extent to which they directly contribute to tumor killing remains unclear. Using a mouse model of cutaneous melanoma, Bawden *et al*. characterized the effector functions of $CD4^+$ T cells in generating protective antitumor immunity. Melanoma-specific $CD4^+$ T cells infiltrated tumors, adopted diverse effector states, and could provide tumor

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CHEMISTRY AUTOMATION

Better conditions for photochemistry

There has been an extraordinary burst of recent research in photochemistry and photocatalysis driven in part by the environmentally benign appeal of light as a source of reactivity. However, many of the studies showcase small-scale reactions, and scale-up relies on a patchwork of different technologies that can require substantial trial and error to optimize. Slattery *et al*. report a combined software and hardware platform that iteratively determines optimal, substrate-specific conditions for photochemical processes in a scalable, flow-based architecture. The closed-loop Bayesian optimization approach enhances overall and space-time yields of a variety of distinct reactions. —JSY

Science p. 382 10.1126/science.adj1817

CANCER

Blood tests for cancer

The aim of cancer-screening programs is to find cancer at early stages, before it is symptomatic, because early treatment can improve the outcome for patients. Blood-based multicancer screening tests are becoming available, but there are substantial challenges in how they should be used. In a Perspective, Micalizzi *et al.* discuss these issues, including the need to counsel people with a positive test, the challenges of identifying the cancer type, and the false-positive rate, which could lead to unnecessary invasive testing and anxiety. They propose that people should be selected for blood-based screening if they are at higher risk of developing cancer, and such individuals could be identified with artificial intelligence–based risk assessment. Although blood-based tests are likely to be an important tool for early cancer detection, how they are implemented must maximize predictive power and minimize harm. —GKA

Science p. 368 10.1126/science.adk1213

EVOLUTION

Mutable fitness

The benefits and costs of mutations that undergo natural selection can change depending on genetic interactions with subsequent mutations. In an enduring experiment, 12 lineages of *Escherichia coli* have been maintained for more than 75,000 generations, with each generation sampled and preserved. Couce *et al*. made transposon insertion libraries in ancestral and evolved strains taken at the 50,000 generation point and measured fitness in competition experiments using these samples. The numbers of beneficial mutations rapidly tailed off during long-term passage, with parallel changes in fitness cost and gene essentiality occurring across the lineages. The authors found nonessential genes that became essential and essential genes that became nonessential in all lineages. Predictability stemmed from the importance of loss-of-function mutations that scale with the length of the target genes. —CA

Science p. 383 10.1126/science.add1417

GAMMA-RAY ASTRONOMY

Electron acceleration in a black hole jet

Quasars contain an accreting supermassive black hole that ejects a jet of plasma moving at relativistic speeds. The acceleration process in relativistic jets is poorly understood. The H.E.S.S. Collaboration observed SS 433, a nearby stellar-mass analog to distant quasars, in teraelectron volt gamma rays and spatially resolved gamma-ray emission at different energies, finding a different distribution from previous x-ray observations (see the Perspective by Bosch-Ramon). By modeling the emission mechanism from plasma moving along the jet, the authors showed that electrons were accelerated to high energies at a shock front located several parsecs from the black hole. The same process might operate in other relativistic jets. —KTS

Science p. 382 10.1126/science.adi2048; see also p. 371 10.1126/science.adn3487

ENVIRONMENTAL POLICY

Identifying protected places

The Clean Water Act is a defining piece of environmental legislation in the US, but the waters that it protects from pollution have never been clearly defined. Greenhill *et al*. developed a machine learning model that uses geospatial data to predict which waters are covered by the Clean Water Act and trained and tested the model with jurisdictional determinations from the US Army Corps of Engineers. This work provides an estimate of the extent of protected waterways, as well as an understanding of the effects of Supreme Court and White House rules that have reinterpreted or changed the regulation. For a subset of sites with high predictive accuracy, their model can also act as decision support tool to expedite permitting. —BEL

Science p. 406 10.1126/science.adi3794

ECOLOGY

Small invader leads to big shifts

Human global activities have led to the movement of species from their origins to distant sites across the world. The influence of these displaced species on the existing ecology of their new location can vary from damaging to positive, and many of their impacts may be much more subtle than one might predict. For example, the big-headed ant, originally described in Mauritius, has spread throughout much of the subtropical and tropical world. Kamaru *et al*. characterized how its presence at the Ol Pejeta Conservancy in Kenya disrupted a mutualism between native ants and acacia trees that led to increased herbivory by elephants and ultimately a shift in lion prey species from zebra to buffalo (see the Perspective by Gaynor). —SNV

Science p. 433 10.1126/science.adg1464; see also p. 370 10.1126/science.adn3484

SYNTHETIC BIOLOGY

Accelerating evolution in *E. coli*

Mutations introduced when copying the genomic DNA of an organism can be selected for if they provide an advantage to the offspring, but high levels of mutation in the genome can cause catastrophic defects and are selected against. Thus, organisms acquire new functions very slowly. Tian *et al*. introduced user-defined DNA into an *Escherichia coli* cell in such a way that it is selectively copied and rapidly mutated by distinct replication machinery without affecting the organism's genome (see the Perspective by Williams and Liu). This approach massively accelerates the evolution of new function from user-defined DNA sequences without passing on catastrophic defects to offspring. —DJ

Science p. 421 10.1126/science.adk1281; see also p. 372 10.1126/science.adn3434

ORGANIC CHEMISTRY

Phosphorus steers a Claisen rearrangement

To the untrained eye, the Claisen rearrangement's manner of swapping connectivity between carbon and oxygen atoms in a six-atom framework looks almost like a magic trick. The reaction is also a challenge to control, because there no obvious binding points for a catalyst

or modes for acceleration. G. Zhang *et al*. now report that a chiral phosphorus-based catalyst can induce enantioselectivity in the Eschenmoser-Claisen variation to produce amides with quaternary stereocenters. Key to the activation protocol is an imine reduction to form a phosphorus–nitrogen bond that sets up the requisite geometry and is later cleaved by a silane to achieve turnover. —JSY

Science p. 395 10.1126/science.adl3369

IMMUNOLOGY

A constraint on inflammasome complexes

The formation of multiprotein complexes called inflammasomes is critical for innate immune responses to infection but can also underlie autoinflammatory disorders. Coombs *et al*. showed that whereas the NLRP3 and NLRP6 members of the NOD-like receptor family formed inflammasomes in cells, NLRP12 did not. Instead, NLRP12 acted as an endogenous inhibitor of NLRP3 inflammasome formation. A disease-associated NLRP12 mutant failed to suppress NLRP3 inflammasome activation, suggesting that the NLRP3 inhibitors currently in clinical trials could be used to treat NLRP12-based autoinflammatory disorders. —JFF

Sci. Signal. (2024) 10.1126/scisignal.abg8145

protection independently of other lymphocytes. CD4+ T cells provided protection through multiple partially redundant cytotoxicity pathways. These results demonstrate that tumor-infiltrating CD4+ T cells are equipped to contribute to tumor control through multiple helper and effector modes. —CO

Sci. Immunol. (2024) 10.1126/scimmunol.adi9517

IMMUNOLOGY

ZEB2 controls the ABCs of ABCs

Age-associated B cells (ABCs) are a distinct subset of B cells that accumulate as we age and during some chronic infections. ABCs also contribute to the pathogenesis of certain autoimmune diseases such as systemic lupus erythematosus and multiple sclerosis. Dai *et al*. report that the transcription factor ZEB2 is critical in driving ABC formation and pathogenicity in both mice and humans. ZEB2 promotes the gene signature, phenotype (e.g., CD11c expression), and function (e.g., phagocytic ability) of ABCs. Moreover, ZEB2 represses MEF2B, a transcription factor that instructs germinal center development, which consequently steers ABCs toward an extrafollicular response. ZEB2's regulation of ABCs also requires JAK-STAT signaling, which suggests that targeting this pathway may reduce ABCs in autoimmune disease. —STS

Science p. 413 10.1126/science.adf8531

AIR POLLUTION

More in the air than we thought

Air pollution from gaseous organic compounds generated by petrochemical extraction typically is estimated using measurements of a subset of those species, volatile organic compounds. He *et al*. showed that this approach can vastly underestimate the true magnitude of the problem. Their aircraft-based measurements of total gas-phase organic carbon concentrations over the Athabasca oil sands region of Alberta, Canada, revealed that emissions from that region alone were much larger than estimates made on the basis of more limited arrays of species by as much as a factor of 64. The underreported species included abundant precursors to secondary air pollution that must be included in organic carbon pollution monitoring and reporting. —HJS

Science p. 426 10.1126/science.adj6233

THIN FILMS

Water release

Freestanding oxide membranes have a variety of interesting applications, but pulling these materials off of the substrate after synthesis can be challenging. J. Zhang *et al*. found a new phase in the strontium aluminum oxide system capable of producing free-standing oxide films that are crack free across relatively large areas. The substrates are water soluble, allowing for an easy method with which to release the oxide films of interest. The material should enable the production of relatively high-quality films for a wide variety of potential applications. —BG

Science p. 388 10.1126/science.adi6620

CANCER

Alu repeats in cancer

Alu elements, a type of short genetic sequence that repeats many times, can be differentially associated with cancerous cells, but their biomarker potential is often overlooked because of the technical challenges stemming from their repetitive nature. Douville *et al*. developed a machine learning approach to profile Alu elements in individuals with or without cancer. The method was designed for high specificity in classifying cancer and was validated in multiple independent cohorts, with solid cancers particularly marked by a reduction in AluS subfamily elements. This study suggests that Alu elements may hold valuable information to increase the likelihood of early cancer detection. —CC

Sci. Transl. Med. (2024) 10.1126/scitranslmed.adi3883

IN OTHER JOURNALS

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

GENOMES

Sea anemones don't do TADs

Sea anemones acquired their name after a terrestrial flowering plant because of their radially symmetrical body plan and colorful appearance. They belong to an ancient phylum called Cnidaria, which includes corals, jellyfish, and hydra. Their body plan contrasts with the more familiar organization of bilateral animals such as humans, and given their deep evolutionary divergence, one might expect their genomic organization to also differ. Zimmermann *et al.* have discovered that the genomes of sea anemone species are organized differently from the way that human DNA is organized in the nucleus. Human cells use a set of molecular motors to arrange DNA into topologically associated domains, or TADs, to compartmentalize gene expression, but no TADs were detected in the two sea anemones species examined. TADs might have evolved in bilaterians to facilitate the contact of different gene-regulatory elements that became separated as the genome got bigger. —DJ

Nat. Commun. (2023) 10.1038/s41467-023-44080-7

WORKFORCE

Scoping out the terrain

How do graduate students navigating the increasing complexity of postgraduate careers view the process of academic professional development (PD)? Cavallo *et al.* saw a trend among their graduate students from feedback indicating that they were interested in PD opportunities even though their attendance at PD sessions was lacking. To investigate this mismatch, seven in-depth interviews with doctoral students from a large R1 university were conducted. Results suggested that the various components of PD offerings contained overlapping programs with little to no explanation for how graduate students could integrate and benefit from them. The authors recommend that institutions support coordination and communication between PD programs and develop a usable map or menu of offerings to graduate students. —MMC

Stud. Grad. Postdr. Educ. (2023) 10.1108/SGPE-03-2022-0022

The starlet sea anemone, *Nematostella* spp., is radially symmetrical and has a different body and genome organization compared with bilaterally symmetrical animals.

NEUROIMMUNOLOGY

Arranging around vulnerability

Peripheral sensory neurons, like the central nervous system, are protected from pathogens by anatomical barriers and immune cells. Lund *et al.* combined imaging with transcriptional analyses of single cells to characterize vascular endothelial cells and macrophages associated with dorsal root ganglia (DRG) in mice. Blood vessels of the DRG exhibited molecular, structural, and functional zonation along the arteriovenous axis. Macrophages expressing the scavenger receptor CD163 localized specifically to the regions of DRG vasculature with the highest blood permeability. These macrophages phagocytosed molecules circulating in the blood and were activated in response to systemic inflammation induced by lipopolysaccharide. A similar population of macrophages was identified in human tissues. —SHR

J. Exp. Med. (2024)
10.1084/jem.20230675

NEUROSCIENCE

Neurodevelopment to neurodegeneration

The brain contains resident immune or myeloid cells called microglia. Variants of the microglia-related gene *TREM2* (triggering receptor expressed on myeloid cells 2) have been associated with increased susceptibility to developing neurodegenerative diseases, including Alzheimer's disease. Although *TREM2* has been shown to mediate many critical microglia functions, its role during development has not been fully characterized. Using a combination of in vitro and in vivo approaches, Tagliatti *et al.* showed that *Trem2* contributes to pyramidal neuron metabolism in the mouse hippocampus during development. The lack of *Trem2* affected many metabolic pathways, ultimately resulting in abnormal neuronal transcriptomic and energetic profiles. Thus, impaired neuronal metabolism during development may determine increased susceptibility to neurodegeneration later in life. —MMa

Immunity (2023)
10.1016/j.immuni.2023.12.002

CARDIOLOGY

Protecting the protectors

A myocardial infarction, or "heart attack," takes place when blood flow, and thus oxygen, to part of the heart is blocked, causing ischemia. Effective treatment consists of restoring blood flow by various means, but reperfusion can stimulate immune infiltration and tissue damage. Not all of these immune cells are harmful, however: Macrophages expressing a protein called MerTK remove dying cells and other debris, helping the tissue heal. Shao *et al.* identified a transcription factor called ATF3 that helps to protect the heart by preventing the loss of MerTK$^+$ macrophages. They also found a possible way to stimulate this pathway and potentially improve cardiac repair in the clinical setting. —YN

Nat. Cardiovasc. Res. (2024)
10.1038/s44161-023-00392-x

MACHINE LEARNING

Accurate free energies in catalysis

Recent advances in machine learning techniques and electronic structure theory enable accurate atomistic simulations at large scales, bringing materials and chemical process modeling into a new era. Nevertheless, it is crucial to identify the optimal combination of these two approaches for a specific system. Using machine learning thermodynamic perturbation theory, Rey *et al.* demonstrated that a high level of theory (here, random phase approximation) is necessary to approach chemical accuracy for activating free-energy barriers of isomerization and cracking alkenes catalyzed by protonic zeolites. This is an important step in chemical processes designed for the valorization of long-chain paraffins, for which numerous lower-level electronic calculations have failed in the past. The proposed methodology could be extended to other types of catalytic chemical reactions and beyond. —YS

Angew. Chem. Int. Ed. (2023)
10.1002/anie.202312392

SOCIOLOGY

Identity shaped by genetic ancestry tests

To overcome limitations of prior studies of the impact of genetic ancestry tests on how people self-identify in terms of race and ethnicity, Roth and Yaylacı conducted a randomized controlled trial. They observed very low rates of change for racial identity and small but significant ethnic identity changes among those receiving test results relative to non-test-takers. Test-takers' aspirations influenced their likelihood of adding an ethnic identity reported by the test. More influential, for both adding and abandoning ethnic identities were the reported ancestry percentages arising from admixture analysis. —BW

Am. J. Sociol. (2023)
10.1086/728819

RESEARCH ARTICLE SUMMARY

EVOLUTION

Changing fitness effects of mutations through long-term bacterial evolution

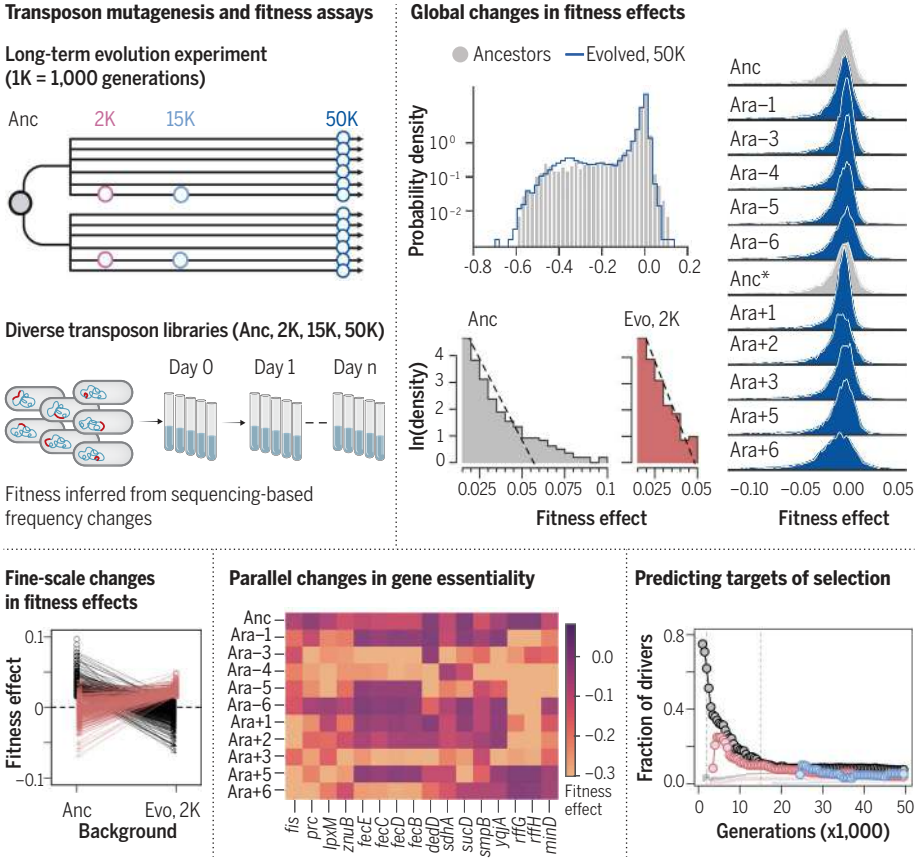
Alejandro Couce*†, Anurag Limdi†, Melanie Magnan, Siân V. Owen, Cristina M. Herren, Richard E. Lenski, Olivier Tenaillon*‡, Michael Baym*‡

INTRODUCTION: Evolution is constrained by the mutations accessible to natural selection. The benefits and costs of these mutations are described by the distribution of fitness effects (DFE). The DFE governs the tempo and mode of adaptation by capturing the fitness landscape of the local mutational neighborhood and reflects the mutational robustness of genotypes. However, the DFE need not remain static over evolution; with every accumulating mutation, the effects and accessibility of subsequent mutations may change through genetic interactions. Understanding how the DFE changes is important for models that seek to explain the speed of adaptation, maintenance of genetic diversity, and pace of the molecular clock.

RATIONALE: We quantified the effects of hundreds of thousands of insertion mutations in 12 populations of *Escherichia coli* through 50,000 generations of experimental evolution. We generated high-coverage transposon insertion libraries in the ancestral and evolved strains and measured the fitness effects of these mutations in bulk competitions. We characterized both the statistical properties of the DFEs and the effects of mutations in specific genes.

RESULTS: We saw no systematic change in the deleterious tail of the DFE. By contrast, the fraction of beneficial mutations declined rapidly, with its form approaching an exponentially distributed tail. At the gene level, we saw frequent changes in the fitness effects of insertion mutations in specific genes. Both the genetic identity and effect sizes of beneficial mutations changed over time. In the deleterious tail, there were frequent changes in the costs of specific mutations and even in gene essentiality. These changes often evolved in parallel across lineages and the changes in essentiality were only partially explained by structural variation. Despite pervasive changes in the fitness effects of particular mutations over time, many targets of selection could still be predicted by combining gene length with the ancestral DFE, owing to the benefit conferred by loss-of-function mutations during early adaptation.

CONCLUSION: Overall, the high-level features of the fitness landscape were largely unchanged over this multi-decade evolution experiment, except for truncation of the beneficial tail of the DFE. Over the short term, the drivers of adaptation were often predictable from the gene-level details of the DFE, especially combined with the length of genes available for beneficial mutations. As the populations accumulated more mutations over longer timescales, pervasive epistasis led to changes in the magnitude and even the sign of the fitness effects of many mutations, making some previously advantageous mutations deleterious and vice versa. Consequently, some evolutionary paths that were inaccessible to the ancestor became accessible to the evolving populations, while others were closed off. Moreover, many of the changes in the fitness effects of particular mutations, both beneficial and deleterious, occurred in parallel across the replicate populations. Thus, some features of the DFEs changed repeatedly and predictably over time, even as the overall form of the fitness landscape was largely unchanged. Taken together, our results demonstrate the dynamic—but often statistically predictable—nature of mutational fitness effects. ■



Changing distribution of fitness effects over evolution. Transposon mutagenesis of *E. coli* strains from a long-term evolution experiment and bulk fitness assays enable characterization of genome-wide and gene-level distribution of fitness effects (DFE). The overall shape of the DFE is conserved, except for a declining beneficial tail, while the effects of specific mutations and gene essentiality often evolve in parallel across populations. The ancestral DFE, combined with gene length, predicts drivers of adaptation.

The list of author affiliations is available in the full article.
*Corresponding author. Email: a.couce@upm.es (A.C.); olivier.tenaillon@inserm.fr (O.T.); baym@hms.harvard.edu (M.B.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
Cite this article as A. Couce *et al.*, *Science* **383**, eadd1417 (2024). DOI: 10.1126/science.add1417

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.add1417

RESEARCH ARTICLE

EVOLUTION

Changing fitness effects of mutations through long-term bacterial evolution

Alejandro Couce[1,2,3]*†, Anurag Limdi[4]†, Melanie Magnan[1], Siân V. Owen[4], Cristina M. Herren[4,5], Richard E. Lenski[6,7], Olivier Tenaillon[1,8]*‡, Michael Baym[4]*‡

The distribution of fitness effects of new mutations shapes evolution, but it is challenging to observe how it changes as organisms adapt. Using *Escherichia coli* lineages spanning 50,000 generations of evolution, we quantify the fitness effects of insertion mutations in every gene. Macroscopically, the fraction of deleterious mutations changed little over time whereas the beneficial tail declined sharply, approaching an exponential distribution. Microscopically, changes in individual gene essentiality and deleterious effects often occurred in parallel; altered essentiality is only partly explained by structural variation. The identity and effect sizes of beneficial mutations changed rapidly over time, but many targets of selection remained predictable because of the importance of loss-of-function mutations. Taken together, these results reveal the dynamic—but statistically predictable—nature of mutational fitness effects.

Evolution in asexual populations is a local process because selection can only act on mutants generated from existing genotypes. Thus, information about the relative fitness of the genotypes that can arise in the mutational neighborhood of the current population is essential for predicting future evolution. The distribution of fitness effects (DFE) captures the properties of an organism's mutational neighborhood: the proportion and magnitude of beneficial mutations determines the tempo and mode of adaptation, whereas the fraction of neutral and deleterious mutations defines the organism's robustness to mutational perturbations. Indeed, the DFE lies at the core of many theories describing fundamental evolutionary phenomena, including the speed of adaptation (*1*), fitness decay in small populations (*2*), the maintenance of genetic variation (*3*), the probability of parallel (*4*) versus divergent (*5*) evolution, the pace of the molecular clock (*6*), and the evolution of sex (*7*) and mutation rates (*8*). However, it is unclear if the general properties of local mutational neighborhoods remain static over long periods of evolution because with each successive mutation in a lineage, the accessibility and effect of subsequent mutations can be altered through genetic interactions (i.e., epistasis) (*9*, *10*).

The evolution of the overall shape of the DFE has received much theoretical and empirical attention. The beneficial tail of a DFE is expected to shorten as beneficial substitutions accumulate in an evolving population. Indeed, experiments with microbes show that the speed of adaptation steadily declines during adaptation to a constant environment (*11*–*13*), but it is generally unclear whether this deceleration reflects a decline in the availability or magnitude of new beneficial mutations (*13*). Besides becoming shorter, Extreme Value Theory predicts, using simple statistical principles, that the beneficial tail should become exponentially distributed as the population approaches a fitness peak (*14*–*17*). Although many studies support this model (*18*–*20*), some have reported non-exponential distributions of beneficial effects and it is unclear whether these exceptions represent populations far away from their fitness peak or, alternatively, the inadequacy of the theory (*21*–*23*). The picture is even more complicated for the deleterious tail of the DFE (*24*, *25*). Selection can favor mechanisms conferring increased robustness to mutational perturbations, especially at high mutation rates and in large populations (*26*–*29*), an idea with mixed support from studies with viruses and yeast (*30*–*32*). By contrast, recent theoretical work suggests that the genetic architecture of complex traits may lead to mutations being on average more detrimental on fitter genetic backgrounds (*33*), consistent with empirical data from crosses among diverse yeast strains (*34*).

However, these predictions address only the global (i.e., macroscopic) form of the DFE, with little attention to the fine-scale (i.e., microscopic) processes underlying changes in its overall shape. In the beneficial tail, the microscopic details may determine the extent to which adaptive pathways are predictable (*35*). For example, in the absence of interactions among mutations, adaptations will shorten the beneficial tail simply by the process of sampling without replacement, and therefore a complete DFE would suffice to specify the probabilities of all possible adaptive pathways in a given environment. By contrast, if each accumulated mutation changes the fitness effects and rank order of the remaining mutations (*36*), then predicting adaptive pathways would be impossible beyond the very short term.

In the deleterious tail, the microscopic details may reveal which physiological processes and genes are important or essential for fitness and how those processes and genes might change over time. Further, those details may provide evidence bearing on whether changes in the deleterious tail are the product of natural selection acting directly on mutational robustness or, alternatively, a byproduct of selection on related physiological processes (*26*). Moreover, the extreme end of this tail contains the set of essential genes whose loss would render the organism inviable. Prior work has shown that gene essentiality can vary greatly between species and even between strains of the same species (*34*–*39*). For instance, about a third of the essential genes in *Escherichia coli* are nonessential in *Bacillus subtilis*, and vice versa (*40*). Essentiality is also malleable over shorter timescales: in *Saccharomyces cerevisiae* and *Staphylococcus aureus,* many essential genes become nonessential following selection for suppressors (*41*, *42*), and horizontal gene transfer alters the essentiality of some core genes in *E. coli* (*39*). To what extent gene essentiality remains constant in the absence of direct selection, environmental change, or recombination is unclear. However, this issue has broad fundamental interest (e.g., understanding species' ecological and geographic ranges) (*43*) and applied consequences (e.g., the quest for the "minimal genome") (*44*).

Empirical studies of the DFE have generally been either small in scale (*45*) or focused on narrow genomic regions (*46*), and they typically lack detailed information on the level of adaptation of a given population to the test environment. Consequently, it has been difficult to distinguish among competing hypotheses about the evolution of the DFE during the course of adaptation, both at the macroscopic and microscopic levels. To address this challenge, one would ideally like to measure the relative fitness of the complete set of genome-wide mutants at multiple time points along a well-characterized adaptive trajectory. To do so, we turned to the Long-Term Evolution Experiment (LTEE), in which twelve populations of *E. coli*

[1]Université Paris Cité and Université Sorbonne Paris Nord, Inserm, IAME, F-75018 Paris, France. [2]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK. [3]Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM), 28223 Madrid, Spain. [4]Department of Biomedical Informatics, and Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA 02115, USA. [5]Department of Marine and Environmental Sciences, Northeastern University, Boston, MA 02115, USA. [6]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA. [7]Program in Ecology, Evolution, and Behavior, Michigan State University, East Lansing, MI 48824, USA. [8]Université Paris Cité, Inserm, Institut Cochin, F-75014 Paris, France.
***Corresponding author. Email: a.couce@upm.es (A.C.); olivier.tenaillon@inserm.fr (O.T.); baym@hms.harvard.edu (M.B.)**
†These authors contributed equally to this work.
‡These authors contributed equally to this work.

have been serially propagated in a glucose-limited minimal medium (*47*) for over 75,000 generations.

To quantify changes in the DFE, we generated genome-wide transposon insertion libraries in strains isolated at several time points from the LTEE, and we measured relative fitness values using high-resolution, bulk competitions. Such insertions typically lead to losses of function; by their nature, spontaneous loss-of-function mutations occur readily and so our approach surveys a large (but not complete) portion of the fitness landscape accessible by single-step mutations. Of note, we also observed two types of more subtle effects. First, an insertion in the C terminus of a gene may cause only a partial loss of function or even a change in function (*48*). We observed several examples of this outcome, including insertions in this region that were not merely tolerated but conferred large fitness benefits (fig. S1). Second, the positions of many beneficial insertions, including in intergenic regions and genes upstream of known targets of adaptation in the LTEE, suggest impacts on gene expression (fig. S1).

Our experimental system covered a large fitness gradient (>70% gains) which was generated by selection of spontaneous mutations in a constant environment, with no horizontal gene transfer (*11*). It is therefore suitable for detecting evolutionary trends in mutational robustness and the size of the essential gene set. Moreover, the most important mutations driving adaptation have been identified from signatures of parallelism in whole-genome sequences (*49*, *50*), allowing predictions based on the DFE at one time point to be compared with the actual fate of mutations observed during later evolution. Lastly, by comparing patterns in changing fitness effects across multiple independently evolving lineages, we can characterize the extent to which changes in the DFE are idiosyncratic or parallel.

Results

High-throughput insertion mutagenesis and fitness measurements

We performed two sets of experiments that analyzed the DFEs of many clones from the LTEE. In one experiment, we focused on changes in mutational robustness and gene essentiality during evolution. To do so, we constructed high-coverage transposon libraries in the LTEE ancestors (REL606 and REL607) and a clone isolated from each population (Ara+1 to Ara+6 and Ara−1 to Ara−6) at 50,000 generations. In the other experiment, we focused on the early, rapid changes in the properties of the beneficial tail. To that end, we made transposon libraries in the ancestor and clones sampled at 2000 and 15,000 generations from two populations (Ara+2 and Ara−1), when fitness had increased by ~25 and ~50%, respectively (*11*). In both experiments, we obtained >100,000 unique insertions, disrupting >78% of the genes with >95% overlap in genes disrupted in the ancestral and evolved libraries (fig. S2).

We estimated the fitness effects of all these mutants as selection coefficients, which we calculated from the frequency trajectories of every allele based on high-throughput sequencing during bulk competition assays under the same conditions as in the LTEE (Fig. 1, Fig. 2A, and Methods). This sequencing-based approach resolves the identity of each mutant at the molecular level; it allows us to interrogate both overall trends and the microscopic details of the locally accessible mutational landscape. We inferred fitness effects relative to a set of reference mutations, which consisted of insertions in known or presumed neutral loci, in the same transposon library (Fig. 1D and Methods). This approach allows relative fitness effects to be compared across the LTEE strains. The resulting fitness estimates were highly reproducible between technical replicates and consistent with independent estimates obtained from pairwise competitions between engineered deletion mutants and their unmutated parents (fig. S3 and data S1).

No systematic changes in the overall shape of the DFE

To investigate whether the overall form changed over time, we first compared the DFEs of the two LTEE ancestors and a clone from each of the 12 populations evolved independently for 50,000 generations. We excluded two evolved samples from further analyses because their fitness measurements were unreliable for technical reasons and therefore not comparable to the ancestor. In Ara+4, the within-gene measurement variability for fitness was extremely high and the correlation between technical replicates was poor (fig. S4A). In Ara−2, a few insertion mutations increased rapidly and outcompeted other mutations (fig. S4, B and C), which made the measurements unreliable and systematically biased [see supplementary materials (SM), text 1, for more details]. The exclusion of populations Ara-2 and Ara+4 from further analyses does not substantively alter our conclusions (fig. S5). Overall, most mutations are nearly neutral (within ~2 to 3% of neutrality, depending on the strain), but in all cases having a much heavier tail of deleterious mutations than beneficial mutations (Fig. 2B), consistent with previous results (*30*–*32*). The aggregate DFEs for the ancestors and evolved lines were nearly identical, except for an excess of mutations that are beneficial ($s > 0.03$, an effect reliably distinguishable from measurement noise) in the ancestral over the evolved backgrounds (0.9 versus 0.5% of all mutations, respectively; Fig. 2C, note the logarithmic scaling). This difference in the supply of beneficial mutations and its evolutionary significance are examined in depth in our second experiment (see below).

There was no systematic directional trend in how the means of the DFEs changed during evolution (*t*-test based on population means: $P = 0.37$). Although the mean fitness effect differed significantly between the ancestor and several evolved lines considered individually (Fig. 2D), these differences varied in their direction (two evolved clones had higher means than the ancestor and three had lower means), and they are primarily driven by noisy measurements in the deleterious tail (fig. S6). Therefore, robustness measured as the overall mean of the DFE of insertion mutations did not systematically change during the 50,000 generations of adaptation.

The constancy of the deleterious tail we observe over time stands in contrast to a study that measured the DFE for 91 insertion mutations in hybrid yeast genotypes with fitness values spanning ~20%, in which deleterious effects were significantly worse in the more-fit backgrounds (*34*). A potentially important difference is that the fitness variation among the yeast backgrounds was generated by crossing two distantly related strains, whereas we use a series of backgrounds from lineages undergoing adaptation to the same environment in which we assessed the fitness effects of the new mutations. In any case, theoretical predictions about the tail of deleterious mutations differ substantially and have been guided mostly by plausibility arguments (*24*, *25*), and so these studies collectively help refine current models by clarifying their assumptions and narrowing the range of parameters.

Parallel changes in fitness effects over evolution

A conserved macroscopic distribution does not preclude microscopic changes in the effects of individual mutations. Therefore, we examined whether and how the fitness effects of the same insertion mutations varied between the ancestor and evolved strains. We restricted this analysis to insertions with fitness effects $s > -0.3$ in both the ancestor and evolved strain, as measurements of extremely deleterious effects have more measurement noise. The fitness effects of some mutations differed between the ancestral and evolved strains, with some becoming more deleterious and others less so (Fig. 3A). Depending on the evolved strain, between 3 and 6% of the mutations had significantly different fitness effects from those in the ancestor (Fig. 3B) and 13% had differential effects in at least one evolved strain.

We observed significant parallelism across the independent lineages in the genes with fitness effects that changed significantly over evolution. We first examined this possibility through hierarchical clustering of mutations that were roughly neutral in the ancestor ($s > -0.05$) and clearly deleterious in an evolved

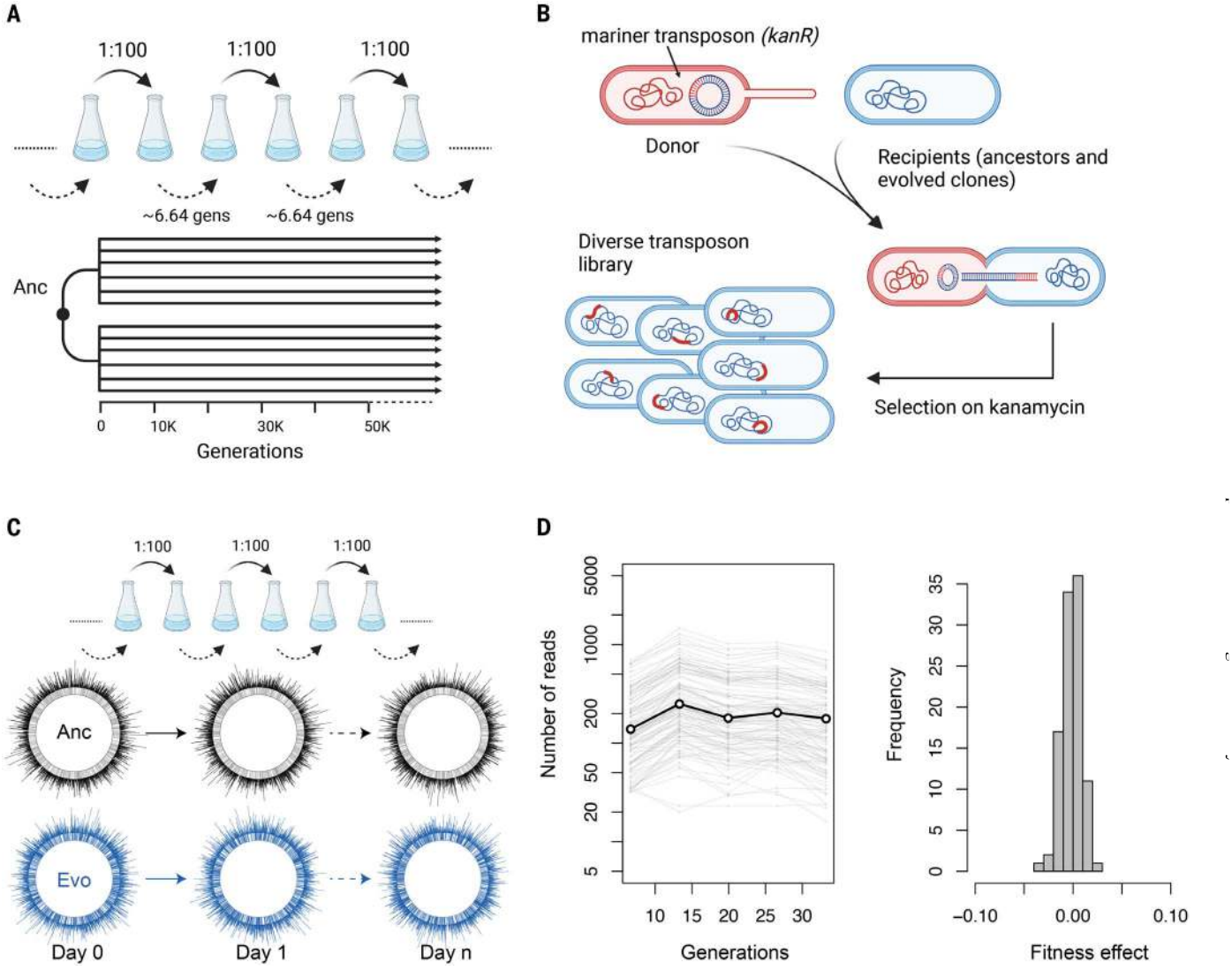


Fig. 1. Schematic representation of mutagenesis and fitness assay pipeline. (**A**) The Long-Term Evolution Experiment (LTEE) is an ongoing experiment in which 12 populations of *E. coli* evolve in and adapt to a glucose-limited minimal medium. (**B**) We created transposon libraries in the LTEE ancestors and clones from the evolving populations by transferring a mariner transposon with a kanR resistance gene, and selecting transconjugants on medium containing kanamycin. (**C**) We then propagated the resulting insertion libraries for several days in the same minimal medium as used in the LTEE and quantified the abundance of mutants over time using sequence data. (**D**) The abundance trajectories of a set of neutral loci were used to normalize coverage depth across time points, providing an internal reference to estimate selection coefficients of mutations (left). The fitness effects for these neutral loci were closely centered around zero (right). Panels (A) to (C) were created partially with Biorender.com.

strain ($-0.3 < s < -0.15$), and vice versa (Fig. 3, C and D). Although many such changes were specific to individual lineages, many others occurred in parallel across multiple lineages. To assess whether the observed parallelism was greater than that expected by chance, we compared the two complementary cumulative distributions of differential effects of gene disruptions in multiple lineages against a null distribution, which we generated by shuffling the fitness profiles of each population 10,000 times. Both the neutral-to-deleterious and deleterious-to-neutral transitions occurred in parallel more often than expected by chance (Fig. 3E). This outcome was insensitive to the chosen cutoff values (fig. S7). These parallel changes across independent lineages indicate that selection acted, directly or indirectly, to influence those changes.

Parallel changes in gene essentiality over evolution

Moving toward the extreme deleterious tail, we next investigated gene essentiality. Strict lethality or an absolute inability to replicate is often difficult to distinguish from extreme growth defects. For this analysis, we therefore define a gene as differentially essential between the ancestor and an evolved clone if (i) the fitness effect of disruption $s > -0.15$ in one strain and $s < -0.3$ in the other, or (ii) mutants were absent in the library prior to the bulk competition in the LTEE medium DM25, suggesting that the gene was essential in LB (see SM). This approach ensured that small changes in fitness effects (say from −0.31 to −0.29) were not counted as changes in essentiality. Also, our choice of $s < -0.3$ emerged from simulated competitions, which indicated that mutations with deleterious effects of this magnitude or larger could not be reliably distinguished from lethality (fig. S8). Using the cutoff $s < -0.3$, we detected 557 genes that were essential in DM25 in the ancestor (see SM).

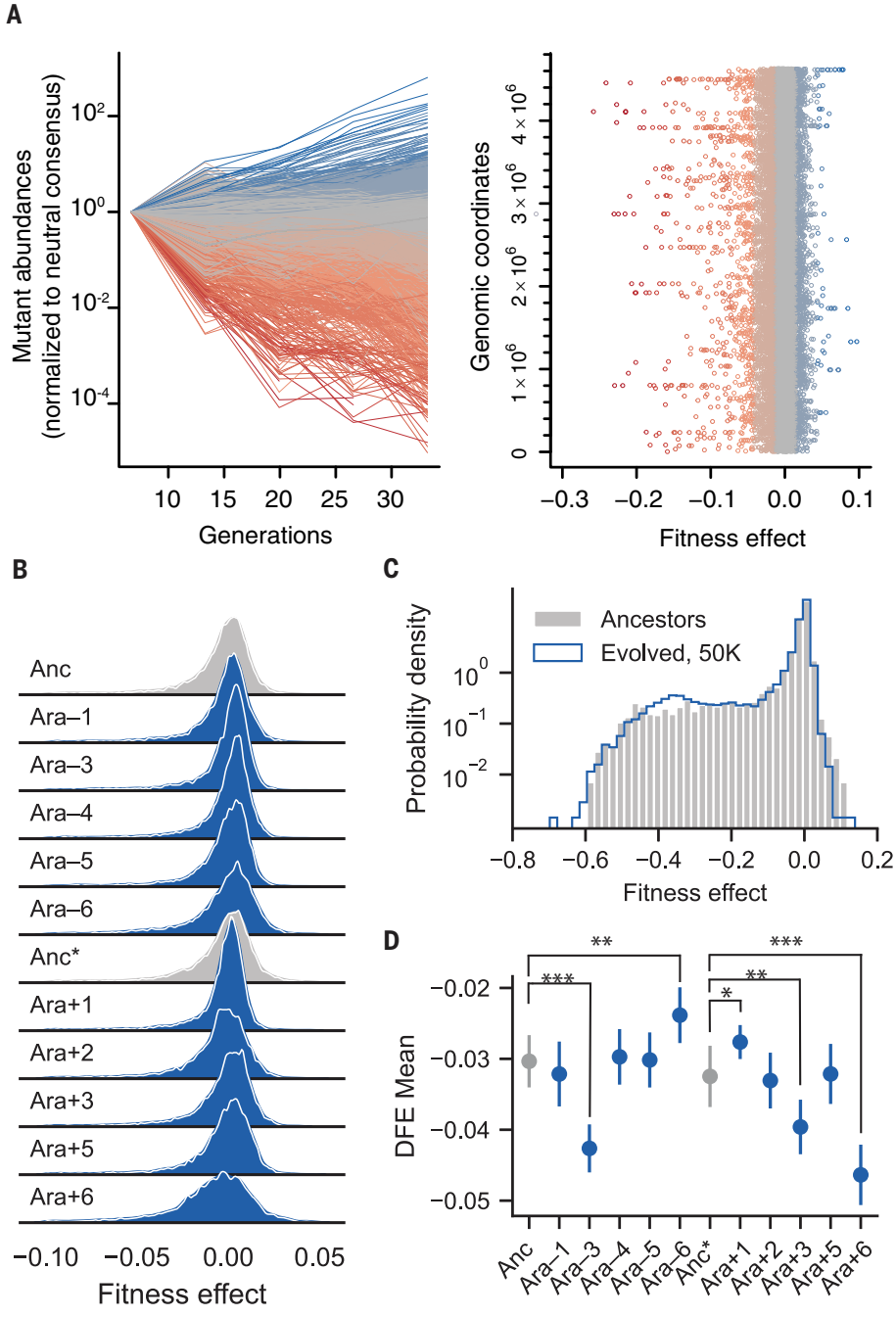


Fig. 2. The overall distribution of fitness effects (DFE) is largely unchanged after 50,000 generations. (**A**) Frequency trajectories of the whole mutant library in the ancestor (left), and mapping of estimated fitness effects along the chromosome (right). Colors indicate fitness effects, from deleterious (red) to beneficial (blue). (**B**) Ridge plot of the overall DFE in the two LTEE ancestors (Anc, REL606; Anc*, REL607) (gray), which differ by a neutral marker, and 50,000-generation clones sampled from each population (blue). We excluded two strains (Ara–2 and Ara+4) from further analyses (see text and fig. S4). The histograms were smoothed using kernel density estimation and are shown with a linear *y*-axis. DFEs are only shown for fitness effects ranging from −0.1 to 0.05, as the density outside these regions is very low. (**C**) Comparison of the aggregated DFEs of the ancestral and evolved strains. Here the histograms are plotted with a logarithmic *y*-axis to show more clearly the deleterious and beneficial tails of the DFEs. (**D**) Means of the DFEs: error bars indicate the 95% confidence interval in the estimate of means given the associated measurement noise in the bulk fitness assays. Statistically significant differences between the evolved lines and ancestors after Bonferroni correction for multiple tests are indicated (*Z*-test; ***$P < 0.001$, **$0.001 < P < 0.005$, *$0.005 < P < 0.05$).

We found genes that went from nonessential to essential and vice versa in all the LTEE lines (Fig. 4A and data S2). We confirmed two examples of differential gene essentiality in DM25 using clean deletion mutants in the ancestor REL606 and Ara–1 (fig. S9 and data S1). In total, 77 nonessential genes became essential in at least one evolved lineage and 97 essential genes became nonessential in at least one lineage, corresponding to ~17% of the essential genes in the ancestor. However, many more genes became nonessential in Ara–6 than in the other evolved lines (Fig. 4C) as a result of gene duplications discussed below. If we exclude Ara–6, then the nonessential-to-essential transition is more common. Indeed, across the other LTEE populations, we observed a significant tendency for more nonessential genes to become essential than the reverse change ($P = 0.0008$, Mann-Whitney U test). This asymmetry suggests that mutational robustness in terms of gene essentiality typically decreased during the LTEE. Both the essential-to-nonessential and nonessential-to-essential transitions occurred in parallel much more often than expected by chance (Fig. 4D). This outcome was insensitive to the exact cutoff values for essentiality (fig. S10) and it persisted when we partitioned essentiality changes by the culture medium (fig. S11). This parallel evolution in gene essentiality again implies that these changes result from selection. It is unclear how selection would act directly on essentiality; instead, this parallelism is presumably a correlated response to selection on gene expression or other metabolic traits.

Gene essentiality has previously been associated with highly expressed genes (*51*–*53*). We therefore examined whether changes in gene essentiality were associated with altered expression levels. We used a recently published RNA-Seq dataset for the LTEE ancestor and evolved strains at 50,000 generations (*54*). Consistent with previous findings, essential genes have higher expression levels on average than nonessential genes (fig. S12A). However, for those genes that became essential or nonessential during the LTEE, we find no significant differences in the normalized expression levels in the ancestor and evolved strains (fig. S12B). This result implies that changes in essentiality are not generally related to altered levels of gene expression.

Changes in gene essentiality could also arise as by-products of other mutations, especially losses or gains of other gene functions. Gene duplications can give rise to robustness by providing functional redundancy (*55*), whereas deletions can increase the essentiality of other genes by eliminating existing redundancies. We examined these possibilities by sequencing the ancestors and 50,000-generation clones with high coverage (>60-fold) to identify all large deletions and duplications in the evolved

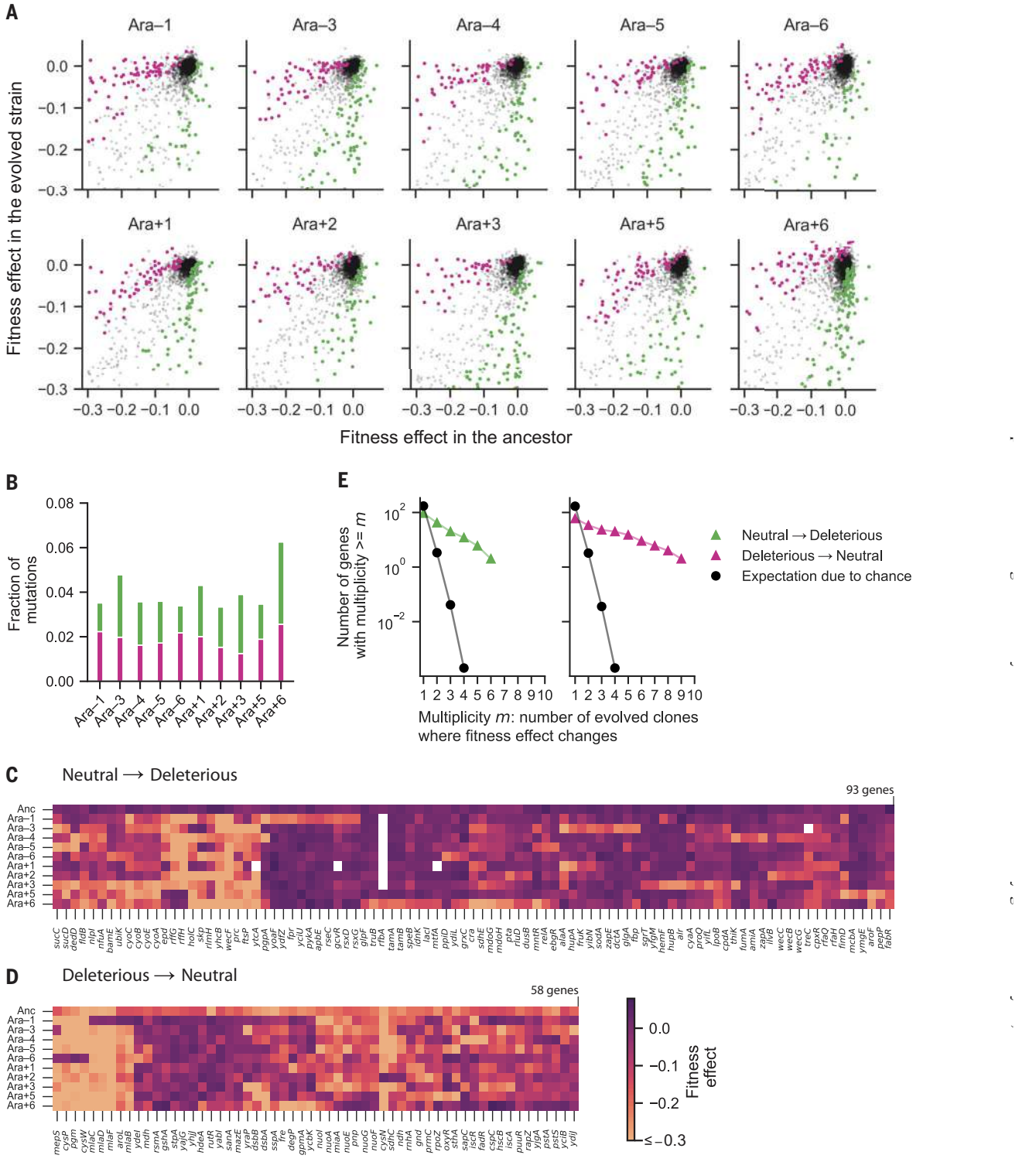


Fig. 3. Extensive and parallel changes in fitness effects of insertion mutations over evolution. (**A**) Pairwise comparison of fitness effects of mutations in nonessential genes ($s > -0.3$) between the ancestor (REL606) and each evolved strain. Purple, more deleterious in the ancestor; green, more deleterious in the evolved strain; Bonferroni corrected P-value < 0.05 (two-tailed Z-test). (**B**) Fraction of mutations (with $s > -0.3$ in both the ancestor and the evolved strain) with significant differences in fitness effects between the ancestor and each evolved clone (Bonferroni corrected P-values < 0.05). (**C** and **D**) Clustered heatmaps showing fitness effects (scale at right) of gene disruptions that became roughly neutral ($s > -0.05$) or clearly deleterious ($-0.3 < s < -0.15$) in at least one 50,000-generation strain. Genes that were deleted during evolution are shown in white. Genes with mutations conferring fitness effects below −0.3 (the threshold for essentiality) were set to −0.3 for the clustering and visualization. (**E**) Parallel changes in fitness effects. We estimated the expected number of parallel changes from chance alone by shuffling the profile of changes in fitness effects 10,000 times and counting how often the same genes had parallel changes (neutral to deleterious or deleterious to neutral) in at least m populations. The expectation is an average over 10,000 simulations and therefore can be < 1.

genomes. We then asked whether changes in gene essentiality were associated with these structural variants and their potential effects on redundancy given homologs in the ancestral genome (data S3). We found some cases where structural variants were associated with changes in gene essentiality. These cases included parallel deletions in most lineages that spanned the *rfb* operon and caused insertions in some paralogs to become essential in the evolved clones (fig. S13A). For most newly essential genes, however, we found no evidence that essentiality was caused by loss of redundant genes. With respect to duplications, the genome from population Ara–6 has two large duplications spanning ~300 and ~25 genes (fig. S13B). Ara–6 alone accounts for the majority of transitions from essential-to-nonessential genes and most of those transitions are found in the duplicated regions (fig. S13C). Further details and analyses are provided in the SM (see "Gains and losses of functional redundancy explain some, but not most, changes in essentiality").

Rapid contraction of the beneficial tail of the DFE

Our first experiment showed substantial changes in the small but critically important beneficial tail of the DFE. We therefore conducted additional experiments focused specifically on this tail and how it changed over evolution. Half of the ~70% fitness gain typically seen at 50,000 generations of the LTEE had already occurred by 5000 generations (*11*). We decided therefore to create transposon libraries in clones sampled at 2000 (2K) and 15,000 (15K) generations, when fitness had increased by ~25 and ~50%, respectively. To increase our resolution near selective neutrality, we divided each locus into five segments of equal length and then pooled the insertions within each segment. This approach expands the range of potentially observable beneficial mutations by enabling detection of polar effects within transcription units, effects linked to regulatory intergenic regions, and potentially subtle effects of insertions in the C-termini of protein-coding genes (fig. S2). As an added benefit, comparing the fitness effects among segments of the same locus helps identify potential artifacts and provides a within-experiment control to quantify the reproducibility of the fitness estimates (see SM, fig. S14).

We first focused on samples obtained from population Ara+2. Figure 5, A to C, shows that the fraction of beneficial insertion mutations is substantially larger in the ancestor than in the evolved backgrounds [6.8% for ancestor (Anc) versus 4.3 and 3.2% for 2K and 15K, respectively; $P < 0.044$ both cases, two-sample Kolmogorov–Smirnov (K-S) test]. By contrast and in agreement with what we observed for the 50,000-generation clones, the deleterious

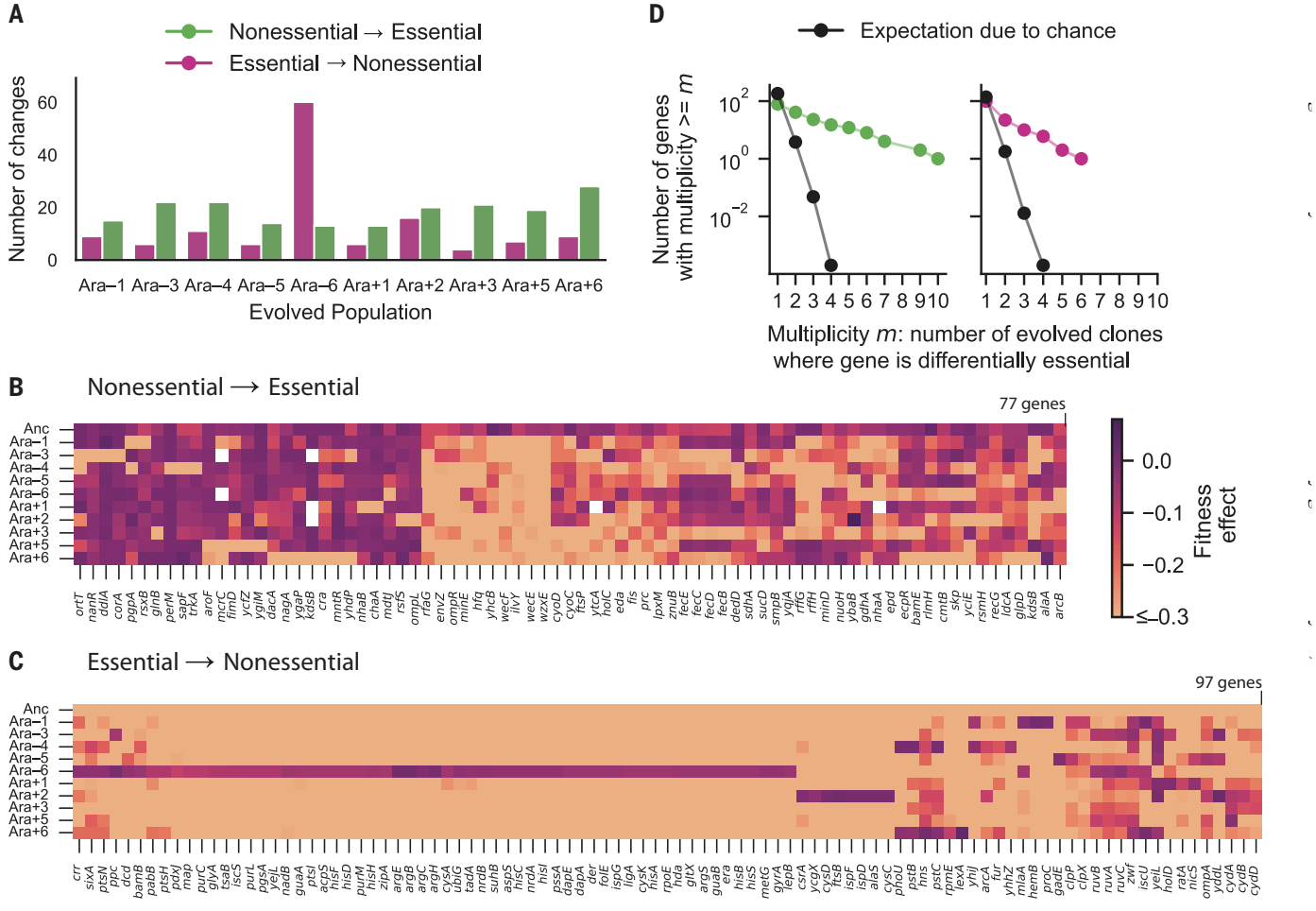


Fig. 4. Extensive and parallel changes in gene essentiality over evolution. (**A**) Number of genes that are differentially essential between the ancestor and each evolved strain. (**B** and **C**) Clustered heatmaps showing fitness effects (scale at right) of genes that evolved to become essential or nonessential in at least one 50,000-generation strain. Genes that were deleted during evolution are shown in white. Genes with mutations conferring fitness effects below −0.3 (the threshold for essentiality) were set to −0.3 for the clustering and visualization. (**D**) Parallel changes in gene essentiality. We estimated the expected number of parallel changes from chance alone by shuffling the profiles of changes in gene essentiality 10,000 times and counting how often the same genes had altered essentiality in at least *m* populations. The expectation is an average over 10,000 simulations and therefore can be < 1.

fraction is essentially constant across the three backgrounds (20.5% for Anc versus 21.0% and 19.6 for 2K and 15K, respectively; $P > 0.076$ both cases, two-sample K-S test). These patterns are consistent with analyses at the level of individual genes for both beneficial and deleterious mutations (Fig. 5D and fig. S15). To examine whether these results depend on the particular lineage, we also measured the DFEs for clones sampled at 2000 and 15,000 generations from population Ara–1, which accumulated a different set of beneficial mutations along its independent adaptive trajectory (see SM, table S1). At least two major features distinguish the evolutionary history of this lineage from that of Ara+2. First, Ara–1 fixed a mutation in *topA* early in the LTEE. Mutations in this gene confer among the largest fitness benefits seen in the LTEE for any single substitution (*56*); they fixed in 5 of the 12 populations but never reached detectable frequency in Ara+2. Second, Ara–1 evolved a mutator phenotype whereas Ara+2 retained the low ancestral mutation rate throughout the experiment; however, Ara–1 became hypermutable only after ~21,000 generations and hence poses no added complications to our analysis of the evolved clones from earlier generations. Despite independent histories, we obtained similar results for these two lineages, at both the macroscopic and microscopic levels (fig. S16). Our findings demonstrate that the contraction of the beneficial tail of the DFE occurred early and quickly as adaptation proceeded. Specifically, the small number of beneficial mutations that accumulated during the first 2000 generations were sufficient to have a significant impact on the adaptive landscape of the evolving population.

An exponential tail of beneficial mutations emerged during adaptation

Extreme Value Theory predicts on statistical grounds that the effects of beneficial mutations should be exponentially distributed when a population is well-adapted to its environment (*1*, *14*). Despite some empirical support (*18*–*20*), the evidence remains inconclusive owing to a severe limitation of most studies: without detailed knowledge of a population's evolutionary history, it is difficult to characterize its level of adaptation to a particular environment (*21*–*23*). Our data, by contrast, can test these ideas. We found that beneficial mutations in the evolved genetic backgrounds are well fit by an exponential distribution whereas this distribution is decisively rejected for the ancestor ($P < 0.001$ for Anc versus $P = 0.571$ and $P = 0.852$ for Ara+2 clones 2K and 15K, respectively; one-sample K-S test). We considered alternative distributions, but the exponential provides the best fit for the evolved backgrounds (see SM, table S2). Note that the exponential distribution is a special case of both the Weibull and gamma distributions, so it is not surprising that the data also fit well to them. These two distributions can be thought of as natural transitional shapes before reaching the limiting case of the exponential distribution. Indeed, the beneficial tail for the ancestor was fit to different degrees by both gamma and Weibull distributions ($P = 0.035$ and $P = 0.29$, respectively; one-sample K-S test), consistent with previous studies of viral and bacterial genotypes thought to be poorly adapted to their test environments (*19*, *21*). Overall, our results support the view that, after an early period of rapid adaptation to a new environment, the distribution of beneficial mutations becomes exponential. Thus, by analyzing changes in the DFE in a temporal series of genetic backgrounds becoming better adapted to their environment, we have reconciled otherwise disparate pieces of evidence relevant to general models of adaptation.

Changing identity of beneficial mutations and sign epistasis

We next sought to understand how changes in the DFE's macroscopic structure emerged from changes at the level of genes and mutations. We found that during the early phase of adaptation, deleterious mutations typically exhibit only slight epistasis across the three focal genetic backgrounds of the Ara+2 lineage (fig. S15). That is, the magnitude of their harmful effects may vary, but deleterious mutations in the ancestor tend to remain deleterious in the evolved backgrounds, consistent with the observed constancy of the deleterious tail (see fig. S17 for more details). By contrast, beneficial mutations are dominated by strong sign-epistatic interactions (Fig. 5D). Only 5.9% of the mutations beneficial in the ancestor are still beneficial at 2000 generations, with most becoming effectively neutral (76.9%) and some deleterious (17.2%) (Fig. 5E at left). This pattern also holds in the reverse direction: most beneficial mutations at 2000 generations are neutral (72.5%) or deleterious (17.9%) in the ancestor (Fig. 5E at left). Similar patterns occur when comparing how fitness effects changed between 2000 and 15,000 generations (Fig. 5E at right). Given the transitory nature of beneficial effects, we asked whether the overall DFE of the initially beneficial mutations retains even a slightly positive tendency at the later time points. In fact, it does not. The DFE of mutations that were beneficial in the ancestor becomes indistinguishable from a random sample of the parent distribution (Fig. 5F at left), and the same holds for the reverse scenario (Fig. 5F at right) ($P > 0.085$ both cases; two-sample K-S test). This regression to the mean persists even when we account for measurement noise around neutrality (fig. S14, C and D).

What explains this turnover in the identity of the beneficial mutations? In a previous study, the first five mutations to fix in one LTEE population were shown to exhibit diminishing-returns epistasis, such that their benefits declined in magnitude as the background fitness increased (*56*). However, it was unlikely a priori that these five mutations would show sign epistasis because they were chosen precisely because their combination was favored by natural selection (*57*). By contrast, another study analyzed the co-occurrence of fixed mutations across 115 lines of *E. coli* that evolved under thermal stress and found that sign epistasis was common (*58*). Moreover, that study found that the prevalence of different types of epistasis reflected the modular architecture of cellular traits: mutations affecting different modules tended to have additive effects whereas those impacting the same module tended to be redundant. We therefore investigated the extent of modularity in our data and found that beneficial mutations in the ancestral background often occurred repeatedly in the same operons (see SM, $P < 0.01$). Mutations in the same operon typically alter the same cellular process and often in similar ways and therefore the potential for redundancy at this functional level provides a simple explanation for why large sets of beneficial mutations disappear and other sets emerge as adaptation proceeds. More generally, the increased prevalence of sign epistasis with adaptation has also been predicted from general properties of the genotype-to-fitness map (*59*).

Target size is an important predictor of the genes that accumulate beneficial mutations

We identified a large set of loci that can produce beneficial mutations, including some known targets for adaptation in the LTEE (e.g., *topA*, *pykF*, *nadR*) (*49*). However, the fate of beneficial mutations in the course of evolution is determined not only by their individual fitness effects but also by their occurrence rate and the nature of their interactions with other beneficial mutations (*34*, *36*, *60*–*63*). Consequently, only a fraction of all possible beneficial mutations will contribute to adaptation in an evolving population. To gain further insight into this issue, we compared our data with metagenomic data previously obtained by sequencing whole-population samples from the 12 LTEE populations over the course of 60,000 generations (*50*). We see a significant but fairly weak correlation between our fitness estimates for mutations in the ancestor and the abundance of corresponding alleles during the LTEE ($r = 0.26$, Fig. 6A), and this correlation largely disappears when using the beneficial effects estimated in the evolved backgrounds. By contrast, the abundance of alleles in the metagenomic data correlates more strongly with the target size of the locus ($r = 0.71$, Fig. 6, B and C, and SM). These patterns are consistent with intense competition among independently

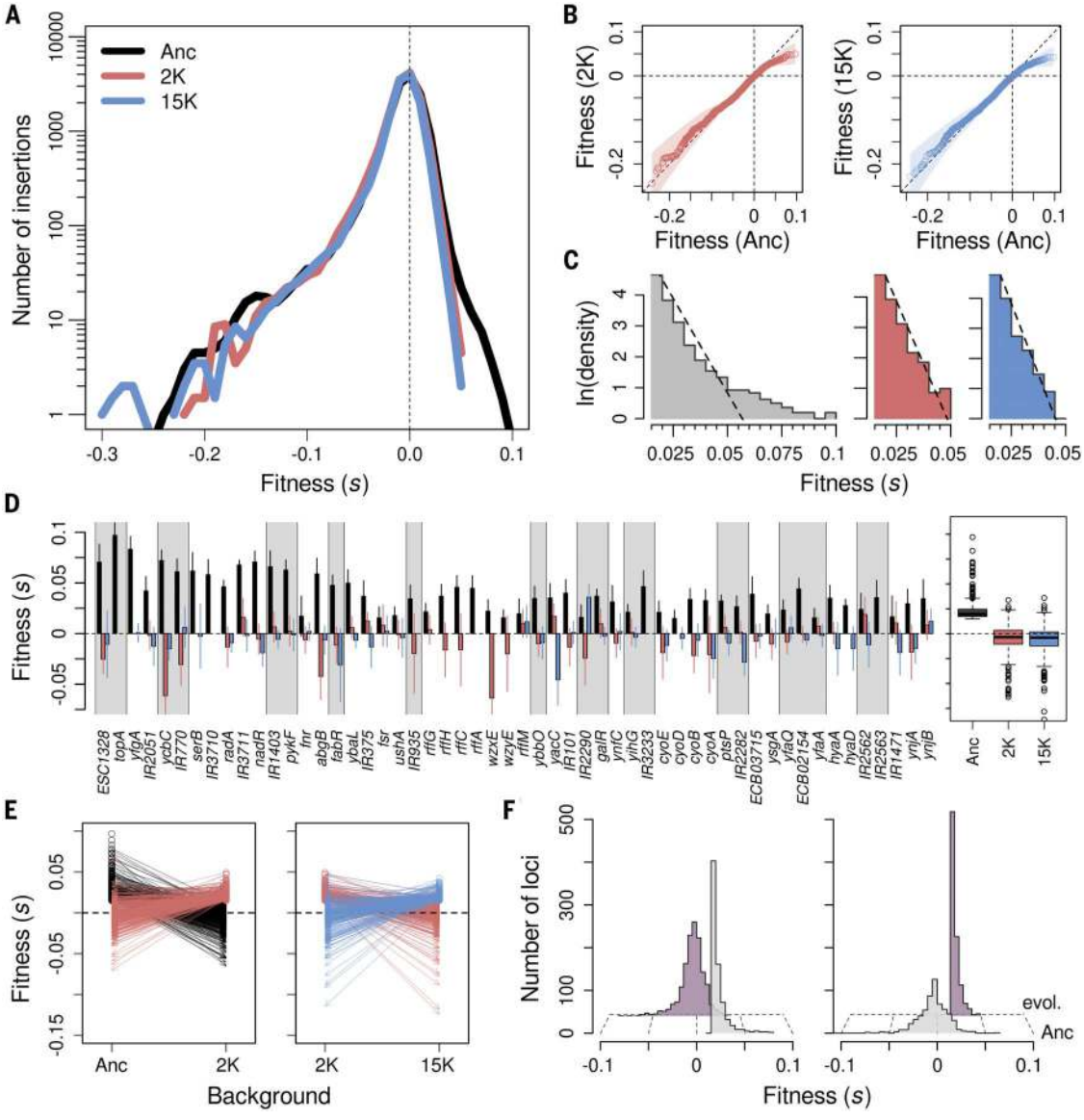


Fig. 5. Rapid contraction of the beneficial fraction over the first 15,000 generations. (**A**) DFEs in the ancestor (black), 2K (red) and 15K (blue) backgrounds from population Ara+2. Note that the logarithmic scaling of the *y*-axis exaggerates minor, nonsignificant differences in the extreme deleterious tails. (**B**) Only the beneficial tails underwent substantial changes during evolution, as indicated by comparing the cumulative fitness distributions for the ancestor and 2K evolved strain (left), and for the ancestor and 15K strain (right). Shaded areas show 95% bootstrapped confidence intervals. (**C**) Beneficial tails rapidly became exponentially distributed. Histograms show the best fits to exponential distributions (dashed lines) in the ancestor (gray), 2K (red), and 15K (blue). Note that all three x-axes use the same scale. (**D**) The genes and intergenic regions with the most beneficial alleles in the ancestral background and their fitness effects in the 2K (red) and 15K (blue) backgrounds. Gray shaded areas indicate members of the same transcription unit. (**E**) Most of the beneficial mutations available to the ancestor became neutral or deleterious in the 2K background (black arrows), whereas most beneficial mutations available in the 2K background were neutral or deleterious in the ancestor (red arrows). The same general pattern occurs when comparing beneficial mutations in the 2K and 15K backgrounds (right panel). (**F**) More than 90% of initially beneficial mutations became neutral or deleterious in later generations (left), and >90% of beneficial mutations from later generations were neutral or deleterious in the ancestor.

arising beneficial mutations (i.e., clonal interference), a pervasive phenomenon in the LTEE (*50*, *64*). Under intense clonal interference, the rate at which particular beneficial mutations occur may shape genomic evolution even more than their fitness effects (*65*). In any case, the best linear model includes target size as the most explanatory single variable but also includes significant contributions from the fitness effects in both the ancestral and 2000-generation genetic backgrounds (Fig. 6C and table S3). Finally, we note that a potentially important factor contributing to the observed weak correlations is that our methods involve insertion mutations, which usually, but not always (fig. S1), cause losses of function. Although losses of unused functions have contributed to adaptation in the LTEE (*49*, *66*), subtle changes that typically require point mutations have also been important in refining some functions (*35*, *49*, *67*).

Predicting future beneficial mutations as adaptation proceeds

Given that sign epistasis is widespread, it is natural to ask for how long the information about the particular loci in the beneficial tail of a

DFE can successfully predict the subsequent steps of adaptation. To address this question, we used the metagenomic data to record the alleles nearing fixation through time and calculated how many corresponded to loci for which we detected beneficial effects. We found that the ancestral DFE predicted most of the loci where mutations became dominant early in the LTEE populations; the predictive power decays rapidly but it was still evident for ~15,000 generations (Fig. 6D). This decay was largely driven by lineages that evolved hypermutability early in the LTEE; when these mutator populations are excluded from the analysis, the ancestral DFE retained significant predictive power through 50,000 generations (fig. S18A). In turn, the DFEs measured in the evolved backgrounds had lower predictive power and it took longer for their predictions to materialize; the latter effect may reflect the declining rate of adaptation. These patterns corroborate work showing that parallel genomic evolution was more common early in the LTEE than in later generations (*49*, *68*).

Finally, why does the ancestral DFE have such predictive power, when it is based on insertion mutations that represent only a limited set of all possible mutations from a functional standpoint? To address this question, we quantified how many loci with frequent beneficial mutations in the LTEE include mutations with presumed loss-of-function effects. To that end, we assumed that nonsense, frameshift, deletions, and insertions cause losses of function. We find that these presumptive inactivating mutations contribute most (>50%) of the early adaptive mutations in the LTEE, and they continue to be a sizable fraction over the long run (~25%, fig. S18B). Of note, another study with *Methylobacterium extorquens* adapted to use methanol as the sole carbon source also found that most early beneficial mutations appear to disrupt functions (*69*). These results support the "coupon-collecting" model of rapid evolution (*50*, *60*), in which "rough-and-ready" loss-of-function mutations dominate the early phase of adaptation to a new environment owing simply to their high rates of occurrence. Under this model, many initially beneficial mutations also become redundant because they inactivate the same functional module. This model implies that fitness effects alone are inadequate for predicting adaptive fixations, but taking target size into account compensates for this uncertainty. This interpretation satisfactorily explains our findings that the initial drivers of adaptation are predictable despite widespread and strong epistasis, and that target size is an important predictor of beneficial alleles that fix early when a population encounters a new environment.

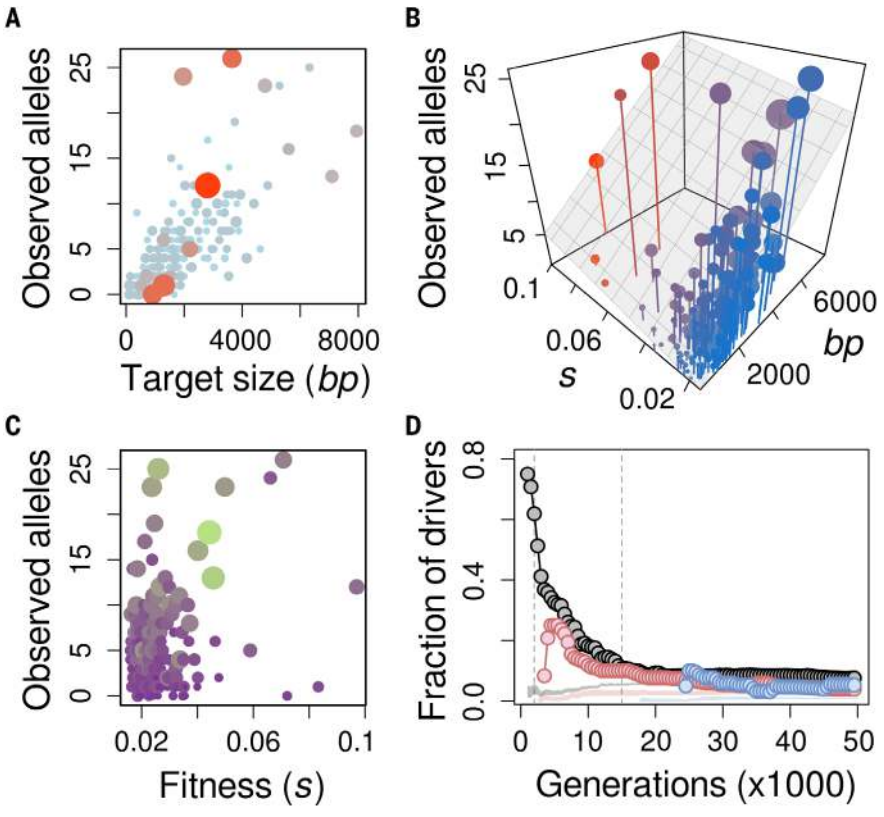


Fig. 6. Determinants of evolutionary outcomes. (**A** and **C**) The prevalence of the observed beneficial mutations in the LTEE is better explained by the mutational target size [(A) area and color of dots represent fitness] than by the magnitude of beneficial fitness effects measured in the ancestor [(C) area and color of dots represent target size]. (**B**) The best linear model for mutation prevalence includes fitness but is more strongly dependent on the mutational target size (area of dots represents target size and color represents fitness). (**D**) The predictive capacity of DFEs as a function of time in the LTEE. Values show the fraction of numerically dominant alleles at each generation that were captured by the DFE measured in the ancestor (black), 2K (red), and 15K (blue) backgrounds. For the ancestor, we measured this fraction across all 12 LTEE populations; for the evolved backgrounds, the fraction includes only the focal population. Shaded areas show the null expectations based on randomly sampled neutral and deleterious mutations.

Conclusions and Discussion

This paper began as two separate projects performed by two different teams, using similar but not identical methods. As we discussed our findings together, we discovered that each project reinforced and complemented the other. They reinforce one another by finding the same evolution of the overall form of the DFE; they are complementary because one project delved deeply into the fine-scale genetic changes in the deleterious tail while the other did so for the beneficial tail. Thus, together we have characterized changes in the DFE over the course of long-term evolution in a new environment at high resolution, including both the distribution's overall form and the effects of specific mutations. At a macroscopic scale, the idiosyncratic shape of the beneficial tail of the DFE became truncated, leading to an exponential distribution as predicted by some models (*14*, *15*). By contrast, there was no discernible change in the deleterious tail of the DFE, and mutational robustness—measured as the mean of the DFE across the replicate populations—was also unchanged over adaptation, suggesting that robustness was not under strong directional selection. With the notable exception of a population that evolved large duplications encompassing many genes, we observed a tendency for more genes to become essential than nonessential, lending some support to the "increasing costs" model of epistasis (*33*, *34*), but this effect disappeared when we examined the entire DFE. Overall, our results paint a complex picture of changing fitness effects that no simple model adequately captures.

At a microscopic scale, we found frequent changes in the fitness effects of particular mutations, even as the overall statistical properties of the DFE remained nearly constant. In the deleterious tail, there were frequent shifts in the effects of specific mutations (~13% of those in nonessential genes) over 50,000 generations, with some mutations becoming more deleterious and others less so. Similarly,

we also observed frequent changes in the identity of beneficial mutations over time, even over just 2000 generations. This dynamic pattern implies that the beneficial tail of the DFE is continually replenished by new and functionally different mutations as adaptation proceeds, even as other mutations lose their advantage. This shifting set of beneficial mutations over time helps to explain the sustained gains in fitness observed over tens of thousands of generations in the LTEE.

Prior work has shown that gene essentiality is not a static property of a species; however, the rate at which it changes is unknown (*37*, *41*, *42*). Here we show that ~3% of the genome had altered essentiality, which is similar to the variation in essentiality across diverse strains of *E. coli* when tested in three environments and often involving horizontally transferred genes (*39*). By contrast, the changes in gene essentiality that we observe in the LTEE happened over a much shorter evolutionary timescale, in the absence of any horizontal transfer, and without applying direct selection to suppress or enhance essentiality. Our demonstration of the fluid nature of essentiality indicates that the foundation of a minimal autonomous genome should not rely on a static snapshot of essentiality, because deleting genes can impact the potential for further genome reductions.

The ability to predict evolution remains elusive, in part because it requires a deep understanding of fitness landscapes and how they change. We found that the beneficial tail of the ancestral DFE is strongly predictive of the actual targets of selection in the LTEE, as inferred from the mutations nearing fixation in metagenomic data, particularly during early adaptation. This predictability reflects the prominent role that loss-of-function mutations had early in the LTEE, which seems applicable to other model systems (*60*, *61*, *63*, *69*). Over the long-term, however, pervasive epistasis resulted in declining predictability of these driver mutations, as the fitness effects of many mutations changed in magnitude and even their sign. Consequently, evolutionary paths that were inaccessible to the ancestor became available, whereas others were closed off, as reported recently in protein evolution (*70*). Because natural selection has steered most of the LTEE populations along similar trajectories, the paths that open or close are often the same across independently evolving lineages. Although we have shown that insertions capture the effects of a substantial fraction of the beneficial mutations in the LTEE, other types of mutations occur in the LTEE that might have more complex effects. For example, point mutations and structural rearrangements may be more likely to generate gains or changes of function, which could lead to more unpredictable outcomes, as seen with the evolution of citrate utilization in one of the 12 LTEE lines (*71*). Taken together, our results demonstrate the dynamic, but statistically predictable, nature of mutational fitness effects; they show that some features of evolutionary trajectories change repeatedly and predictably over time, even as the macroscopic features of the fitness landscape remain largely unchanged.

Methods

We used two suicide-plasmid delivery systems to construct the transposon libraries in the ancestor and several evolved clones from the *E. coli* long-term evolution experiment (LTEE) (Table S4). We then passaged the transposon libraries in DM25, the medium in which the populations have evolved (*47*), for 4 to 8 days, and we then isolated genomic DNA from the pool of mutants. In the first set of experiments discussed in the main text, we followed an approach we refer to as UMI-TnSeq that uses the mariner transposon carried by the pSC189 plasmid (*72*, *73*). We used this method to disrupt all genes in the ancestor and the 50,000-generation clonal isolates from all 12 LTEE populations. The genomic regions adjacent to the insertion site were captured using a tagmentation-based approach. To control for potential PCR bias, we attached unique molecular identifiers (UMIs) to individual molecules during PCR amplification (see Detailed Experimental Protocols in SM). In the second set of experiments, we used the INSeq methodology (*74*), focusing on the ancestor and the 2000- and 15,000-generation clones from two LTEE populations, called Ara+2 and Ara–1. We chose these populations because they neither evolved hypermutability nor diversified into stably coexisting lineages during the first 15,000 generations. Many other LTEE populations evolved one or both of these features, which would complicate testing our hypotheses (*50*).

After estimating the frequency of insertion mutants in the transposon libraries from bulk sequencing over the course of the fitness assays, we estimated the relative fitness of each mutant using linear regression of ln(frequency) of each mutant against the number of generations of selection during the assay. In the UMI-TnSeq analysis, we calculated the fitness effects of disrupting a given gene by averaging over all insertion sites in its interior (excluding the initial 10 and final 25% of the gene). In the INSeq analysis, we calculated fitness effects at the level of sub-genic regions by dividing each locus into five equally sized segments, while requiring a minimum size of 100 bp per segment.

There are two main differences between the UMI-TnSeq and INSeq approaches. First, polar effects within transcription units are expected to be more accentuated with the INSeq approach, because the 1.5-Kb insert carries two transcriptional terminators after the kanamycin resistance gene. The second difference concerns how regions adjacent to the insertion site are identified. The INSeq transposon encodes recognition sequences for the restriction enzyme MmeI, which cuts 20 bp away from its binding site and thus allows the capture of the 14 bp adjacent to the insertion site (fig. S19). This approach should minimize PCR bias because the genomic fragments are of uniform length, thus reducing the need to add UMIs during PCR. We also performed a replicate experiment with the Ara+2 samples to show that applying the UMI-TnSeq methodology to the INSeq transposon libraries yields essentially the same results (fig. S20).

REFERENCES AND NOTES

1. C. O. Wilke, The speed of adaptation in large asexual populations. *Genetics* **167**, 2045–2053 (2004). doi: 10.1534/genetics.104.027136; pmid: 15342539
2. M. Lynch, R. Bürger, D. Butcher, W. Gabriel, The mutational meltdown in asexual populations. *J. Hered.* **84**, 339–344 (1993). doi: 10.1093/oxfordjournals.jhered.a111354; pmid: 8409355
3. D. Charlesworth, B. Charlesworth, M. T. Morgan, The pattern of neutral molecular variation under the background selection model. *Genetics* **141**, 1619–1632 (1995). doi: 10.1093/genetics/141.4.1619; pmid: 8601499
4. L.-M. Chevin, G. Martin, T. Lenormand, Fisher's model and the genomics of adaptation: Restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution* **64**, 3213–3231 (2010). doi: 10.1111/j.1558-5646.2010.01058.x; pmid: 20662921
5. J. J. Welch, A. Eyre-Walker, D. Waxman, Divergence and polymorphism under the nearly neutral theory of molecular evolution. *J. Mol. Evol.* **67**, 418–426 (2008). doi: 10.1007/s00239-008-9146-9; pmid: 18818860
6. T. Ohta, The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* **23**, 263–286 (1992). doi: 10.1146/annurev.es.23.110192.001403
7. J. R. Peck, G. Barreau, S. C. Heath, Imperfect genes, Fisherian mutation and the evolution of sex. *Genetics* **145**, 1171–1199 (1997). doi: 10.1093/genetics/145.4.1171; pmid: 9093868
8. O. Tenaillon, B. Toupance, H. Le Nagard, F. Taddei, B. Godelle, Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics* **152**, 485–493 (1999). doi: 10.1093/genetics/152.2.485; pmid: 10353893
9. M. L. M. Salverda *et al.*, Initial mutations direct alternative pathways of protein evolution. *PLOS Genet.* **7**, e1001321 (2011). doi: 10.1371/journal.pgen.1001321; pmid: 21408208
10. D. Aggeli, Y. Li, G. Sherlock, Changes in the distribution of fitness effects and adaptive mutational spectra following a single first step towards adaptation. *Nat. Commun.* **12**, 5193 (2021). doi: 10.1038/s41467-021-25440-7; pmid: 34465770
11. M. J. Wiser, N. Ribeck, R. E. Lenski, Long-term dynamics of adaptation in asexual populations. *Science* **342**, 1364–1367 (2013). doi: 10.1126/science.1243357; pmid: 24231808
12. B. H. Good, M. M. Desai, The impact of macroscopic epistasis on long-term evolutionary dynamics. *Genetics* **199**, 177–190 (2015). doi: 10.1534/genetics.114.172460; pmid: 25395665
13. A. Couce, O. A. Tenaillon, The rule of declining adaptability in microbial evolution experiments. *Front. Genet.* **6**, 99 (2015). doi: 10.3389/fgene.2015.00099; pmid: 25815007
14. J. H. Gillespie, Molecular evolution over the mutational landscape. *Evolution* **38**, 1116–1129 (1984). doi: 10.2307/2408444; pmid: 28555784
15. H. A. Orr, The distribution of fitness effects among beneficial mutations. *Genetics* **163**, 1519–1526 (2003). doi: 10.1093/genetics/163.4.1519; pmid: 12702694
16. P. Joyce, D. R. Rokyta, C. J. Beisel, H. A. Orr, A general extreme value theory model for the adaptation of DNA sequences under strong selection and weak mutation. *Genetics* **180**, 1627–1643 (2008). doi: 10.1534/genetics.108.088716; pmid: 18791255
17. C. J. Beisel, D. R. Rokyta, H. A. Wichman, P. Joyce, Testing the extreme value domain of attraction for distributions of beneficial fitness effects. *Genetics* **176**, 2441–2449 (2007). doi: 10.1534/genetics.106.068585; pmid: 17565958
18. R. Kassen, T. Bataillon, Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nat. Genet.* **38**, 484–488 (2006). doi: 10.1038/ng1751; pmid: 16550173

19. R. Sanjuán, A. Moya, S. F. Elena, The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8396–8401 (2004). doi: 10.1073/pnas.0400146101; pmid: 15159545
20. E. E. Wrenbeck, L. R. Azouz, T. A. Whitehead, Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat. Commun.* **8**, 15695 (2017). doi: 10.1038/ncomms15695; pmid: 28585537
21. R. D. H. Barrett, R. C. MacLean, G. Bell, Mutations of intermediate effect are responsible for adaptation in evolving *Pseudomonas fluorescens* populations. *Biol. Lett.* **2**, 236–238 (2006). doi: 10.1098/rsbl.2006.0439; pmid: 17148371
22. M. J. McDonald, T. F. Cooper, H. J. E. Beaumont, P. B. Rainey, The distribution of fitness effects of new beneficial mutations in *Pseudomonas fluorescens*. *Biol. Lett.* **7**, 98–100 (2011). doi: 10.1098/rsbl.2010.0547; pmid: 20659918
23. D. R. Rokyta *et al*., Beneficial fitness effects are not exponential for two viruses. *J. Mol. Evol.* **67**, 368–376 (2008). doi: 10.1007/s00239-008-9153-x; pmid: 18779988
24. D. P. Rice, B. H. Good, M. M. Desai, The evolutionarily stable distribution of fitness effects. *Genetics* **200**, 321–329 (2015). doi: 10.1534/genetics.114.173815; pmid: 25762525
25. L. Perfeito, A. Sousa, T. Bataillon, I. Gordo, Rates of fitness decline and rebound suggest pervasive epistasis. *Evolution* **68**, 150–162 (2014). doi: 10.1111/evo.12234; pmid: 24372601
26. J. A. G. M. de Visser *et al*., Perspective: Evolution and detection of genetic robustness. *Evolution* **57**, 1959–1972 (2003). doi: 10.1111/j.0014-3820.2003.tb00377.x; pmid: 14575319
27. R. Sanjuán, J. M. Cuevas, V. Furió, E. C. Holmes, A. Moya, Selection for robustness in mutagenized RNA viruses. *PLOS Genet.* **3**, e93 (2007). doi: 10.1371/journal.pgen.0030093; pmid: 17571922
28. C. O. Wilke, J. L. Wang, C. Ofria, R. E. Lenski, C. Adami, Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412**, 331–333 (2001). doi: 10.1038/35085569; pmid: 11460163
29. J. D. Bloom *et al*., Evolution favors protein mutational robustness in sufficiently large populations. *BMC Biol.* **5**, 29 (2007). doi: 10.1186/1741-7007-5-29; pmid: 17640347
30. M. S. Johnson, M. M. Desai, Mutational robustness changes during long-term adaptation in laboratory budding yeast populations. *eLife* **11**, e76491 (2022). doi: 10.7554/eLife.76491; pmid: 35880743
31. A. Butković, R. González, I. Cobo, S. F. Elena, Adaptation of turnip mosaic potyvirus to a specific niche reduces its genetic and environmental robustness. *Virus Evol.* **6**, veaa041 (2020). doi: 10.7554/eLife.76491; pmid: 32782826
32. I. S. Novella, J. B. Presloid, C. Beech, C. O. Wilke, Congruent evolution of fitness and genetic robustness in vesicular stomatitis virus. *J. Virol.* **87**, 4923–4928 (2013). doi: 10.1128/JVI.02796-12; pmid: 23408631
33. G. Reddy, M. M. Desai, Global epistasis emerges from a generic model of a complex trait. *eLife* **10**, e64740 (2021). doi: 10.7554/eLife.64740; pmid: 33779543
34. M. S. Johnson, A. Martsul, S. Kryazhimskiy, M. M. Desai, Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science* **366**, 490–493 (2019). doi: 10.1126/science.aay4199; pmid: 31649199
35. R. J. Woods *et al*., Second-order selection for evolvability in a large *Escherichia coli* population. *Science* **331**, 1433–1436 (2011). doi: 10.1126/science.1198914; pmid: 21415350
36. D. M. Weinreich, R. A. Watson, L. Chao, Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005). doi: 10.1111/j.0014-3820.2005.tb01768.x; pmid: 16050094
37. G. Rancati, J. Moffat, A. Typas, N. Pavelka, Emerging and evolving concepts in gene essentiality. *Nat. Rev. Genet.* **19**, 34–49 (2018). doi: 10.1038/nrg.2017.74; pmid: 29033457
38. L. Parts *et al*., Natural variants suppress mutations in hundreds of essential genes. *Mol. Syst. Biol.* **17**, e10138 (2021). doi: 10.15252/msb.202010138; pmid: 34042294
39. F. Rousset *et al*., The impact of genetic diversity on gene essentiality within the *Escherichia coli* species. *Nat. Microbiol.* **6**, 301–312 (2021). doi: 10.1038/s41564-020-00839-y; pmid: 33462433
40. B.-M. Koo *et al*., Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. *Cell Syst.* **4**, 291–305.e7 (2017). doi: 10.1016/j.cels.2016.12.013; pmid: 28189581
41. G. Liu *et al*., Gene essentiality is a quantitative property linked to cellular evolvability. *Cell* **163**, 1388–1399 (2015). doi: 10.1016/j.cell.2015.10.069; pmid: 26627736
42. F. Rosconi *et al*., A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. *Nat. Microbiol.* **7**, 1580–1592 (2022). doi: 10.1038/s41564-022-01208-7; pmid: 36097170
43. O. M. Maistrenko *et al*., Disentangling the impact of environmental and phylogenetic constraints on prokaryotic within-species diversity. *ISME J.* **14**, 1247–1259 (2020). doi: 10.1038/s41396-020-0600-z; pmid: 32047279
44. E. Martínez-García, V. de Lorenzo, The quest for the minimal bacterial genome. *Curr. Opin. Biotechnol.* **42**, 216–224 (2016). doi: 10.1016/j.copbio.2016.09.001; pmid: 27660908
45. R. Sanjuán, Mutational fitness effects in RNA and single-stranded DNA viruses: Common patterns revealed by site-directed mutagenesis studies. *Philos. Trans. R. Soc. B* **365**, 1975–1982 (2010). doi: 10.1098/rstb.2010.0063; pmid: 20478892
46. H. Kemble, P. Nghe, O. Tenaillon, Recent insights into the genotype-phenotype relationship from massively parallel genetic assays. *Evol. Appl.* **12**, 1721–1742 (2019). doi: 10.1111/eva.12846; pmid: 31548853
47. R. E. Lenski, M. R. Rose, S. C. Simpson, S. C. Tadler, Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138**, 1315–1341 (1991). doi: 10.1086/285289
48. E. C. A. Goodall *et al*., The essential genome of *Escherichia coli* K-12. *mBio* **9**, e02096-17 (2018). doi: 10.1128/mBio.02096-17; pmid: 29463657
49. O. Tenaillon *et al*., Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* **536**, 165–170 (2016). doi: 10.1038/nature18959; pmid: 27479321
50. B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, M. M. Desai, The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017). doi: 10.1038/nature24287; pmid: 29045390
51. J.-F. Gout, D. Kahn, L. Duret; Paramecium Post-Genomics Consortium, The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLOS Genet.* **6**, e1000944 (2010). doi: 10.1371/journal.pgen.1000944; pmid: 20485561
52. J. L. Cherry, Expression level, evolutionary rate, and the cost of expression. *Genome Biol. Evol.* **2**, 757–769 (2010). doi: 10.1093/gbe/evq059; pmid: 20884723
53. J. Zhang, J.-R. Yang, Determinants of the rate of protein sequence evolution. *Nat. Rev. Genet.* **16**, 409–420 (2015). doi: 10.1038/nrg3950; pmid: 26055156
54. J. S. Favate, S. Liang, A. L. Cope, S. S. Yadavalli, P. Shah, The landscape of transcriptional and translational changes over 22 years of bacterial adaptation. *eLife* **11**, e81979 (2022). doi: 10.7554/eLife.81979; pmid: 36214449
55. M. Lynch, J. S. Conery, The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000). doi: 10.1126/science.290.5494.1151; pmid: 11073452
56. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193–1196 (2011). doi: 10.1126/science.1203801; pmid: 21636772
57. F. Blanquart, G. Achaz, T. Bataillon, O. Tenaillon, Properties of selected mutations and genotypic landscapes under Fisher's geometric model. *Evolution* **68**, 3537–3554 (2014). doi: 10.1111/evo.12545; pmid: 25311558
58. O. Tenaillon *et al*., The molecular diversity of adaptive convergence. *Science* **335**, 457–461 (2012). doi: 10.1126/science.1212986; pmid: 22282810
59. D. Greene, K. Crona, The changing geometry of a fitness landscape along an adaptive walk. *PLOS Comput. Biol.* **10**, e1003520 (2014). doi: 10.1371/journal.pcbi.1003520; pmid: 24853069
60. P. A. Lind, E. Libby, J. Herzog, P. B. Rainey, Predicting mutational routes to new adaptive phenotypes. *eLife* **8**, e38822 (2019). doi: 10.7554/eLife.38822; pmid: 30616716
61. R. L. Moran *et al*., Selection-driven trait loss in independently evolved cavefish populations. *Nat. Commun.* **14**, 2557 (2023). doi: 10.1038/s41467-023-37909-8; pmid: 37137902
62. D. R. Rokyta, P. Joyce, S. B. Caudle, H. A. Wichman, An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nat. Genet.* **37**, 441–444 (2005). doi: 10.1038/ng1535; pmid: 15778707
63. A. K. Hottes *et al*., Bacterial adaptation through loss of function. *PLOS Genet.* **9**, e1003617 (2013). doi: 10.1371/journal.pgen.1003617; pmid: 23874220
64. D. E. Deatherage, J. E. Barrick, High-throughput characterization of mutations in genes that drive clonal evolution using multiplex adaptome capture sequencing. *Cell Syst.* **12**, 1187–1200.e4 (2021). doi: 10.1016/j.cels.2021.08.011; pmid: 34536379
65. M. M. Desai, D. S. Fisher, A. W. Murray, The speed of evolution and maintenance of variation in asexual populations. *Curr. Biol.* **17**, 385–394 (2007). doi: 10.1016/j.cub.2007.01.072; pmid: 17331728
66. V. S. Cooper, D. Schneider, M. Blot, R. E. Lenski, Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *J. Bacteriol.* **183**, 2834–2841 (2001). doi: 10.1128/JB.183.9.2834-2841.2001; pmid: 11292803
67. R. Maddamsetti *et al*., Core genes evolve rapidly in the long-term evolution experiment with *Escherichia coli*. *Genome Biol. Evol.* **9**, 1072–1083 (2017). doi: 10.1093/gbe/evx064; pmid: 28379360
68. R. E. Lenski, Convergence and divergence in a long-term experiment with bacteria. *Am. Nat.* **190** (S1), S57–S68 (2017). doi: 10.1086/691209; pmid: 28731830
69. H.-H. Chou, H.-C. Chiu, N. F. Delaney, D. Segrè, C. J. Marx, Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192 (2011). doi: 10.1126/science.1203799; pmid: 21636771
70. Y. Park, B. P. H. Metzger, J. W. Thornton, Epistatic drift causes gradual decay of predictability in protein evolution. *Science* **376**, 823–830 (2022). doi: 10.1126/science.abn6895; pmid: 35587978
71. Z. D. Blount, J. E. Barrick, C. J. Davidson, R. E. Lenski, Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**, 513–518 (2012). doi: 10.1038/nature11514; pmid: 22992527
72. S. L. Chiang, E. J. Rubin, Construction of a mariner-based transposon for epitope-tagging and genomic targeting. *Gene* **296**, 179–185 (2002). doi: 10.1016/S0378-1119(02)00856-9; pmid: 12383515
73. L. Ferrières *et al*., Silent mischief: Bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J. Bacteriol.* **192**, 6418–6427 (2010). doi: 10.1128/JB.00621-10; pmid: 20935093
74. A. L. Goodman *et al*., Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**, 279–289 (2009). doi: 10.1016/j.chom.2009.08.003; pmid: 19748469
75. A. Limdi, S. V. Owen, C. Herren, M. Baym, LTEE-TnSeq-processed-data, version v1, Zenodo; (2022). doi: 10.5281/zenodo.6547537
76. A. Couce, LTEE-INSeq-processed-data, version v1, Zenodo; (2023). doi: 10.5281/zenodo.7985455
77. A. Limdi, baymlab/2022_Limdi-TnSeq-LTEE: November 18 2023, version v1.0, Zenodo; (2023). doi: 10.5281/zenodo.10155524
78. A. Couce, ACouce/LTEE2022: Couce-LTEE2022 Code, version V1, Zenodo; (2023). doi: 10.5281/zenodo.10236681

ACKNOWLEDGMENTS

We thank M. Johnson, M. Desai, A. Murray, T. Bernhardt, C. Souque, F. Rossine, and D. Eaton for valuable feedback and discussions; A. Launay, A. Baron, R. Fernandes, D. Roux, T. Truong, J. Sher, T. Jagdish, K. Brinda, and N. Quiñones-Olvera for technical assistance. We also thank the Bauer Core Facility at Harvard University and the Service of Biochemistry at Bichat Hospital for providing some of the sequencing, and the O2 cluster from the Research Computing Group at Harvard Medical School for supporting some of the computational work. All data used in this analysis is free to access. **Funding:** M.B. acknowledges support from the NIGMS of the National Institutes of Health (R35GM133700), the David and Lucile Packard Foundation, the Pew Charitable Trusts, and the Alfred P. Sloan Foundation. This work was partially supported by the European Commission under the 7th Framework Program (ERC grant 310944 to O.T.), *Foundation pour le Recherche Médicale* (EQU201903007848) (to O.T.), *Agence Nationale pour la Recherche* ANR GeWiEp (ANR-18-CE35-0005-0) (to O.T.), and the Horizon 2020 Framework Programme (MSCA-IF 750129) (to A.C.). A.C. acknowledges support from a Comunidad de Madrid "Talento" Fellowship (2019-T1/BIO-12882) and the Agencia Estatal de Investigación (Proyectos de I+D+i, PID2019-110992GA-I00; Centros de Excelencia "Severo Ochoa", SEV-2016-0672 and CEX2020-000999-S). A.L. acknowledges support from the Molecules, Cells, and Organisms Graduate Program, Harvard University. R.E.L. acknowledges support from the US National Science Foundation (DEB-1951307) and the John Hannah endowment at Michigan State University. **Author contributions:** A.C., A.L., O.T., and M.B. conceived the

project. A.C., A.L., M.M., and S.V.O. designed and conducted experiments. A.C., A.L., C.M.H., R.E.L., O.T., and M.B. designed statistical analyses; A.C. and A.L. analyzed the data; R.E.L. directs the LTEE and provided strains and critical feedback on interpretation; all authors wrote and revised the manuscript.
Competing interests: The authors declare no competing interests.
Data and materials availability: Raw sequencing reads are available from the NCBI BioProject database (PRJNA814281 and PRJNA979973). Processed data are available from Zenodo (75, 76); source code for the sequencing pipeline, downstream analyses, and figure generation are available from GitHub (77, 78).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add1417
Materials and Methods
Supplementary Text
Figs. S1 to S25
Tables S1 to S8
References (79–97)
MDAR Reproducibility Checklist
Data S1 to S5

Submitted 25 May 2022; resubmitted 18 January 2023
Accepted 12 December 2023
10.1126/science.add1417

RESEARCH ARTICLE SUMMARY

CHEMISTRY AUTOMATION

Automated self-optimization, intensification, and scale-up of photocatalysis in flow

Aidan Slattery†, Zhenghui Wen†, Pauline Tenblad†, Jesús Sanjosé-Orduna, Diego Pintossi, Tim den Hartog, Timothy Noël*

INTRODUCTION: Photocatalysis exploits light for driving reactivity under mild conditions, contributing to advancements in synthetic methods for pharmaceuticals, agrochemicals, and materials. Nonetheless, challenges persist in optimizing, replicating, and scaling these techniques. These challenges stem from practical considerations such as uneven light absorption and experimental variability, alongside chemical complexities such as poorly understood reaction mechanisms and intricate interactions among various variables. These phases in advancing photocatalytic processes are crucial yet time-consuming components of contemporary chemical manufacturing, requiring expertise and precision owing to their intricate and sensitive nature.

RATIONALE: In response to the need for efficient optimization of complex photocatalytic reaction conditions, we have developed a robotic platform named RoboChem. RoboChem facilitates the self-optimization, intensification, and scale-up of photocatalytic transformations. By integrating readily available hardware, customized software, and a Bayesian optimization (BO) algorithm, this platform offers a hands-free and safe solution, mitigating associated challenges. Operating autonomously, RoboChem eliminates the requirement for extensive expertise in photocatalysis or scaling processes to achieve optimal results. This renders RoboChem a valuable collaborative robotic platform suitable for any synthetic organic chemistry laboratory, irrespective of users' specific familiarity with photocatalysis.

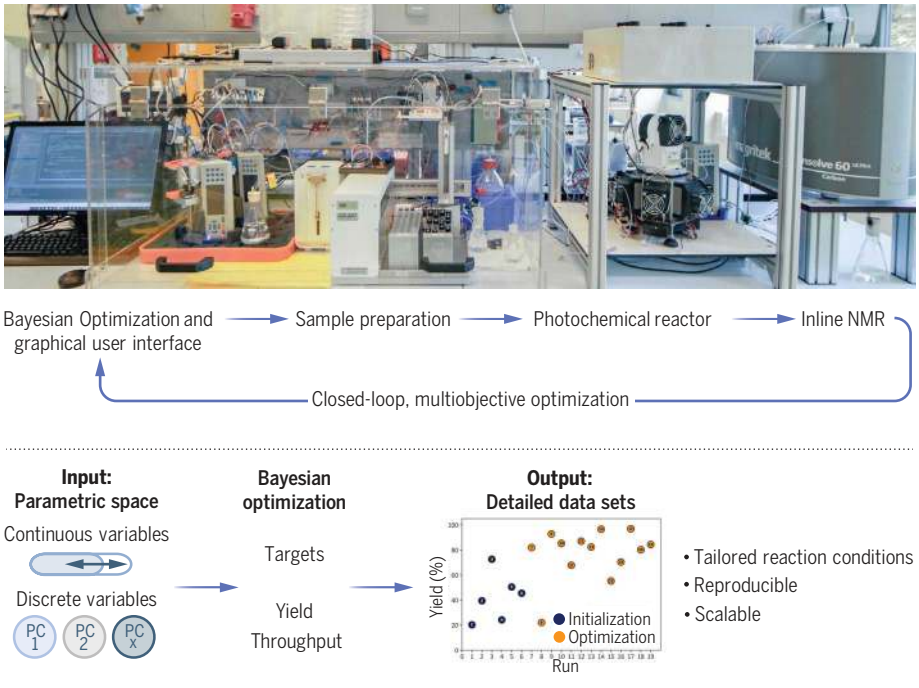
RESULTS: The robotic platform incorporates several key components, including a liquid handler, syringe pumps, a tunable continuous-flow photoreactor, cost-effective Internet of Things devices, and an in-line nuclear magnetic resonance (NMR) system. It uses a closed-loop BO approach to systematically explore a chosen parameter space encompassing both discrete and continuous variables. Consequently, the platform excels in identifying optimal reaction conditions that maximize either yield, throughput, or a combination thereof.

Operating within a continuous flow microreactor, the platform effectively addresses mass, heat, and photon transport considerations, resulting in the generation of well-structured datasets. These datasets capture both positive and negative results, thereby highlighting the influence of specific variables on the targeted objective function.

Furthermore, the optimal conditions identified by the platform have been successfully scaled up within the same continuous flow photoreactor. Manual isolation processes have been applied to obtain meaningful quantities of pure isolated compounds. Notably, the isolated yields closely align with the NMR yields obtained by the platform, validating its high precision and reliability.

The platform's capabilities were demonstrated across a diverse set of 19 molecules, covering various facets of photocatalysis, such as hydrogen atom transfer photocatalysis, photoredox catalysis, and metallaphotocatalysis. Notably, human involvement was limited to the definition of the parametric space, the preparation of stock solutions and the isolation of pure compounds. The effectiveness of the platform stems from its BO algorithm, which efficiently captures intricate interdependencies among different reaction variables. Consequently, the platform consistently identified optimal reaction conditions that either matched or exceeded those obtained through manual approaches. As a result, the RoboChem platform stands out from conventional synthetic methods by tailoring reaction conditions to the specific needs of each substrate. This capability enables a thorough assessment of the applicability and limitations of the reported transformations, ultimately enhancing their value for potential industrial implementation.

CONCLUSION: The RoboChem robotic platform expedites and streamlines the optimization of photocatalytic transformations, simultaneously enhancing safety and liberating researchers to focus on other creative facets of chemistry. ■



RoboChem: a benchtop robotic platform for the self-optimization, intensification and scale-up of photocatalytic transformations.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: t.noel@uva.nl
†These authors contributed equally.
Cite this article as A. Slattery *et al.*, *Science* **383**, eadj1817 (2024). DOI: 10.1126/science.adj1817

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adj1817

RESEARCH ARTICLE

CHEMISTRY AUTOMATION

Automated self-optimization, intensification, and scale-up of photocatalysis in flow

Aidan Slattery[1]†, Zhenghui Wen[1]†, Pauline Tenblad[1]†, Jesús Sanjosé-Orduna[1], Diego Pintossi[1], Tim den Hartog[1,2,3], Timothy Noël[1]*

The optimization, intensification, and scale-up of photochemical processes constitute a particular challenge in a manufacturing environment geared primarily toward thermal chemistry. In this work, we present a versatile flow-based robotic platform to address these challenges through the integration of readily available hardware and custom software. Our open-source platform combines a liquid handler, syringe pumps, a tunable continuous-flow photoreactor, inexpensive Internet of Things devices, and an in-line benchtop nuclear magnetic resonance spectrometer to enable automated, data-rich optimization with a closed-loop Bayesian optimization strategy. A user-friendly graphical interface allows chemists without programming or machine learning expertise to easily monitor, analyze, and improve photocatalytic reactions with respect to both continuous and discrete variables. The system's effectiveness was demonstrated by increasing overall reaction yields and improving space-time yields compared with those of previously reported processes.

Photocatalysis has greatly advanced synthetic methods by leveraging a range of distinct mechanistic pathways, such as single electron transfer, energy transfer, and hydrogen atom transfer (HAT) (*1*). Its inherently mild nature allows for seamless integration with other catalytic processes, facilitating distinct transformations achievable only through the synergistic action of multiple catalysts (*2*). Despite these advancements, the field still grapples with substantial hurdles in optimization, replication, and scalability of these methods (Fig. 1) (*3*).

These difficulties partly arise from the chemical complexity of photocatalysis, involving poorly understood reaction mechanisms and limited understanding of the photophysics underpinning the observed reactivity (*4*). Additionally, the complex synergistic interactions between different reaction variables often go unnoticed in traditional academic optimization strategies (*5*). When a promising reaction is initially identified, the focus shifts to refining conditions for that specific substrate by using the "one-factor-at-a-time" (OFAT) method. This approach entails systematically adjusting individual variables, such as ligands, bases, solvents, or in rare cases, light intensity, retaining the best result before proceeding to the next. Although design-of-experiments (DoE) strategies are increasingly adopted in industrial settings, applying them to each substrate within a scope is resource intensive. To save time, conditions optimized for a benchmark substrate are often generically applied to others, leading to suboptimal yields or selectivity. This is because each substrate has distinct molecular characteristics, such as differing steric and electronic properties and the presence or absence of sensitive functional groups, all of which substantially influence the reaction's outcome.

Another challenge in developing photocatalytic methods lies in the variability of experimental setups, leading to substantial batch-to-batch inconsistencies and limited scalability (*6*). In photocatalysis, photons act as central reactants, meaning that the reaction rate and stability of reagents closely depend on light intensity. According to the Lambert-Beer law, light intensity diminishes rapidly as it travels through a photocatalytic reaction mixture. Therefore, traditional batch scale-up strategies, which simply enlarge reactor dimensions, are ineffective, as major portions of the reactor receive insufficient light. Furthermore, uneven light distribution can cause irreproducibility, extended reaction times, and unwanted byproduct formation. Flow reactors, integrated with high-power light sources, have thus become crucial for effectively scaling up photocatalytic transformations, even in industrial settings (*7*, *8*). These reactors guarantee uniform high light intensity across the entire reactor cross-section, enhancing reaction kinetics and reducing reaction times, a concept known as process intensification (*9*). However, variations in light sources and reactor geometry mean that even with flow reactor technology, reoptimization is often necessary to ensure compatibility with the specific photocatalytic transformation.

In response to the challenge of rapidly optimizing complex photocatalytic reaction conditions, we sought to develop a multipurpose robotic platform, called RoboChem, that enables the self-optimization, intensification, and scale-up of photocatalytic transformations (Fig. 1). This platform overcomes associated challenges by integrating off-the-shelf hardware and customized software, providing a hands-off solution. Our open-source platform combines a liquid handler, syringe pumps, a high-powered photoreactor, inexpensive Internet of Things (IoT) devices, and an in-line benchtop nuclear magnetic resonance (NMR) system to enable automated and data-rich optimization. By using a continuous-flow capillary photoreactor, our platform ensures highly reproducible data collection, effectively mitigating issues related to mass, heat, and photon transport that often contribute to irreproducibility in photocatalytic transformations (*10*, *11*).

Further, to account for complex intercorrelations between reaction variables, optimization algorithms such as DoE and statistical modeling can be integrated into the platform. However, for complex nonlinear relationships such as those encountered in photocatalytic reactions, machine learning proves to be a more effective approach (*12*). Its capacity to rapidly and efficiently analyze vast amounts of data enables the identification of underlying patterns and the extraction of meaningful conclusions (*13*). Thus, combining machine learning with reaction automation is advantageous (*14*). Given that our platform operates as a linear system (i.e., not parallelized), minimizing the number of experiments required to reach optimal conditions was crucial. For this reason, we turned to Bayesian optimization (BO), which has gained popularity in the chemistry community owing to its capacity to optimize black-box functions (*15*–*17*).

As an automated flow chemistry setup, our platform is capable of exploring large regions of the experimental and chemical space within a relatively short period, making it well suited to address complex optimization problems encountered in photocatalytic reaction scope elaboration. The RoboChem platform distinguishes itself from common synthetic method practices by tailoring the reaction conditions to the specific requirements of every substrate, thereby enabling a clear evaluation of the applicability and limitations of the reported transformations, resulting in increased value for industrial implementation.

Because the capillary photoreactor is equipped with high-power LEDs, of which the light intensity can be adjusted to meet specific photochemical needs, this setup enables the production of

[1]Flow Chemistry Group, van 't Hoff Institute for Molecular Sciences (HIMS), University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands. [2]Zuyd University of Applied Sciences, Nieuw Eyckholt 300, 6419 DJ Heerlen, Netherlands. [3]Netherlands Organisation for Applied Scientific Research (TNO), High Tech Campus 25, 5656 AE Eindhoven, Netherlands.
*Corresponding author. Email: t.noel@uva.nl
†These authors contributed equally.

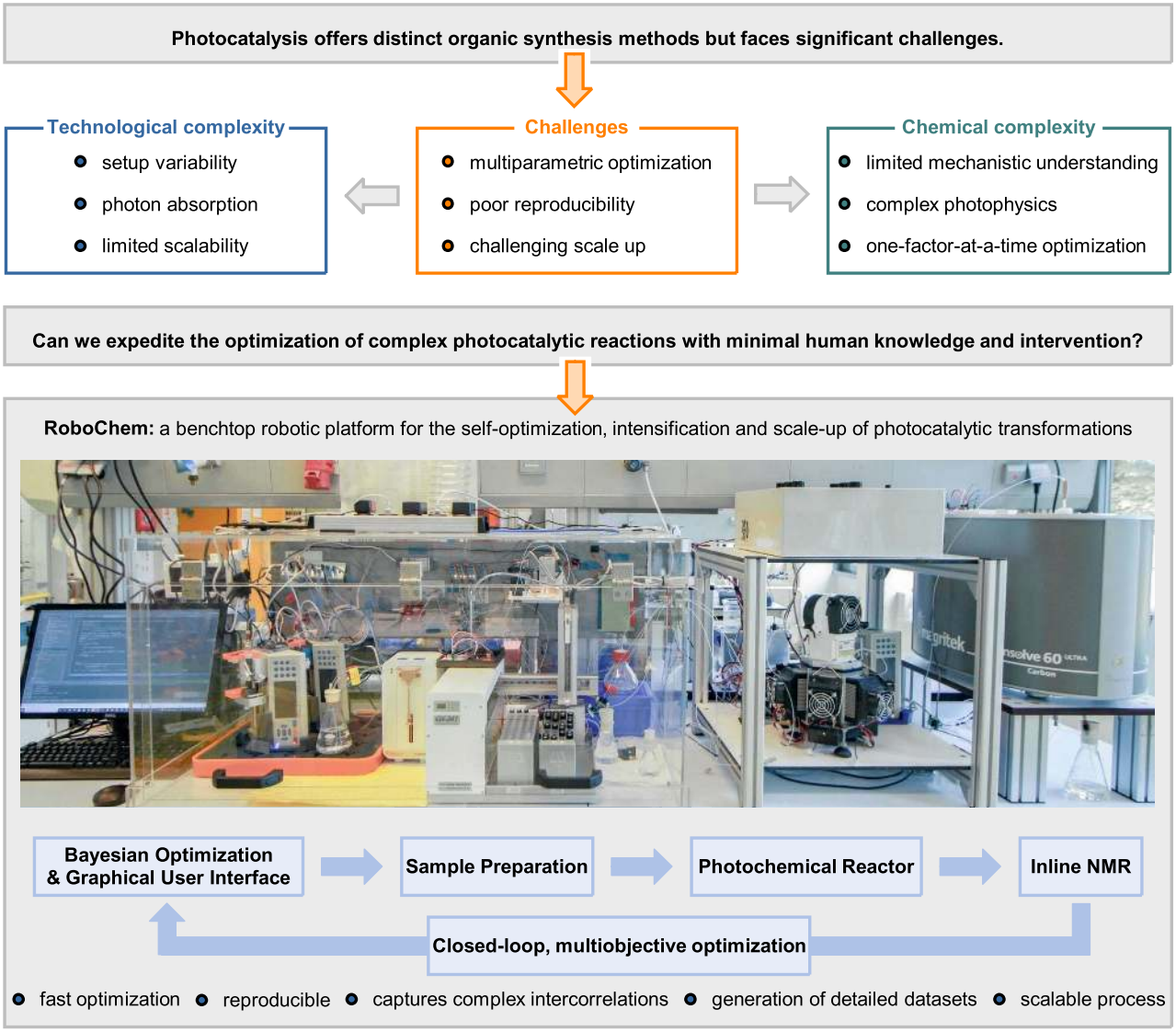


Fig. 1. RoboChem: a benchtop platform for the closed-loop, multiobjective optimization of photocatalytic systems. Shown are the challenges associated with optimization, replication, and scalability of photocatalysis, as well as the robotic platform and its workflow.

materials in substantial quantities. Consequently, results obtained from small-scale experiments can be seamlessly scaled up in the same reactor to produce tens to hundreds of grams of material per day, bridging the gap between laboratory research and practical application. Operating entirely autonomously, the platform further eliminates the need for in-depth expertise in photocatalysis or scaling processes to achieve optimal outcomes. This makes RoboChem an effective collaborative robotic platform, suitable for use in any synthetic organic chemistry laboratory, regardless of the users' specific knowledge in photocatalysis. In this work, we demonstrate the general applicability of RoboChem to the optimization of a diverse set of photocatalytic transformations, including hydrogen atom transfer (HAT) photocatalysis, photoredox catalysis, and metallaphotocatalysis, which are relevant to medicinal and crop protection chemistry.

RoboChem platform

The RoboChem platform can be divided into three distinct workflows: the controller, the planner, and the user input (Fig. 2A). The hardware controller guides the physical platform, encompassing tasks such as preparing the reaction mixture, executing the experiment, and conducting subsequent in-line analysis. The planner, which is a machine learning model, is responsible for determining the optimal experiments to run by selecting parameters and communicating them to the controller. The results are then fed back to the machine learning model, which subsequently recommends the next experiment. Last, the graphical user interface (GUI) allows users to input the necessary parameters, launch the optimization campaign, and initiate the process.

Platform – Controller

RoboChem is controlled by custom Python code and uses open-source libraries (Fig. 2A) with off-the-shelf instruments and devices. By coupling a liquid handler, syringe pumps, switching valves, a high-power continuous-flow photoreactor, as well as simple IoT devices such as phase sensors and ultrasonic detectors with an in-line 60-MHz benchtop NMR spectrometer for data-rich optimization, we have come up with a workflow to easily and efficiently optimize and intensify

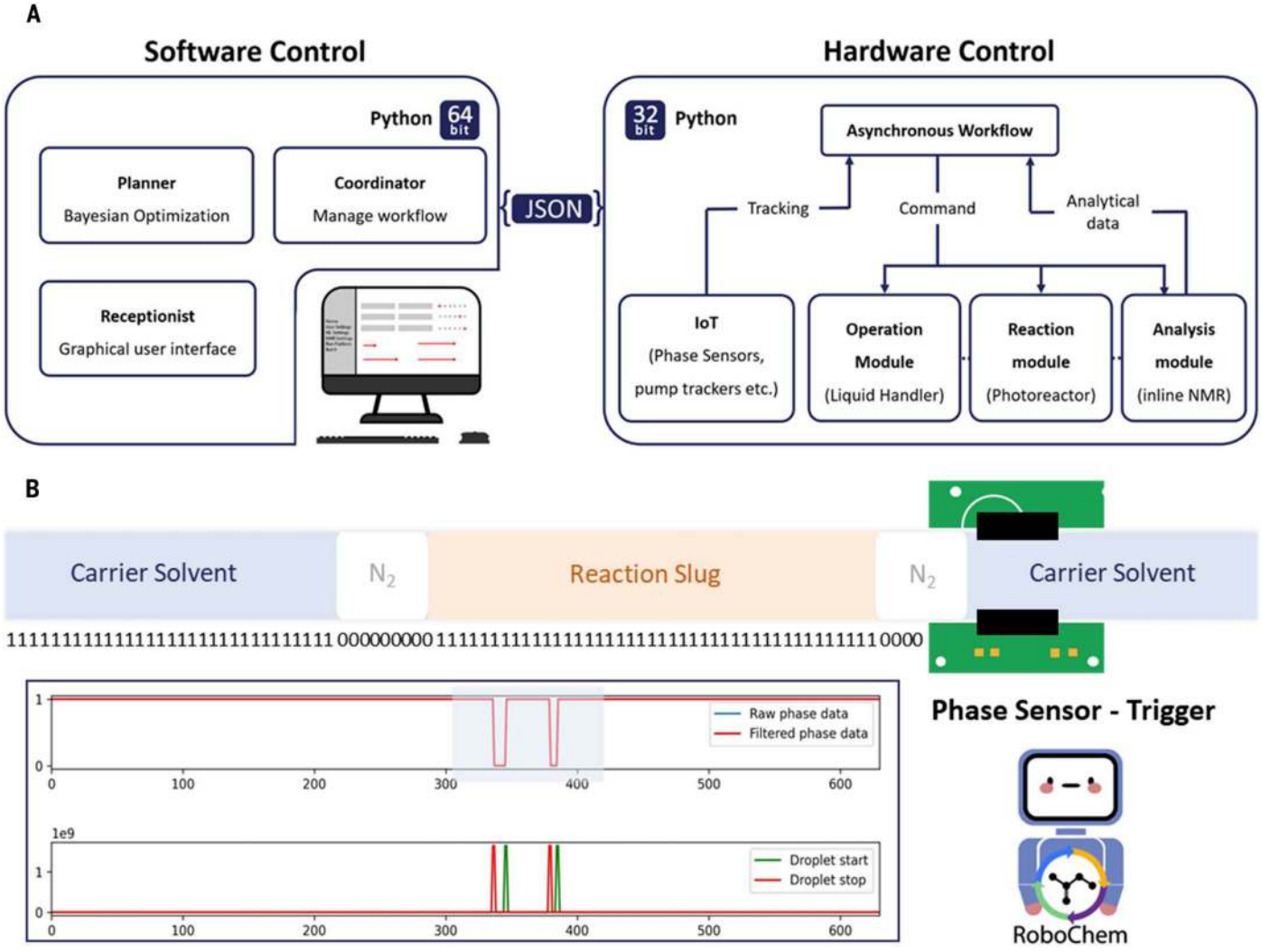


Fig. 2. Automated robotic platform for the self-optimization, intensification, and scale-up of photochemistry in flow. (**A**) High-level view of platform architecture. (**B**) Reaction tracking on the platform carried out by phase sensors to allow for timely triggering of events. A reaction slug is tracked as it passes over a phase sensor and investigated by an algorithm to form a trigger for the next phase of the optimization cycle.

photochemical processes. Each generated reaction slug (650 μL) represents a discrete set of reaction conditions (*18*, *19*), and the reactions are executed sequentially: sample preparation, reaction under the specified conditions, and automated analysis and processing. We used NMR for data analysis, which accurately integrates yields without requiring prior calibration of the analytics by using an analytically pure sample. This approach enabled us to conduct most reactions without the need for an internal standard, opting instead for an externally calibrated standard with a peak in a similar parts-per-million range as the target molecule. Although a 60-MHz benchtop NMR spectrometer was selected as the analytical technique, the platform is easily adaptable to accommodate other analytical techniques, such as Raman spectroscopy, infared spectroscopy, high-performance liquid chromatography (HPLC)–ultraviolet (UV) visible light, HPLC–mass spectrometry (MS), or gas chromatography–MS. Because the volume of the reaction slug is determined by the analytical method, this parameter could be substantially reduced by an order of magnitude if an alternative technique like HPLC were to be used. (*20*)

The RoboChem platform operates in a closed-loop manner, driven by the BO algorithm. This iterative process involves the BO algorithm proposing experiments, which are automatically executed and analyzed. The obtained results are then fed back into the BO algorithm, which generates a new set of conditions for further optimization (*21*, *22*). By harnessing the capabilities of a photomicroreactor equipped with high-intensity LEDs (*23*) and the ability to make computer-controlled power output adjustments, we can minimize reaction times and substantially enhance the throughput of the platform. This increased efficiency reduces the time required for a comprehensive optimization run.

By using a series of phase sensors and a dedicated algorithm to detect the passage of a reaction slug (*24*), the platform can efficiently track its position (Fig. 2B). This cost-effective approach enables precise control over the reaction as it traverses the system. The ability to accurately monitor the reaction's location eliminates the need to hardcode pump volumes, allowing for seamless compatibility with reactors of varying sizes without requiring any code modifications. Moreover, this tracking capability facilitates precise "parking" of the reaction slug in the NMR system for analysis, optimizing reagent usage by minimizing the required quantity for each reaction condition.

The entire system is conveniently located on a standard laboratory benchtop and is enclosed within a custom-designed, closed suction box, eliminating the need for placement within a fume hood during reaction runs. The system's design facilitates three distinct operating modes:

(i) Single experiment: Conducting a reaction under specific conditions, whether for the purpose of yield or productivity discovery, or as part of a scope entry.

(ii) Self optimization: Automating the optimization process for a single reaction or multiple reactions consecutively.

(iii) Scale-up: Exploiting the optimized conditions obtained through self-optimization for efficient scaling up of the reaction.

Bayesian optimization – Planner

The platform leverages BO, a machine learning–based approach, to optimize chemical reactions. BO is a probabilistic model-based method designed to efficiently identify the maximum (or minimum) of an unknown black-box function (*25*). It constructs a probability model of the function using carefully selected samples, which guides the search process by suggesting the next point to evaluate. The BO model incorporates both exploitation and exploration strategies. Exploitation involves investigating areas predicted to have the highest value, whereas exploration focuses on exploring points where the model has limited knowledge. This dual approach prevents the model from becoming trapped in local maxima or minima. The iterative process continues with the model being updated after each new evaluation of the function until a predetermined threshold is reached or a specific number of experiments have been conducted.

The BO model was implemented with the open-source Python package Dragonfly, developed by Kandasamy and collaborators (*22*, *26*). The initial runs are chosen by using Latin-hypercube sampling (*27*, *28*). The researchers define the input variables (parameters to be tuned) and the objective to be optimized. The platform supports both single-objective and multiobjective optimization, targeting yield and/or throughput. In single-objective optimization, the model identifies the global maximum of the reaction. In multiobjective optimization, the model finds a set of nondominated solutions known as the Pareto front (*29*). To assess the progress of the optimization problem, the platform tracks the hypervolume after each run (*30*). In cases of interrupted runs or the desire to build upon previously executed experiments, the platform allows for further optimization from that point (for further details, see the supplementary materials).

The platform's integration of machine learning effectively reduces the reliance on human resources (*31*, *32*). Once the experiments are set up and the optimization process is initiated, the platform operates independently. The machine learning model autonomously determines the next set of experiments to run, and the corresponding commands are automatically transmitted to the platform. As a result, the platform can run continuously, including overnight, freeing up the chemist to focus on other tasks. Throughout this study, human intervention was mainly needed for replenishing solutions. Notably, we successfully prevented clogging issues by using a compact reaction slug that efficiently contained any minor precipitates. After this reaction phase, we maintained a continuous flow of a carrier solvent, such as acetonitrile or DMSO, to remove any remaining precipitates. It is important to note that while we did not encounter clogging in our experiments, the possibility of clogging cannot be ruled out if a flow-incompatible chemical space is chosen.

Graphical user interface – User input

A key aspect of the platform's design is the development of an intuitive graphical user interface (GUI) that enables chemists without programming or machine learning expertise to easily navigate the system. The GUI provides functionality for creating new experiments, storing all the settings and results for an optimization run. It also allows users to generate the required positional and sample data used by the platform and liquid handler to prepare reaction slugs.

The liquid handler can accommodate multiple stock solutions of the same type. As the platform consumes stock solutions, it automatically tracks the quantity used. When one stock solution is depleted, the platform seamlessly transitions to the next vial containing a remaining stock solution. The GUI defines the entire chemical space to be explored under the Machine Learning Settings page.

In the Run Platform tab, a button initiates the platform, and the GUI continuously tracks the results. For single-objective optimizations, the GUI presents a chart displaying the objective function (yield or throughput) against the number of runs. In multiobjective optimizations, it provides a plot of yield versus throughput and includes a graph tracking the hypervolume.

The GUI performs validations to ensure all necessary files and chemical spaces are properly defined before allowing the platform to run. If multiple reagents are added to a subcategory, the GUI automatically treats them as discrete variables for the optimization process.

After each run, the data consisting of both the input parameters and the output values are automatically stored in a JSON file. These JSONs have been converted to Microsoft Excel (.xlsx) files for easier data manipulation and are available on Zenodo as .xlsx files (*33*) and in the supplementary materials in table format, in line with the FAIR principles for scientific data archiving (*34*). These datasets have the potential to be used in future projects; the data are of high quality owing to the absence of mass, heat, and photon transport issues. Given that all experiments are conducted on the same platform (including scale-up), experimental error is significantly reduced, ensuring that NMR-based yield estimations closely align with those achieved after isolation on a 1- to 5-mmol scale. Another advantage is the accumulation of data for negative results, which are not commonly published but nevertheless are important for the development of machine learning models (*35*).

The primary focus of the RoboChem platform is to identify optimal reaction conditions for photocatalytic transformations. Our versatile platform caters to both single- and multiobjective optimization problems, offering synthetic chemists the ability to maximize yield, productivity, and other relevant objective functions. To accomplish this, we selected five distinct photocatalytic reactions, covering a total of 19 substrates, for optimization. For each case, we compared the yield and productivity reported in the literature with the conditions determined by the artificial intelligence (AI)–assisted RoboChem platform. The reaction conditions discovered by the AI were subsequently used to scale up the transformations.

Single-objective optimization for photocatalytic HAT alkylation

We began our testing and validation of the platform with a Giese-type reaction entailing photocatalytic HAT activation of hydrocarbons (*36*). The reaction was conducted in the flow photomicroreactor, using tunable 0- to 144-W, 365-nm–emitting light-emitting diodes (LEDs). This choice allowed us to evaluate robust and well-established chemistry in our laboratory (Fig. 3). Five optimization variables were selected for the reaction [benzylidenemalononitrile concentration, tetrahydrofuran (THF) loading, photocatalyst (TBADT) loading, light intensity, and residence time]. A total of 19 experiments were conducted in a closed-loop fashion continuously for 4 hours. The initiation phase involved six experiments, serving as a preliminary scan of the reaction space. Subsequently, the BO algorithm recommended one new experimental condition at a time for a further 13 experiments, aiming to maximize the objective function [yield (%)]. Within nine experiments, the platform achieved a yield of more than 90% and began converging on the optimal conditions for the chemistry, resulting in a yield exceeding 95% for the desired product. Notably, the reaction manifested a detrimental effect of high light intensity, with the optimal range found to be between 20 and 50% of full power (28 to 72 W

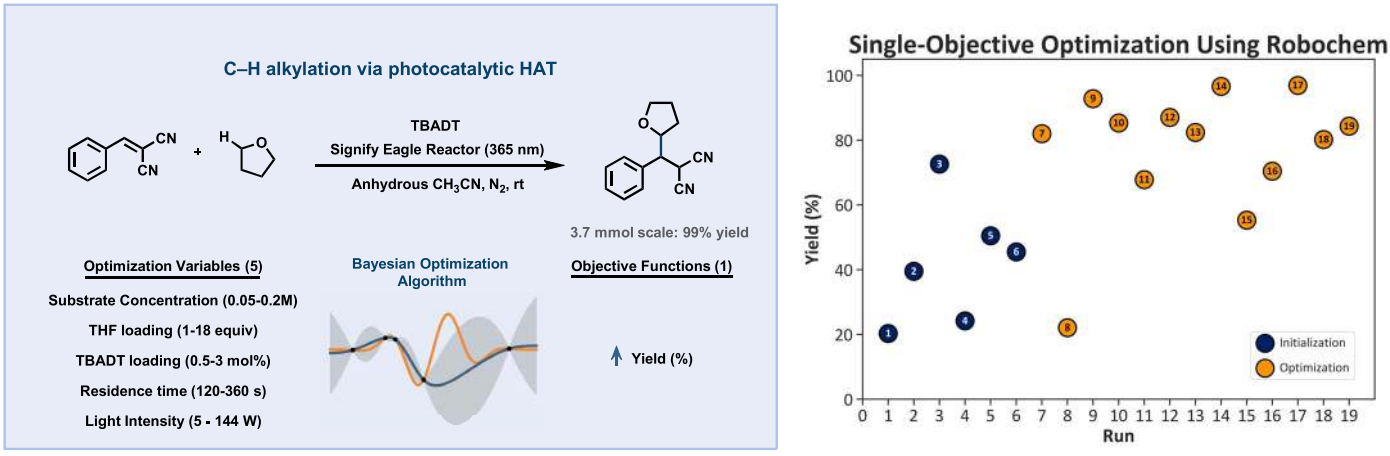


Fig. 3. Single-objective optimization of the photocatalytic HAT-alkylation of benzylidenemalononitrile with tetrahydrofuran (THF). rt, room temperature.

optical input power). These optimal conditions were then used for scaling up the transformation with the same capillary photoreactor, confirming the AI-determined yield with an isolated yield of 99% (3.7-mmol scale).

Single- and multiobjective optimization of C–H trifluoromethylthiolation of C(sp^3)–H and C(sp^2)–H bonds through decatungstate-enabled HAT

Having validated the automated AI-driven photochemical platform in a single objective optimization problem, we aimed next to investigate its capability for optimizing various photocatalytic processes in multiobjective fashion, seeking to simultaneously optimize yield and throughput. Consequently, the reaction conditions found by the AI model would be readily suitable for subsequent scale-up. As an initial benchmark, we selected the decatungstate-mediated trifluoromethylthiolation of C(sp^3)–H and C(sp^2)–H bonds through HAT, as reported in (*37*). The incorporation of the trifluoromethylthiol group in drug-like molecules holds high value in medicinal chemistry, offering high lipophilicity (as indicated by the Hansch parameter of π_R = 1.44) and electronegativity, which enhances the pharmacokinetic properties and optimizes the interaction between the active compound and its target.

In the trifluoromethylthiolation campaign (Fig. 4), five reaction parameters and two objective functions were optimized simultaneously. The photochemistry was conducted in the continuous-flow photomicroreactor, which used perfluoroalkoxy (PFA) tubing with a 0.8-mm I.D. and a total volume of 2.85 mL. To provide the necessary light source, a chip-on-board (COB) UV LED system with a tunable light intensity ranging from 0 to 144 W of optical power was used. The screening chemical space encompassed five continuous parameters: N-(trifluoromethylthio)phthalimide (Phth-SCF_3) concentration, H-donor equivalents, TBADT photocatalyst loading, residence time, and light intensity. The objective functions chosen for optimization were either yield (%) or simultaneously the yield (%) and throughput (mmol h^{-1}) of the SCF_3-bearing molecules. To ensure fair comparisons between different reactor systems, we chose to convert the productivities into space-time yield (STY) ($g{\cdot}L^{-1}{\cdot}h^{-1}$). This normalization factorizes the reactor volume, allowing for a more equitable assessment of performance across varying reactor sizes.

For each substrate, a total of 18 to 36 experiments were conducted within an 8- to 16-hour timeframe. This series comprised eight initialization experiments followed by refinement experiments for each optimization campaign until a sufficient yield or hypervolume was achieved. Notably, substantial yield improvements were observed compared with their respective model counterparts in batch reactions. The platform also demonstrated a remarkable increase in productivities, ranging from 70 to 100 times higher. Next, the reaction conditions selected by the AI model were successfully used for scale-up to 5 mmol. In all cases, the isolated yields obtained during the scale-up process closely matched the NMR yields observed with the AI-optimized reaction conditions.

Upon further analysis of the AI-discovered reaction conditions, several interesting observations emerged. The AI algorithm refined the reaction conditions to achieve optimal reactivity and selectivity for each specific substrate. In this context, the BO algorithm identified experimental conditions that deviated substantially from the standard conditions reported in (*37*). One notable finding was the substantial differences in reaction or residence time and light intensity, depending on the molecule being optimized. Comparing the results obtained for trifluoromethylthiolated Sclareolide (**5**) and Ambroxide (**6**), it is evident that the catalyst loading and light intensity are significantly lower for Ambroxide. This result can be rationalized by the fact that Ambroxide can undergo an additional reaction with another equivalent of Phth-SCF_3, resulting in a double-functionalized final product. However, such a reaction is not possible with Sclareolide, as the α-to-O carbon position is blocked by the carbonyl group. By reducing the catalyst loading and light intensity, the AI algorithm successfully enhanced the yield and selectivity of the monofunctionalized product (**6**).

Multiobjective optimization of oxytrifluoromethylation of alkenes by using photocatalytic single electron transfer

Next, we directed our focus to the oxytrifluoromethylation of alkenes through a three-component process using photocatalytic single electron transfer with $Ru(bpy)_3(PF_6)_2$, as reported by (*38*). In the oxytrifluoromethylation campaign (Fig. 5), we simultaneously optimized five reaction parameters: styrene concentration, CF_3 source loading, photocatalyst loading, residence time, and light intensity. Two objective functions [yield (%) and throughput (mmol h^{-1})] were targeted for optimization. Similarly, for each substrate, a total of 14 to 25 experiments were conducted within a 3- to 10-hour timeframe. The optimization process used rapid ^{19}F NMR analysis (2 min per measurement) for molecules **7**, **8**, **9**, and **11**. However, molecule **10** required a longer optimization time of 19 hours owing to the use of ^{1}H NMR for quantification. To ensure high accuracy, a 16-min analysis window was allocated per reaction. As previously described, our experimental procedure involved six initialization experiments, followed by refinement experiments for each optimization campaign.

Similar to previous experiments, the photochemistry was conducted in the flow photomicroreactor by using PFA tubing with a 0.8-mm I.D. However, for this campaign, chip-on-board blue LEDs with a tunable light intensity ranging from 0 to 188 W were manually installed to

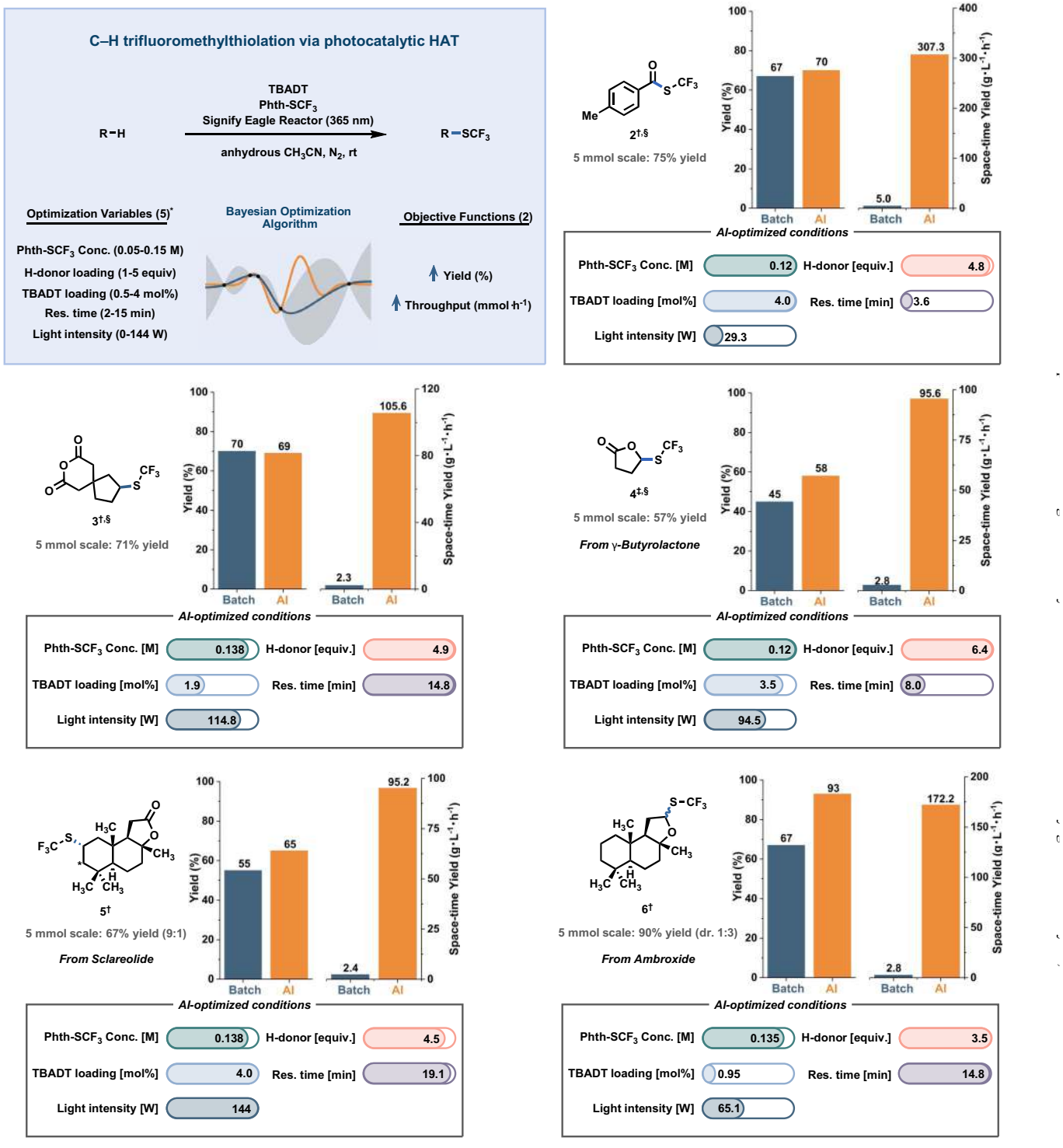


Fig. 4. Substrate scope and associated summary for the C–H trifluoromethylthiolation enabled by RoboChem. *Outer bounds of the chemical space explored for all experiments given (for more experimental details and exact chemical spaces explored for each experiment, see supplementary materials section S5.3). †Yield and space-time yield comparisons made directly from the literature. Conditions: PhthSCF_3 concentration (conc.) (0.2 M), H-donor equiv. [1.5 equiv. for aldehydes, 2.5 equiv. for substrates with activated C(sp^3)–H bonds, 5 equiv. for substrates with unactivated C(sp^3)–H bonds], TBADT loading (2.5 mol%), residence time (16 hours, 6 hours for aldehydes), light (385 nm, 1.2 W), N_2, 25°C. ‡Yield and space-time yield comparisons made to reactions carried out by us under literature conditions with homemade batch reactor with 40-W 370-nm Kessil lamp; yields were determined by quantitative NMR (supplementary materials section S5.1). §Experiments run as single objective with the objective function to increase yield. Res. time, residence time.

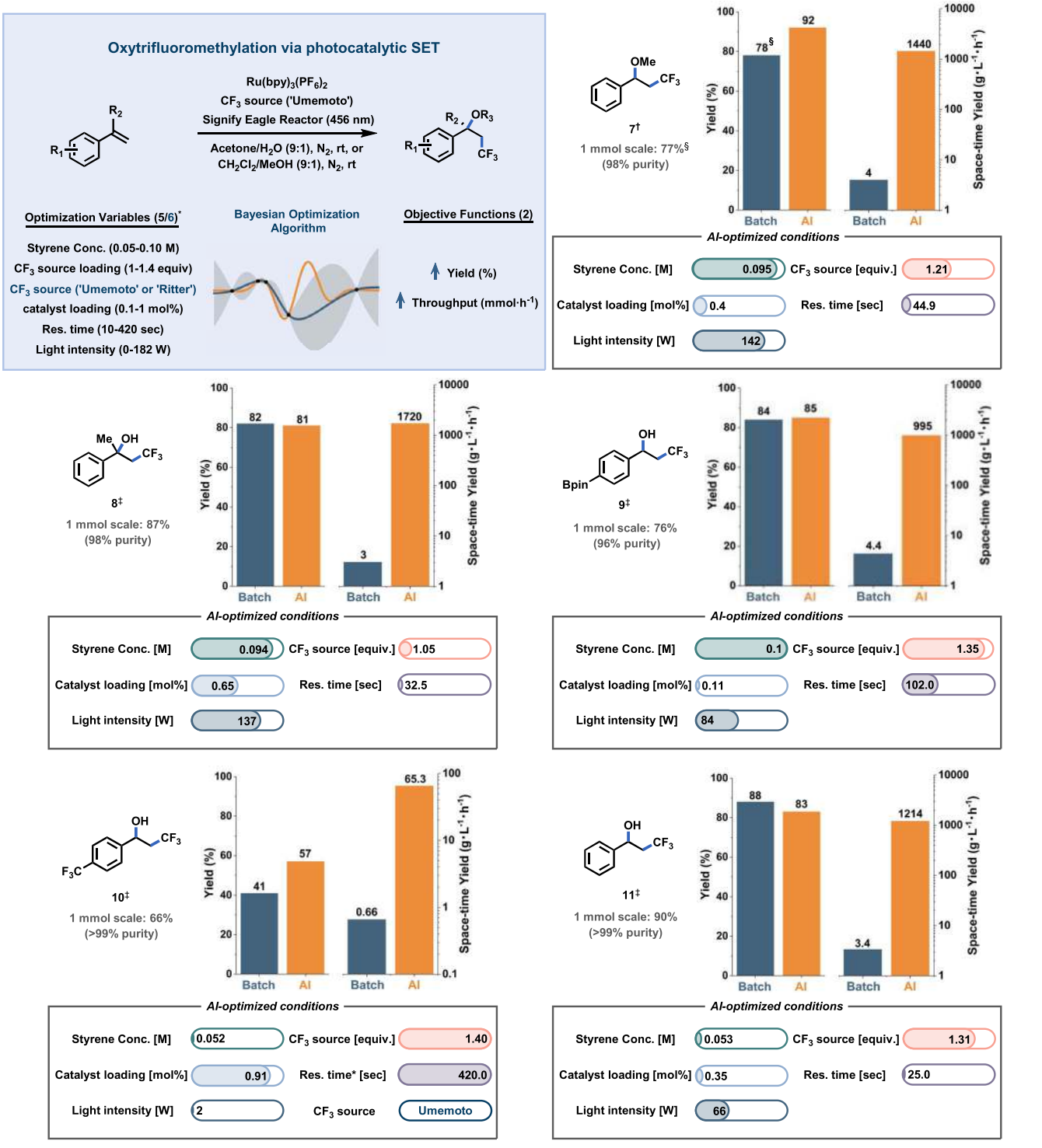


Fig. 5. Substrate scope and associated summary for the three-component oxytrifluoromethylation of alkenes enabled by RoboChem. *Outer bounds of the chemical space explored for all experiments given (for more experimental details and exact chemical spaces explored for each experiment, see supplementary materials section S6.2). †Yield and space-time yield comparisons made directly from the literature. Conditions: Styrene conc. (0.05 M), *fac*-Ir(ppy)$_3$ loading (0.5 mol%), Umemoto loading (1.05 equiv), CH_2Cl_2:MeOH (9:1), 425-nm blue LED (3 W), 2.5 hours. ‡Yield comparisons made directly from the literature. Conditions: Styrene conc. (0.05M), *fac*-Ir(ppy)$_3$ loading (0.5 mol%), Umemoto loading (1.1 equiv), acetone:H_2O (9:1), 425-nm blue LED (3 W), 2 to 4 hours. §Yield > 95% by quantitative NMR.

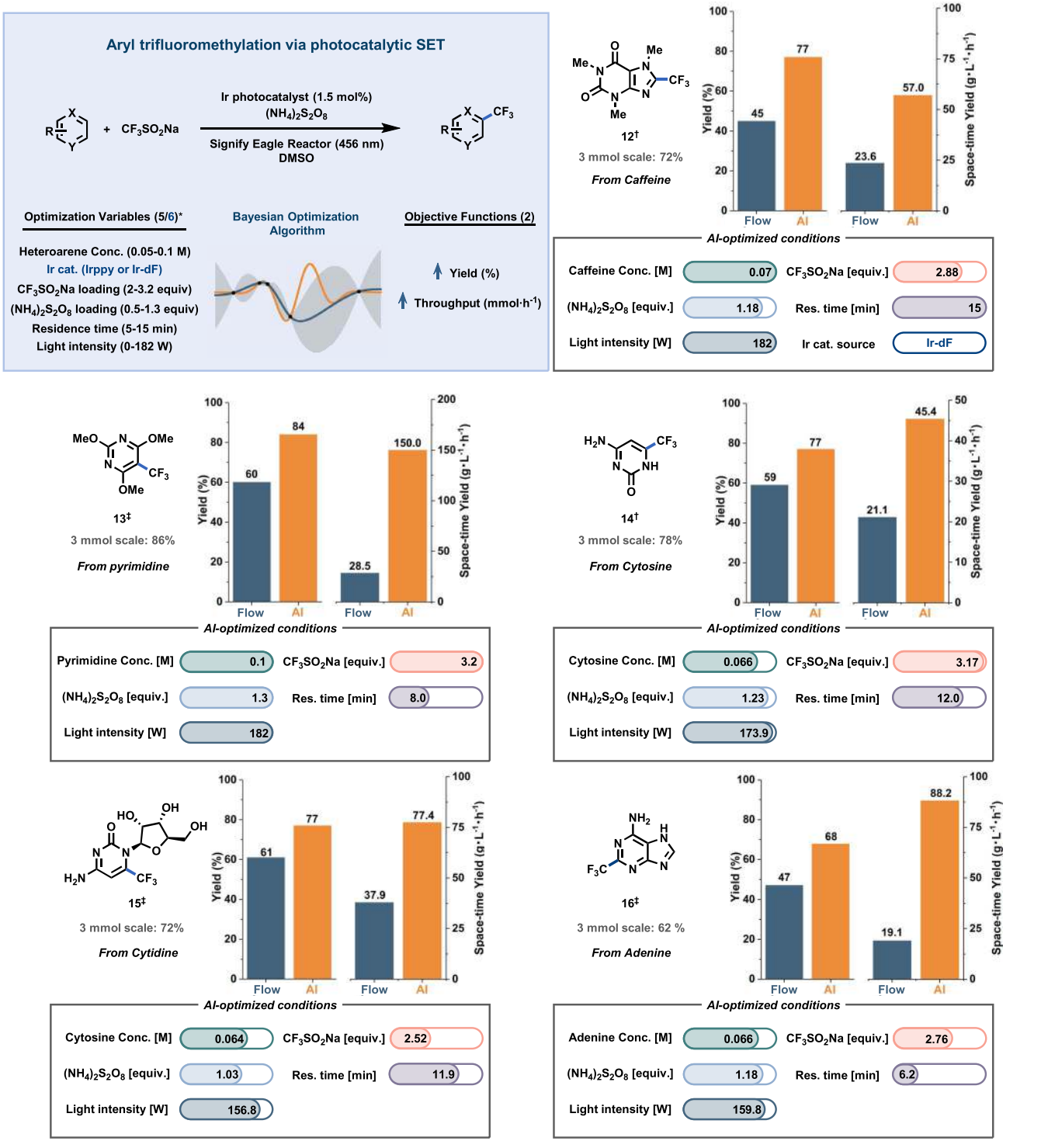


Fig. 6. Substrate scope and associated summary for aryl trifluoromethylation through single electron transfer enabled by RoboChem. *Outer bounds of the chemical space explored for all experiments given (for more experimental details and exact chemical spaces explored for each experiment, see supplementary materials section S7.3). †Yield comparisons made directly from the literature. Conditions: Heteroarene conc. (0.1 M), Ir-dF, {Ir[dF(CF_3)ppy]$_2$}(dtbpy)]PF_6 (1 mol%), CF_3SO_2Na equiv. (3 equiv.), $(NH_4)_2S_2O_8$ equiv. (1 equiv.), residence time (30 min), Vapourtec reactor (456 nm, 60 W). ‡Yield and space-time yield comparisons made to reaction carried out by us under literature conditions with Vapourtec Photoreactor (456 nm, 60 W); yields were determined by quantitative NMR (supplementary materials section S7.2).

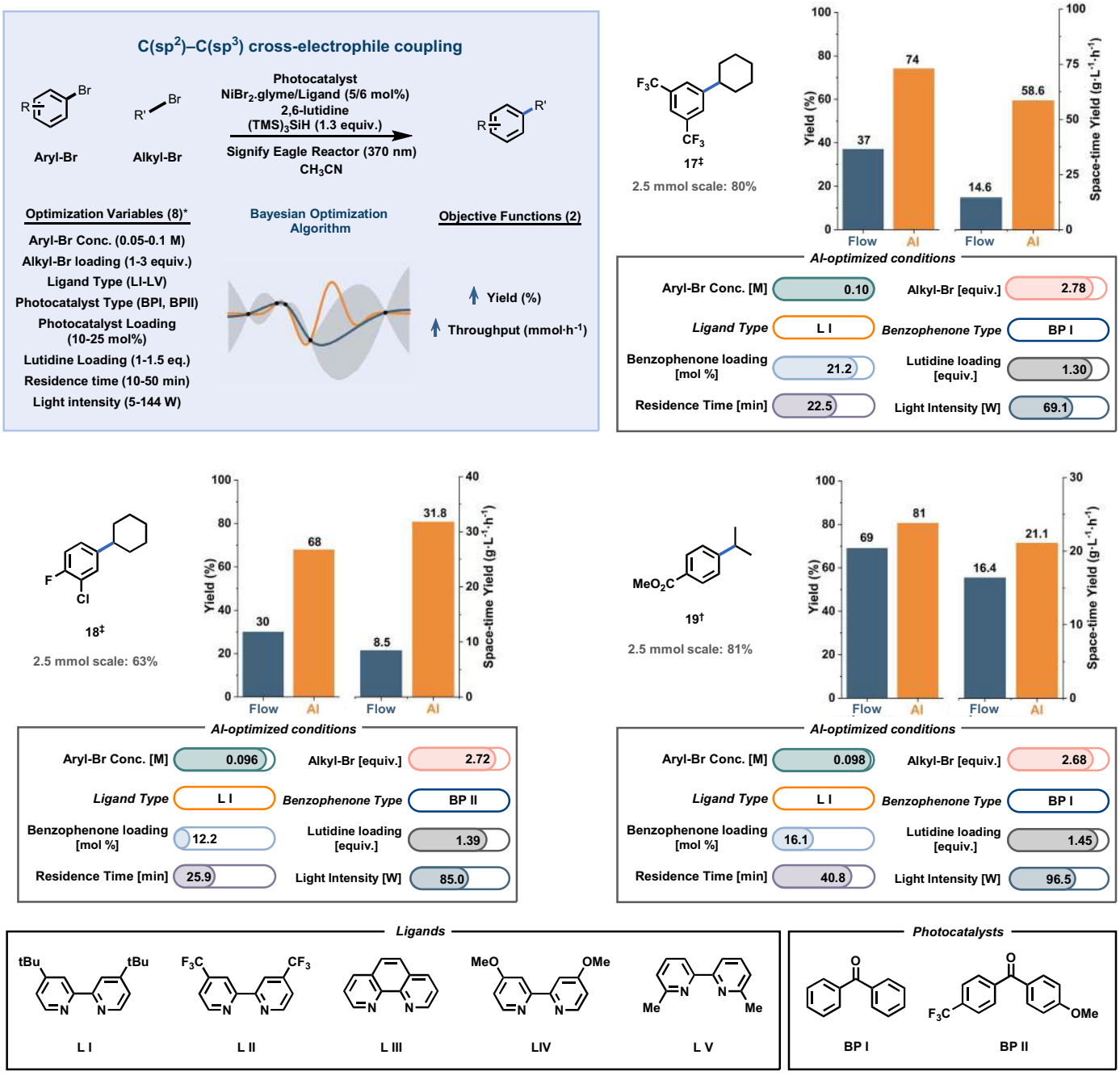


Fig. 7. Substrate scope and associated summary for C(sp^2)–C(sp^3) cross-electrophile coupling enabled by RoboChem. *Outer bounds of the chemical space explored for all experiments given (for more experimental details and exact chemical spaces explored for each experiment, see supplementary materials section S8.3). †Yield comparisons made directly from the literature. Conditions: Aryl bromide conc. (0.1 M), Alkyl bromide (2.5 equiv.), BP II (20 mol%), $NiBr_2$·dtbbp (5 mol%), 2,6-lutidine (1.1 equiv.), $(TMS)_3SiH$ (1.5 equiv.), residence time (45 min), Vapourtec reactor (456 nm, 16 W). ‡Yield and space-time yield comparisons made to reaction carried out by us in duplicate under literature conditions with Vapourtec UV 150 Photoreactor (456 nm, 16 W); yield was determined by quantitative NMR (supplementary materials section S8.2).

match the absorption maximum of the $Ru(bpy)_3$ photocatalyst. Owing to the short residence times, which were as low as 10 s, the internal volume of the photoreactor had to be reduced to 0.26 mL. This adjustment was necessary because at very low residence times (10 s) with a larger internal volume (>3 ml), the syringe pumps struggled to cope with the increased pressure drop across the platform.

The RoboChem platform successfully performed reaction optimization, resulting in conditions that produced outcomes closely aligned with those of the model batch reactions. A significant increase in space-time yield of up to 565-fold was achieved, demonstrating substantial potential for scale-up in the flow reactor. During the scale-up process, a slight improvement in yield was observed compared with that of the optimization carried out on the platform. This can be attributed to the fact that, for scale-up, the internal volume of the reactor was multiplied by a factor of six, whereas the residence time remained the same. Consequently, an associated sixfold increase in flow rate was necessary to maintain the desired residence time, leading to improved mass transfer facilitated by a higher Reynolds number. This

phenomenon accounts for the observed increase in yield compared with that of the platform conditions.

The results indicate a significant dependence of this chemistry on the sustained power applied during the reaction, with higher-power or longer residence time conditions resulting in noticeably lower yields due to photon-induced product degradation. Remarkably, one reaction condition exhibited optimal performance at the lowest "turned-on" power output; specifically, molecule **9** at 2-W optical output. Notably, the choice of the lowest light intensity, particularly when contrasted with the high light intensity used in other examples, further highlights the challenging nature of prediction in this context. Additionally, during the optimization of molecule **9**, various CF_3 sources were screened, including trifluoromethyl thianthrenium triflate (*39*) and Umemoto's reagent, serving as a test case to evaluate discrete variables. The algorithm determined that Umemoto's reagent was the optimal choice for this transformation.

Multiobjective optimization of aryl trifluoromethylation

To provide another example, our objective was to optimize the visible-light photocatalytic trifluoromethylation of highly functionalized heteroarenes developed by our group and researchers from Janssen pharmaceuticals (Fig. 6) (*40*). In our original report, the reaction was carried out in a commercially available Vapourtec UV-150 flow reactor. In the flow photomicroreactor equipped with blue LEDs, we scanned a search space consisting of five reaction parameters (heteroarene concentration, CF_3SO_2Na loading, oxidant loading, residence time, and light intensity), targeting two objective functions [yield (%) and throughput (mmol h^{-1})] for optimization. During the optimization of caffeine trifluoromethylation, we also incorporated a categorical variable to screen for the appropriate photocatalyst. This highlights the capability of the RoboChem platform to evaluate and optimize both discrete and continuous variables. For each substrate, a total of 17 to 35 experiments (including six initialization steps) were conducted within an 11- to 24-hour timeframe.

In this specific example, the RoboChem platform focused on optimizing a diverse range of densely functionalized substrates that hold major interest in drug discovery programs (Fig. 6). Despite the original work being conducted in a flow system, we observed a substantial enhancement in both yield and productivity. This improvement can be attributed to the platform's capacity to tailor the reaction conditions to each substrate individually, coupled with the use of a more potent light source in the flow photomicroreactor. It is well-recognized that increasing the light intensity can inherently influence performance, particularly in the realm of photochemistry (*41*). However, if this transition were the primary factor in enhanced performance, rather than the AI optimization, we would expect to see more uniformity in the optimized conditions across all literature-to-RoboChem transitions. Contrary to this, our findings demonstrate notable variability in these conditions, suggesting a significant and distinct contribution from AI refinement.

Multiobjective optimization of $C(sp^2)$–$C(sp^3)$ cross-electrophile coupling

In our final case study, we focused on optimizing complex photocatalytic transformations that use synergistic catalytic cycles, such as metallaphotocatalysis. The union between photocatalysis and transition-metal catalysis offers distinct reactivity, facilitating otherwise difficult bond formations (*42*). We prioritized carbon-carbon bond formation because of its importance in the pharmaceutical and agrochemical sectors. Specifically, we studied the cross-electrophile coupling of alkyl and aryl bromides to achieve $C(sp^2)$–$C(sp^3)$ bond formation (*43*). Mechanistic insights revealed that the reaction requires a combination of benzophenone HAT photocatalysis, silyl radical–induced halogen atom transfer, and nickel-catalyzed cross-coupling. (*44*, *45*) Given the vast potential chemical space and the delicate balance between the three catalytic cycles, we believed that identifying the optimum would be a serious hurdle for chemists, making it an apt test for the RoboChem platform. In our optimization campaign (Fig. 7), eight reaction parameters were concurrently adjusted: aryl halide concentration, alkyl halide loading, selection among five nickel ligand sources (**L I** to **L V**), choice between two benzophenone photocatalysts (**BP I** and **BP II**), benzophenone and 2,6-lutidine loadings, residence time, and light intensity. We optimized two objective functions: yield (%) and throughput (mmol h^{-1}). For every substrate, either 45 or 60 experiments were conducted, with total optimization durations ranging from 41 to 58 hours. Each experimental phase comprised an initial 20 runs, followed by 25 or 40 runs for optimization.

Initially, we sought to enhance the yield of a substrate that underperformed under literature conditions (*45*). After 60 experiments spanning 58 hours with RoboChem, we successfully elevated the yield of compound **17** from 37 to 77%. Analysis of the data pinpointed the choice of ligand as an important variable. **L V** was suboptimal, likely owing to the steric interference of its two methyl groups, highlighting the sensitivity of the nickel catalytic cycle toward steric hindrance. **L II**, **L III**, and **L IV** performed adequately but were outperformed by **L I**. This observation was swiftly recognized by the algorithm, which selected **L I** in 31 of the 40 optimization runs. **BP I** emerged as the preferred HAT catalyst, despite underwhelming results under literature conditions when compared with **BP II** (*45*). Moreover, higher light intensities, considered detrimental in literature model reactions (*45*), proved beneficial.

Subsequently, we undertook optimization for compounds **18** and **19**. Both endeavors yielded valuable datasets with both positive and negative results (detailed in supplementary materials), which led ultimately to tailored reaction conditions, including the selection of HAT catalyst **BP II** for compound **18**, that substantially improved yields. This result underscores the potential pitfalls of optimizing OFAT on a model substrate, which may inadvertently narrow the chemistry to that molecule, limiting broader applicability. As in prior case studies, our NMR yields closely mirrored isolated yields, underscoring our platform's precision and reproducibility. This also highlights the benefit of using flow technology to control mass, heat, and photon transport across varied reaction scales (*6*), an advantage not available with analogous automated batch reactor systems.

Conclusions and outlook

As shown in this work, RoboChem constitutes a versatile robotic platform for the self-optimization, intensification, and/or scale-up of a diverse set of photocatalytic reactions. Operated through a BO algorithm, this platform is able to explore the presented parameter space, ultimately furnishing customized reaction conditions attuned to the specific needs of each substrate. By substantially reducing the need for human intervention, RoboChem not only increases operational safety but also liberates researchers to dedicate more time to the more creative aspects of chemistry, thereby freeing them from the drudgery of reaction optimization and intensification tasks.

The modularity of the robotic platform is an asset of our design, and we foresee its integration with different types of flow reactors and process analytical technologies in the future. Moreover, by individually optimizing reaction parameters and generating datasets that include both optimal and suboptimal conditions, intricate relationships between the targeted reaction parameters, the substrate structures, and the objective functions can be uncovered. The ability to automatically generate rich datasets, obtained within a highly reproducible reactor environment, can contribute to the future digitization of synthetic chemistry.

Methods summary

For a more detailed description of the preparation and running of the platform, and

generation of stock solutions, see the supplementary materials. A short summary is provided below.

Analytics preparation

A calibration curve was established for the benchtop NMR by using a reference standard featuring a molecule within a comparable parts-per-million range as the peaks of interest in the product (e.g., α,α,α-trifluorotoluene for the photocatalytic oxytrifluoromethylation and the aryl trifluoromethylation). These calibration measurements were conducted under the exact NMR conditions as those used during the campaign. Subsequently, a scripting file was generated to automate the processing of the product's NMR spectrum. Additionally, the NMR instrument underwent shimming procedures in preparation for the campaign's commencement.

User input

The experiment was then initialized by using the GUI to input the settings for the optimization campaign (chemical space, stock solution concentrations, NMR settings, scripting file, etc.).

Platform operation

Stock solutions were prepared according to the procedures detailed in the supplementary materials. The solutions were subsequently transferred to 4-ml glass vials and then loaded into the liquid handler. From the GUI, the system was then run. The liquid handler makes up the reaction solution which was automatically introduced into the continuous-flow photochemical reactor. Upon exiting the reactor, the reaction slug was analyzed by benchtop NMR, and subsequently, the objective function (yield or throughput) was calculated. This information was sent back to the BO algorithm, which defines the next set of conditions. Every campaign kicks off with a Latin Hypercube sampling initialization, after which optimization toward maximizing the objective function happens. The optimization cycle was stopped by the operator after the hypervolume appeared to reach a maximum or when the total amount of runs indicated at the outset was realized. During the optimization campaign, the remaining amount of stock solutions and carrier solvent were checked periodically and refilled if necessary. If the stock solutions are not refilled in time, the platform will automatically stop the sequence until refilling.

REFERENCES AND NOTES

1. M. H. Shaw, J. Twilton, D. W. C. MacMillan, Photoredox Catalysis in Organic Chemistry. *J. Org. Chem.* **81**, 6898–6926 (2016). doi: 10.1021/acs.joc.6b01449; pmid: 27477076
2. K. L. Skubi, T. R. Blum, T. P. Yoon, Dual Catalysis Strategies in Photochemical Synthesis. *Chem. Rev.* **116**, 10035–10074 (2016). doi: 10.1021/acs.chemrev.6b00018; pmid: 27109441
3. T. Noël, E. Zysman-Colman, The promise and pitfalls of photocatalysis for organic synthesis. *Chem Catal.* **2**, 468–476 (2022). doi: 10.1016/j.checat.2021.12.015
4. D. M. Arias-Rotondo, J. K. McCusker, The photophysics of photoredox catalysis: A roadmap for catalyst design. *Chem. Soc. Rev.* **45**, 5803–5820 (2016). doi: 10.1039/C6CS00526H; pmid: 27711624
5. C. J. Taylor *et al*., A Brief Introduction to Chemical Reaction Optimization. *Chem. Rev.* **123**, 3089–3126 (2023). doi: 10.1021/acs.chemrev.2c00798; pmid: 36820880
6. S. D. A. Zondag, D. Mazzarella, T. Noël, Scale-Up of Photochemical Reactions: Transitioning from Lab Scale to Industrial Production. *Annu. Rev. Chem. Biomol. Eng.* **14**, 283–300 (2023). doi: 10.1146/annurev-chembioeng-101121-074313; pmid: 36913716
7. K. C. Harper *et al*., Commercial-Scale Visible Light Trifluoromethylation of 2-Chlorothiophenol Using CF_3I Gas. *Org. Process Res. Dev.* **26**, 404–412 (2022). doi: 10.1021/acs.oprd.1c00436
8. C. Bottecchia *et al*., Manufacturing Process Development for Belzutifan, Part 2: A Continuous Flow Visible-Light-Induced Benzylic Bromination. *Org. Process Res. Dev.* **26**, 516–524 (2022). doi: 10.1021/acs.oprd.1c00240
9. T. Van Gerven, G. Mul, J. Moulijn, A. Stankiewicz, A review of intensification of photocatalytic processes. *Chem. Eng. Process.* **46**, 781–789 (2007). doi: 10.1016/j.cep.2007.05.012
10. L. Buglioni, F. Raymenants, A. Slattery, S. D. A. Zondag, T. Noël, Technological Innovations in Photochemistry for Organic Synthesis: Flow Chemistry, High-Throughput Experimentation, Scale-up, and Photoelectrochemistry. *Chem. Rev.* **122**, 2752–2906 (2022). doi: 10.1021/acs.chemrev.1c00332; pmid: 34375082
11. D. Cambié, C. Bottecchia, N. J. W. Straathof, V. Hessel, T. Noël, Applications of Continuous-Flow Photochemistry in Organic Synthesis, Material Science, and Water Treatment. *Chem. Rev.* **116**, 10276–10341 (2016). doi: 10.1021/acs.chemrev.5b00707; pmid: 26935706
12. J. Freiesleben, J. Keim, M. Grutsch, Machine learning and Design of Experiments: Alternative approaches or complementary methodologies for quality improvement? *Qual. Reliab. Eng. Int.* **36**, 1837–1848 (2020). doi: 10.1002/qre.2579
13. M. I. Jordan, T. M. Mitchell, Machine learning: Trends, perspectives, and prospects. *Science* **349**, 255–260 (2015). doi: 10.1126/science.aaa8415; pmid: 26185243
14. C. Houben, A. A. Lapkin, Automatic discovery and optimization of chemical processes. *Curr. Opin. Chem. Eng.* **9**, 1–7 (2015). doi: 10.1016/j.coche.2015.07.001
15. F. Häse, L. M. Roch, A. Aspuru-Guzik, Chimera: Enabling hierarchy based multi-objective optimization for self-driving laboratories. *Chem. Sci.* **9**, 7642–7655 (2018). doi: 10.1039/C8SC02239A; pmid: 30393525
16. A. D. Clayton *et al*., Bayesian Self-Optimization for Telescoped Continuous Flow Synthesis. *Angew. Chem. Int. Ed.* **62**, e202214511 (2023). doi: 10.1002/anie.202214511; pmid: 36346840
17. B. J. Shields *et al*., Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**, 89–96 (2021). doi: 10.1038/s41586-021-03213-y; pmid: 33536653
18. D. Perera *et al*., A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science* **359**, 429–434 (2018). doi: 10.1126/science.aap9112; pmid: 29371464
19. H. W. Hsieh, C. W. Coley, L. M. Baumgartner, K. F. Jensen, R. I. Robinson, Photoredox Iridium-Nickel Dual-Catalyzed Decarboxylative Arylation Cross-Coupling: From Batch to Continuous Flow via Self-Optimizing Segmented Flow Reactor. *Org. Process Res. Dev.* **22**, 542–550 (2018). doi: 10.1021/acs.oprd.8b00018
20. C. Avila *et al*., Automated stopped-flow library synthesis for rapid optimisation and machine learning directed experimentation. *Chem. Sci.* **13**, 12087–12099 (2022). doi: 10.1039/D2SC03016K; pmid: 36349112
21. C. P. Breen, A. M. K. Nambiar, T. F. Jamison, K. F. Jensen, Ready, Set, Flow! Automated Continuous Synthesis and Optimization. *Trends Chem.* **3**, 373–386 (2021). doi: 10.1016/j.trechm.2021.02.005
22. K. Kandasamy *et al*., Tuning Hyperparameters without Grad Students: Scalable and Robust Bayesian Optimisation with Dragonfly. *J. Mach. Learn. Res.* **21**, 1–27 (2020).
23. T. Wan *et al*., Accelerated and Scalable C(sp^3)-H Amination via Decatungstate Photocatalysis Using a Flow Photoreactor Equipped with High-Intensity LEDs. *ACS Cent. Sci.* **8**, 51–56 (2022). doi: 10.1021/acscentsci.1c01109; pmid: 35106372
24. Noel Research Group, Robochem, v1, Zenodo (2023); https://doi.org/10.5281/zenodo.10261472.
25. B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, N. de Freitas, Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* **104**, 148–175 (2016). doi: 10.1109/JPROC.2015.2494218
26. A. M. K. Nambiar *et al*., Bayesian Optimization of Computer-Proposed Multistep Synthetic Routes on an Automated Robotic Flow Platform. *ACS Cent. Sci.* **8**, 825–836 (2022). doi: 10.1021/acscentsci.2c00207; pmid: 35756374
27. W. Li, L. Lu, X. Xie, M. Yang, A novel extension algorithm for optimized Latin hypercube sampling. *J. Stat. Comput. Simul.* **87**, 2549–2559 (2017). doi: 10.1080/00949655.2017.1340475
28. M. D. McKay, R. J. Beckman, W. J. Conover, Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* **21**, 239–245 (1979).
29. A. D. Clayton *et al*., Algorithms for the self-optimisation of chemical reactions. *React. Chem. Eng.* **4**, 1545–1554 (2019). doi: 10.1039/C9RE00209J
30. S. Daulton, M. Balandat, E. Bakshy, Differentiable Expected Hypervolume Improvement for Parallel Multi-Objective Bayesian Optimization. arXiv:2006.05078 [stat.ML] (2020).
31. S. Steiner *et al*., Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, eaav2211 (2019). doi: 10.1126/science.aav2211; pmid: 30498165
32. S. V. Ley, D. E. Fitzpatrick, R. J. Ingham, R. M. Myers, Organic synthesis: March of the machines. *Angew. Chem. Int. Ed.* **54**, 3449–3464 (2015). doi: 10.1002/anie.201410744; pmid: 25586940
33. Noel Research Group, Robochem Datasets, v1, Zenodo (2023); https://doi.org/10.5281/zenodo.10262296.
34. M. D. Wilkinson *et al*., The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016). doi: 10.1038/sdata.2016.18; pmid: 26978244
35. M. P. Maloney *et al*., Negative Data in Data Sets for Machine Learning Training. *J. Org. Chem.* **88**, 5239–5241 (2023). doi: 10.1021/acs.joc.3c00844; pmid: 37126471
36. Z. Wen *et al*., Optimization of a Decatungstate-Catalyzed C(sp^3)-H Alkylation Using a Continuous Oscillatory Millistructured Photoreactor. *Org. Process Res. Dev.* **24**, 2356–2361 (2020). doi: 10.1021/acs.oprd.0c00235; pmid: 33100815
37. T. E. Schirmer, A. B. Rolka, T. A. Karl, F. Holzhausen, B. König, Photocatalytic C-H Trifluoromethylthiolation by the Decatungstate Anion. *Org. Lett.* **23**, 5729–5733 (2021). doi: 10.1021/acs.orglett.1c01870; pmid: 34260256
38. Y. Yasu, T. Koike, M. Akita, Three-component oxytrifluoromethylation of alkenes: Highly efficient and regioselective difunctionalization of C=C bonds mediated by photoredox catalysts. *Angew. Chem. Int. Ed.* **51**, 9567–9571 (2012). doi: 10.1002/anie.201205071; pmid: 22936394
39. H. Jia, A. P. Häring, F. Berger, L. Zhang, T. Ritter, Trifluoromethyl Thianthrenium Triflate: A Readily Available Trifluoromethylating Reagent with Formal CF_3^+, $CF_3^•$, and CF_3^- Reactivity. *J. Am. Chem. Soc.* **143**, 7623–7628 (2021). doi: 10.1021/jacs.1c02606; pmid: 33985330
40. I. Abdiaj, C. Bottecchia, J. Alcazar, T. Noël, Visible-Light-Induced Trifluoromethylation of Highly Functionalized Arenes and Heteroarenes in Continuous Flow. *Synthesis (Stuttg.)* **49**, 4978–4985 (2017). doi: 10.1055/s-0036-1588527
41. T. D. Svejstrup *et al*., Effects of Light Intensity and Reaction Temperature on Photoreactions in Commercial Photoreactors. *ChemPhotoChem* **5**, 808–814 (2021). doi: 10.1002/cptc.202100059
42. A. Y. Chan *et al*., Metallaphotoredox: The Merger of Photoredox and Transition Metal Catalysis. *Chem. Rev.* **122**, 1485–1542 (2022). doi: 10.1021/acs.chemrev.1c00383; pmid: 34793128
43. D. A. Everson, D. J. Weix, Cross-electrophile coupling: Principles of reactivity and selectivity. *J. Org. Chem.* **79**, 4793–4798 (2014). doi: 10.1021/jo500507s; pmid: 24820397
44. P. Zhang, C. C. Le, D. W. C. MacMillan, Silyl Radical Activation of Alkyl Halides in Metallaphotoredox Catalysis: A Unique Pathway for Cross-Electrophile Coupling. *J. Am. Chem. Soc.* **138**, 8084–8087 (2016). doi: 10.1021/jacs.6b04818; pmid: 27263662
45. A. Luridiana *et al*., The Merger of Benzophenone HAT Photocatalysis and Silyl Radical-Induced XAT Enables Both Nickel-Catalyzed Cross-Electrophile Coupling and 1,2-Dicarbofunctionalization of Olefins. *ACS Catal.* **12**, 11216–11225 (2022). doi: 10.1021/acscatal.2c03805; pmid: 36158902

ACKNOWLEDGMENTS

Funding: We gratefully acknowledge the generous funding provided by the Dutch Research Council (NWO) under the Open Technology Program (Multi-Modal Photochemistry, no. 18433) (A.S., D.P., and T.N.). Z.W. was supported by the China Scholarship Council (CSC, no. 201808440313). Furthermore, we extend our gratitude for the generous funding received from the European Union H2020 research and innovation program under the

European Research Council program (FlowHAT, no. 101044355) (T.N.), Marie S. Curie Grant Agreement (PhotoReAct, no. 956324) (P.T. and T.N.), and MSCA Individual Fellowship program (SELECTFLOW, no. 101061835) (J.S.O.). **Author contributions:** A.S., Z.W., and D.P. developed, assembled and maintained the hardware of the robotic platform; A.S., Z.W., P.T., and D.P. developed the code of the controller that steered the robotic platform; P.T. implemented and assessed the BO algorithm; A.S., Z.W., J.S.O., T.d.H., and T.N. selected and supervised the chemistry to be run on the robotic platform; A.S., Z.W., and J.S.O. made the stock solutions, evaluated the reaction progress, and carried out the product isolation; P.T. implemented the GUI; A.S., P.T., and T.N. drafted the manuscript with input from all authors; and T.N. conceived the idea for this work and secured the funding.
Competing interests: The authors declare no competing interests.
Data and materials availability: All the code for the automated platform is available at https://github.com/Noel-Research-Group/Robochem_v1.0. An archived version of the code is available on Zenodo (*24*). Tabulated data are also available on Zenodo (*33*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj1817
Materials and Methods
Figs. S1 to S120
Tables S1 to S57
NMR Spectra
References (*46*–*50*)

Submitted 11 June 2023; resubmitted 12 October 2023
Accepted 13 December 2023
10.1126/science.adj1817

RESEARCH ARTICLE

QUANTUM OPTICS

Observation of millihertz-level cooperative Lamb shifts in an optical atomic clock

Ross B. Hutson[1,2]*, William R. Milner[1,2], Lingfeng Yan[1,2], Jun Ye[1,2]*, Christian Sanner[3]

Collective couplings of atomic dipoles to a shared electromagnetic environment produce a wide range of many-body phenomena. We report on the direct observation of resonant electric dipole-dipole interactions in a cubic array of atoms in the many-excitation limit. The interactions produce spatially dependent cooperative Lamb shifts when spectroscopically interrogating the millihertz-wide optical clock transition in strontium-87. We show that the ensemble-averaged shifts can be suppressed below the level of evaluated systematic uncertainties for optical atomic clocks. Additionally, we demonstrate that excitation of the atomic dipoles near a Bragg angle can enhance these effects by nearly an order of magnitude compared with nonresonant geometries. Our work demonstrates a platform for precise studies of the quantum many-body physics of spins with long-range interactions mediated by propagating photons.

Studies of quantum many-body physics naturally arise in the context of quantum sensing. For any quantum sensor, the amount of extractable information regarding a metrological quantity of interest is fundamentally limited by the number of accessible qubits (*1*, *2*). This creates a generic incentive to build devices capable of manipulating and characterizing quantum systems of ever-increasing size (*3*–*5*). Because interactions within the system or with the environment typically scale with system size, the main challenges are then twofold: How can interactions be controlled to reduce systematic effects, and/or how can they be leveraged to generate useful entanglement?

In the context of atomic clocks, notable progress toward probing larger numbers of atoms while avoiding systematic effects resulting from contact interactions has been made by trapping atoms in three-dimensional (3D) optical lattices with at most one atom per lattice site (*6*–*9*). Nonetheless, clock shifts caused by long-range resonant dipole-dipole interactions have loomed just beyond experimental detectability (*10*–*12*).

The classical electric field, evaluated at a position $\mathbf{b}$, generated by a point dipole $\mathbf{d_a} \propto e^{-i\omega t}$, oscillating at an angular frequency ω, and localized at a position $\mathbf{a}$, is given by $\mathbf{E_a}(\mathbf{b}) = k^3 e^{ikr}\{[\mathbf{d_a} - \hat{\mathbf{r}}(\hat{\mathbf{r}}\cdot\mathbf{d_a})]/(kr) + [3\hat{\mathbf{r}}(\hat{\mathbf{r}}\cdot\mathbf{d_a}) - \mathbf{d_a}][1/(kr)^3 - i/(kr)^2]\}/4\pi\varepsilon_0$, where $\mathbf{r} = \mathbf{r_{ba}} = \mathbf{b} - \mathbf{a}$, $r = |\mathbf{r}|$, $\hat{\mathbf{r}} = \mathbf{r}/r$, ε_0 is the vacuum permittivity, and $k = \omega/c$ with c being the speed of light (*13*). A second, freely oscillating dipole $\mathbf{d_b}$ localized at position $\mathbf{b}$ will then dynamically evolve according to the interaction term $H_{\mathbf{ba}} = -\mathbf{d_b}\cdot\mathbf{E_a}(\mathbf{b})$, whose real and imaginary parts, respectively, lead to a frequency shift and damping of the initial excitation. These interactions form the basis of classical linear optics (*13*–*15*).

An ensemble of indistinguishable (pseudo–)spin-½ systems, with internal ground and excited states labeled $|g\rangle$ and $|e\rangle$, respectively, can analogously be described using the formalism of quantum optics, where the reduced density matrix $\hat{\rho}$ evolves in time t according to the master equation $\partial_t\hat{\rho} = \mathcal{L}_{\text{free}}[\hat{\rho}] = \mathcal{L}_1[\hat{\rho}] + \mathcal{L}_2[\hat{\rho}]$ with the Liouvillian superoperator describing collective electromagnetic interactions given in Lindblad form as (*16*)

$$\mathcal{L}_2[\hat{\rho}] = -i\sum_{\mathbf{a},\mathbf{b}} V_{\mathbf{ba}}\left(\hat{S}^\dagger_{\mathbf{b}}\hat{S}_{\mathbf{a}}\hat{\rho} - \hat{S}_{\mathbf{a}}\hat{\rho}\hat{S}^\dagger_{\mathbf{b}}\right) + \text{H.c.} \quad (1)$$

and generic single-spin dynamics governed by $\mathcal{L}_1[\hat{\rho}]$. Here, $\hat{S}^\dagger_{\mathbf{a}} = \zeta^*_{\mathbf{a}}|e\rangle_{\mathbf{a}}\langle g|_{\mathbf{a}}$ is the raising operator for the spin at $\mathbf{a}$, with $\zeta_{\mathbf{a}}$ being an arbitrary phase factor satisfying $|\zeta_{\mathbf{a}}|^2 = 1$. The coefficients $V_{\mathbf{ba}}$ of the effective Hamiltonian $\hbar\sum_{\mathbf{a},\mathbf{b}} V_{\mathbf{ba}}\hat{S}^\dagger_{\mathbf{b}}\hat{S}_{\mathbf{a}} + \text{H.c.}$, where $\hbar$ is the reduced Planck constant, are obtained by quantizing the dipole moments $\mathbf{d_a} \to \langle g|\hat{\mathbf{d}}|e\rangle\zeta_{\mathbf{a}}\hat{S}^\dagger_{\mathbf{a}} + \text{H.c.}$ appearing in the classical interaction terms $H_{\mathbf{ba}}$, applying the rotating wave approximation, and negating the homogeneous self-interaction energy (Lamb shift) $\text{Re}(V_{\mathbf{aa}}) \to 0$. The characteristic energy scale of $\mathcal{L}_2$ is set by the excited-state spontaneous decay rate $\Gamma = -2\text{Im}(V_{\mathbf{aa}})$, which is on the order of $2\pi \times 1$ mHz for the clock transition in neutral strontium-87.

Equation 1 has long been known to contain the physics of cooperative decay (*17*) and cooperative Lamb shifts (*18*, *19*), with these effects being subsequently observed in a wide variety of physical systems (*20*–*38*). However, in the context of optical frequency metrology, resonant dipole-dipole interactions are typically neglected because of the relatively low atomic densities and weak transition strengths.

[1]JILA, NIST, and University of Colorado, Boulder, CO 80309, USA. [2]Department of Physics, University of Colorado, Boulder, CO 80309, USA. [3]Department of Physics, Colorado State University, Fort Collins, CO 80523, USA.
*Corresponding author. Email: ross.hutson@colorado.edu (R.B.H.); jun.ye@colorado.edu (J.Y.)

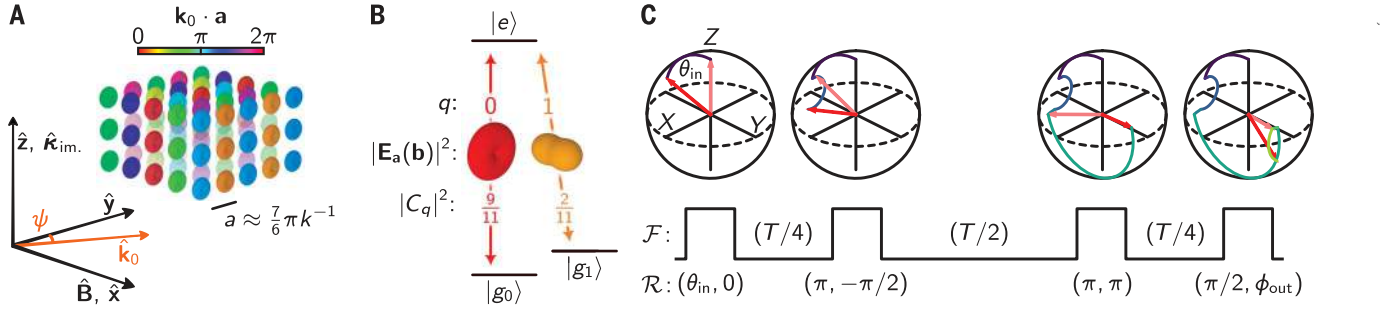


Fig. 1. Experimental setup. (**A**) Atomic dipoles on a cubic lattice, indexed by their positions $\mathbf{a}$, are excited with spatially dependent phases $\mathbf{k}_0\cdot\mathbf{a}$. Proximity to the Bragg condition $\mathbf{k}_0\cdot a\hat{\mathbf{y}} = \pi$ leads to long-range phase ordering along the $\hat{\mathbf{y}}$ axis. (**B**) Far-field ($kr_{\mathbf{ba}} \gg 1$) single-atom radiation patterns $|\mathbf{E_a}(\mathbf{b})|^2$ for the two ($q \in \{0,1\}$) spectroscopically resolved $|e\rangle \leftrightarrow |g_q\rangle$ transitions. (**C**) Resonant laser pulses $\mathcal{R}(\theta,\phi)$ rotate the atomic states by an angle θ about the $(X_{\mathbf{a}}\cos\phi + Y_{\mathbf{a}}\sin\phi)$ axis. The various pulses and free-evolution periods $\mathcal{F}(t)$ are chosen such that the output-state projection $\langle\hat{Z}_{\mathbf{a}}\mathcal{U}\rangle$ is proportional only to terms in $\mathcal{L}_{\text{free}}$ that scale antisymmetrically with $\cos\theta_{\text{in}}$, namely those caused by the resonant dipole-dipole interactions in $\mathcal{L}_2$.

Experimental approach for observing cooperative Lamb shifts

We have observed that such interactions cannot always be neglected in an atomic clock. We performed Ramsey spectroscopy on a quantum-degenerate Fermi gas of neutral strontium-87 in a cubic optical lattice with a total free-evolution period of $T = 2$ s. Subsequently imaging the Ramsey interferometer output, the observed population differences served as a proxy for spatially inhomogeneous "clock shifts" of the atomic resonance. Our model, based on the resonant dipole-dipole interactions contained in Eq. 1, accurately reproduces the observed shifts over a wide range of external control parameters.

The spatial distribution of the interferometer signal is shown to be dependent on the relative drive phases and radiation patterns of the atomic dipoles, which we control by varying the angle of incidence of the probe laser and polarization q of the probe photons, respectively. Additionally, the shifts are shown to scale with the strengths of the atomic dipole moments through changes in the pulse area of the initial interferometer pulse and the relative strength of the probed transition. Despite the presence of a relatively large technical dephasing rate $\gamma \approx 2\pi \times 34$ mHz $\gg \Gamma$ caused by Raman scattering of optical lattice photons (*39*), we relied on the precision of the atomic clock to divide a Ramsey fringe by more than a part in 10^3 to measure millihertz-level cooperative Lamb shifts in the limit $\Gamma T \ll 1$ over a wide range of initial excitation fractions.

Viewed in the context of modern quantum optics (*10*, *40*–*46*), periodic arrays of atomic dipoles are a promising platform for studies of many-body physics because they are thought to host a rich Hilbert space containing long-lived quasiparticles (*44*, *45*, *47*). Although cooperative Lamb shifts in the multiple-excitation limit have recently been observed, modeling of the observed shift was unsuccessful because decoherence mechanisms that significantly affect the dynamics at long interrogation times $\Gamma t \gg 1$ were not accounted for (*48*). Our work suggests that optical atomic clocks are natural platforms for studies of many-body quantum optics given that all parameters contained in $\mathcal{L}_{\text{free}}$ are systematically characterized, and coherent manipulations and projective measurements on timescales much shorter than Γ^{-1} are readily achievable.

Experimental overview and optical spectroscopy

As previously described (*49*, *50*) and schematically represented in Fig. 1A, in a shot-based experiment with a cycle time of 12.5 s, a single-component Fermi-degenerate gas of $N_{\text{tot}} \approx 10^4$ strontium-87 atoms is loaded into the ground band of a cubic optical lattice and initialized into the $|e\rangle = |5s5p\,{}^3P_0, F = 9/2, m_F = -9/2\rangle$ electronic state. The optical lattice is formed with a lattice constant of $a_{\text{lat}} \approx 407$ nm by interfering retro-reflected Gaussian laser beams, with 60-μm $1/e^2$ radii and peak depths of $k_B \times 12$ μK (where k_B is the Boltzmann constant), along each of the $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ axes. At these depths, tunneling rates are ~10 mHz between neighboring sites. Indexing the lattice sites by their positions $\mathbf{a} = a_{\text{lat}}(x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}})$ for integer $\{x, y, z\}$, in situ tomographic imaging (*50*) allows for the reconstruction of the site-wise atomic filling fractions $n_{\mathbf{a}}$, revealing a Fermi-Dirac distribution with a fitted peak density of $n_0 \approx 0.80$, root mean square (RMS) radii of $(w_{\hat{\mathbf{x}}'}, w_{\hat{\mathbf{y}}'}, w_{\hat{\mathbf{z}}'}) \approx (3.4\,\mu\text{m}, 4.3\,\mu\text{m}, 2.4\,\mu\text{m})$, and a mean entropy per atom of $2.0 \times k_B$ (*51*).

Clock spectroscopy is then performed on the $5s^2\,{}^1S_0 \leftrightarrow 5s5p\,{}^3P_0$ so-called clock transition at $\nu = kc/2\pi \approx 429$ THz using laser light that is phase stabilized to a cryogenic-silicon optical cavity (*9*, *52*). The probe light propagates with a wave-vector $\mathbf{k}_0 = k(\hat{\mathbf{x}}\sin\psi + \hat{\mathbf{y}}\cos\psi)$, motivating the choice of local frame $\zeta_{\mathbf{a}} = e^{-i(\omega t - \mathbf{k}_0\cdot\mathbf{a})}$. Resonant pulses with a $2\pi \times 50$ Hz Rabi frequency, and variable pulse areas θ and phase shifts ϕ, perform global rotations of the atomic state $\hat{\rho} \to \mathcal{R}(\theta,\phi)\hat{\rho} = \exp\left[-i\theta\sum_{\mathbf{a}}(\hat{S}^{\dagger}_{\mathbf{a}}e^{-i\phi} + \hat{S}_{\mathbf{a}}e^{i\phi})/2\right]\hat{\rho}$ + H.c.. A homogeneous $B \approx 290$ μT magnetic field applied along the $\hat{\mathbf{x}}$ axis creates a 540-Hz differential Zeeman splitting between the two available ground states, $|g_q\rangle = |5s^2\,{}^1S_0, F = 9/2, m_F = -9/2 + q\rangle$ for $q \in \{0, 1\}$ representing the polarization of an absorbed photon in the spherical basis such that their respective resonances with the excited state are spectroscopically resolved. As represented in Fig. 1B, each subspace exhibits distinct far-field ($kr_{\mathbf{ba}} \gg 1$) radiation patterns $|\mathbf{E}_{\mathbf{a}}(\mathbf{b})|^2$ owing to differences in the magnitudes C_q ($C_0 = \sqrt{9/11}$ and $C_1 = \sqrt{2/11}$) and orientations $\mathbf{e}_q$ [$\mathbf{e}_0 = \hat{\mathbf{x}}$ and $\mathbf{e}_1 = (\hat{\mathbf{y}} - i\hat{\mathbf{z}})/\sqrt{2}$] of the atomic dipole moments

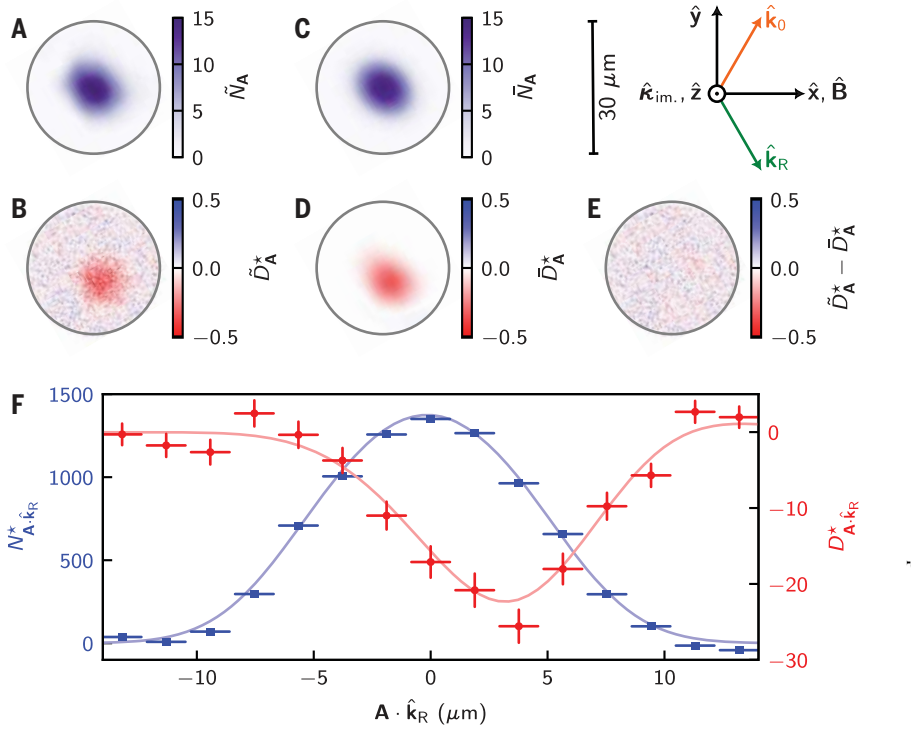


Fig. 2. Imaging cooperative Lamb shifts. (**A** and **B**) Sum $\tilde{N}_{\mathbf{A}}$ and rectified difference $\tilde{D}^{\star}_{\mathbf{A}} = \tilde{D}_{\mathbf{A}}/\cos\theta_{\text{in}}$ signals averaged over the subset of data corresponding to the angular momentum preserving the $q = 0$ transition and resonant angle of incidence $\psi = 29.5°$. Deviations of $\tilde{D}^{\star}_{\mathbf{A}}$ from zero indicate the presence of clock shifts that scale antisymmetrically with $\cos\theta_{\text{in}}$; the differences in the spatial profiles of $\tilde{N}_{\mathbf{A}}$ and $\tilde{D}^{\star}_{\mathbf{A}}$ indicate the presence of long-range, anisotropic interactions. (**C** and **D**) The modeled sum signal $\bar{N}_{\mathbf{A}}$ (C) is obtained by fitting $\tilde{N}_{\mathbf{A}}$ to a Fermi-Dirac distribution, which is then fed into Eq. 2 to obtain the modeled difference signal $\bar{D}_{\mathbf{A}}$ (D) using no other free parameters (*51*). (**E**) The residuals of the subtraction $\tilde{D}^{\star}_{\mathbf{A}} - \bar{D}^{\star}_{\mathbf{A}}$. The maximum of $D^{\star}_{\mathbf{A}}$ is spatially offset from the maximum of N_A along the direction $\hat{\mathbf{k}}_R = \hat{\mathbf{k}}_0 - 2\left(\hat{\mathbf{k}}_0\cdot\hat{\mathbf{y}}\right)\hat{\mathbf{y}}$ owing to constructive interference of the reflected probe light along k_R. (**F**) 1D projections of the sum $N_{\mathbf{A}\cdot\hat{\mathbf{k}}_R}$ and rectified difference $D^{\star}_{\mathbf{A}\cdot\hat{\mathbf{k}}_R}$ signals are obtained by projecting the images in (A) to (D) onto $\hat{\mathbf{k}}_R$. The measured signals $\tilde{N}_{\mathbf{A}\cdot\hat{\mathbf{k}}_R}$ and $\tilde{D}^{\star}_{\mathbf{A}\cdot\hat{\mathbf{k}}_R}$ are shown as blue circles and red squares, respectively. Vertical error bars represent 1σ standard errors, and horizontal bars show the 2-μm binwidth of the projections onto $\hat{\mathbf{k}}_R$. The modeled signals $\bar{N}_{\mathbf{A}\cdot\hat{\mathbf{k}}_R}$ and $\bar{D}^{\star}_{\mathbf{A}\cdot\hat{\mathbf{k}}_R}$ are shown as blue and red lines, respectively.

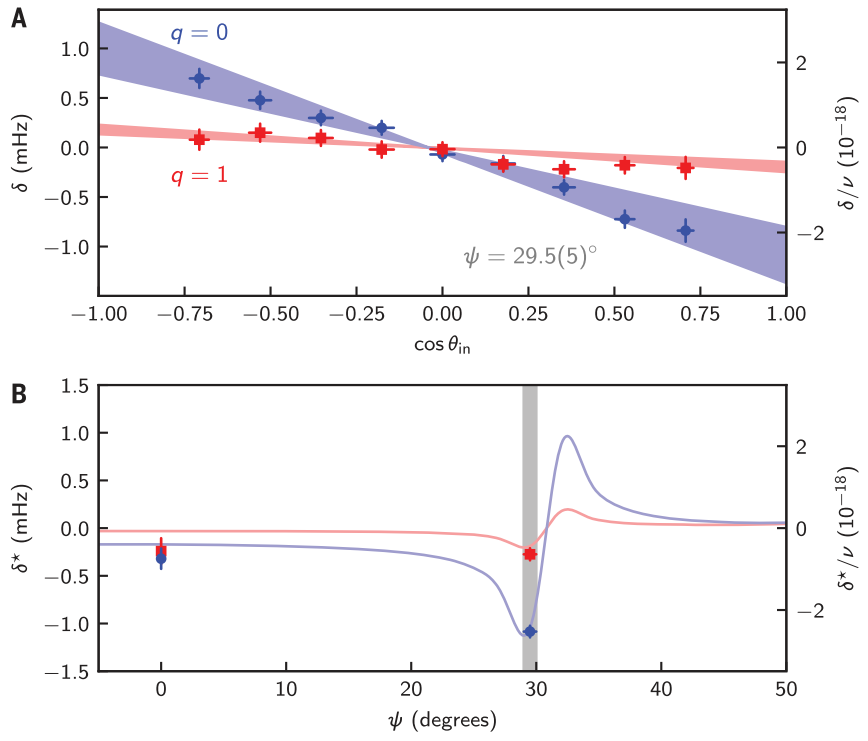


Fig. 3. Controlling cooperative Lamb shifts. (**A**) Scaling of the ensemble-averaged shift δ versus the initial spin projection $\cos\theta_{in}$ for $\psi = 29.5(5)°$ and both q = 0 (blue) and q = 1 (red). Data points show the measured shifts $\bar{\delta}$, with vertical error bars representing 1σ standard errors and horizontal error bars representing 2% observed fluctuations in the initial pulse area θ_{in}. Shaded regions show the modeled shifts $\bar{\delta}$, propagating the experimental uncertainties in ψ and θ_{in}. (**B**) Angle-of-incidence dependence of the shift sensitivity $\delta^* = \delta/\cos\theta_{in}$ to changes in the initial spin projection. Data points and solid lines show the measured and modeled shift sensitivities, respectively. The vertical gray bar represents the angle of incidence used in (A).

$\langle e|\hat{\mathbf{d}}|g_q\rangle = dC_q\mathbf{e}_q$ (*12*), where d^2 is proportional to the excited state's natural decay rate $\Gamma = k^3d^2/3\pi\varepsilon_0\hbar = 2\pi\times 1.35(3)$mHz (*53*). For each angle of incidence, the polarization of the probe laser is adjusted to obtain roughly equal Rabi coupling strengths for each transition at fixed intensity.

Figure 1C depicts the time evolution of the atomic state, as represented on the Bloch sphere with vector components $\hat{X}_\mathbf{a} = \hat{S}^\dagger_\mathbf{a} + \hat{S}_\mathbf{a}$, $\hat{Y}_\mathbf{a} = -\mathrm{i}(\hat{S}^\dagger_\mathbf{a} - \hat{S}_\mathbf{a})$, and $\hat{Z}_\mathbf{a} = \hat{S}^\dagger_\mathbf{a}\hat{S}_\mathbf{a} - \hat{S}_\mathbf{a}\hat{S}^\dagger_\mathbf{a}$ throughout the spectroscopic sequence $\mathcal{U}$. Starting with the initial conditions $\langle\hat{S}^\dagger_\mathbf{a}\hat{S}_\mathbf{a}\rangle = n_\mathbf{a}$ and $\langle\hat{S}_\mathbf{a}\hat{S}^\dagger_\mathbf{a}\rangle = 0$, the first pulse $\mathcal{R}(\theta_{in}, 0)$ rotates the population imbalance of the atomic state $\langle\hat{Z}_\mathbf{a}\rangle \to n_\mathbf{a}\cos\theta_{in}$. The elastic contribution of $\mathcal{L}_2$ to the subsequent free-evolution $\mathcal{F}(t) = e^{\mathcal{L}_{free}t}$ can be intuited, for short times $\Gamma t \ll 1$, as an Ising-type interaction that rotates each atom's Bloch vector about the $Z_\mathbf{a}$ axis at a rate of $-\cos\theta_{in}\sum_\mathbf{b} n_\mathbf{b}n_\mathbf{a}\mathrm{Re}(V^*_\mathbf{ba})$.

The two "spin-echo" pulses $\mathcal{R}(\pi, \phi)$ preserve the coherent dynamics generated by $\mathcal{L}_2$ while suppressing the various single-particle dephasing mechanisms contained in $\mathcal{L}_1[\hat{\rho}] = -\mathrm{i}\sum_\mathbf{a}(\Delta\omega_\mathbf{a} - \mathrm{i}\gamma/2)(\hat{S}^\dagger_\mathbf{a}\hat{S}_\mathbf{a}\hat{\rho} - \hat{S}_\mathbf{a}\hat{\rho}\hat{S}^\dagger_\mathbf{a}) + \mathrm{H.c.}$, where $\Delta\omega_\mathbf{a}$ is the relative detuning of atom $\mathbf{a}$ with respect to the probe laser. The dominant contribution to $\Delta\omega_\mathbf{a}$ arises from frequency drifts of the probe laser on the order of 1 Hz between daily measurements of the transition resonance frequencies. Differential ac Stark shifts varying with the local lattice intensity also contribute to $\Delta\omega_\mathbf{a}$ yet are limited to the sub–10-mHz level by optimizing the optical frequency of each trapping beam (*7*). These detunings do not directly affect the final state because the spin-echo pulses anticommute with time evolution under $\mathcal{L}_1$ in the limit $\gamma \to 0$, whereas the spin-echo pulses approximately commute with evolution under $\mathcal{L}_2$ for $\Gamma t \ll 1$. Finite γ leads to a decay in both the single-atom coherences $\langle\hat{S}^\dagger_\mathbf{a}\rangle \propto e^{-\gamma t/2}$ and excited state populations $\langle\hat{S}^\dagger_\mathbf{a}\hat{S}_\mathbf{a}\rangle \propto e^{-\gamma t}$. For an increasing number of spin-echo pulses, the time-averaged longitudinal decay asymptotically approaches $\langle\hat{Z}_\mathbf{a}\rangle \propto e^{-\gamma t/2}$. The final $\mathcal{R}(\pi/2, \phi_{out})$ pulse maps the interaction-induced phase shifts of the coherences onto the difference in electronic populations.

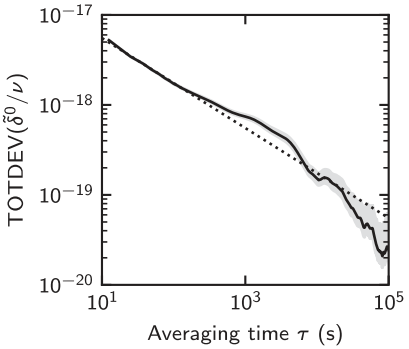


Fig. 4. Instability of clock shift evaluation. The solid black line shows the total deviation (TOTDEV) (*62*) of the spin-balanced clock shift $\bar{\delta}^0/\nu$ as a function of averaging time τ. The shaded region shows the 1σ confidence interval. The dotted black line shows a fit to the data for τ < 100 s, assuming a white noise floor.

A diffraction-limited imaging system with a 1.3-μm $1/e^2$ resolution then records the column-integrated populations of the ground $\tilde{N}^g_\mathbf{A}$ and excited $\tilde{N}^e_\mathbf{A}$ states onto a scientific complementary metal oxide semiconductor (sCMOS) camera through absorption imaging along $\boldsymbol{\kappa}_{im.} = \hat{\mathbf{z}}$ (*8*, *51*). Here, $\mathbf{A}$ indexes the sensor pixels by their positions. Although the collective spin vector does not necessarily evolve linearly in time when subject to Eq. 1 (*10*, *12*), the ensemble-averaged frequency shift may be computed, in the limit of small acquired phase shifts, as $\bar{\delta} = (2\pi T)^{-1}\sum_\mathbf{A}\tilde{D}_\mathbf{A}/\sum_\mathbf{A}\tilde{N}_\mathbf{A}$, where $\tilde{N}_\mathbf{A} = (\tilde{N}^e_\mathbf{A} + \tilde{N}^g_\mathbf{A})$ and $\tilde{D}_\mathbf{A} = (\tilde{N}^e_\mathbf{A} - \tilde{N}^g_\mathbf{A})/\mathcal{C}$ are the sum and difference signals, respectively, and $\mathcal{C} = \sin(\theta_{in})\sin(\phi_{out})e^{-\gamma T/2}$ is the interferometric sensitivity. Approximately 10^4 experimental shots were recorded over a period of 2 weeks while independently modulating the four parameters: $\sin\phi_{out} \in \{-1, 1\}$, $\cos\theta_{in} \in \left[-1/\sqrt{2}, 1/\sqrt{2}\right]$, $q \in \{0, 1\}$, and $\psi \in \{0°, 29.5(5)°\}$.

Experimental evidence for cooperative Lamb shifts

We compare experimentally derived quantities (denoted by symbols covered with a tilde, ~) with their modeled equivalents (denoted by symbols covered with a bar, –) obtained by substituting $\tilde{N}^g_\mathbf{A} \to \bar{N}^g_\mathbf{A} = \sum_{\mathbf{a}\|\mathbf{A}}\left(n_\mathbf{a} - \langle\hat{Z}_\mathbf{a}\mathcal{U}\rangle\right)/2$ and $\tilde{N}^e_\mathbf{A} \to \bar{N}^e_\mathbf{A} = \sum_{\mathbf{a}\|\mathbf{A}}\left(n_\mathbf{a} + \langle\hat{Z}_\mathbf{a}\mathcal{U}\rangle\right)/2$, where $\mathbf{a}\|\mathbf{A}$ denotes the set of all atoms whose image is projected onto the pixel $\mathbf{A}$, and the expectation value of $\hat{Z}_\mathbf{a}$ at the output of the interferometer is computed as

$$\begin{aligned}\langle\hat{Z}_\mathbf{a}\mathcal{U}\rangle &= n_\mathbf{a}\mathcal{C}[J_\mathbf{a}\cos(\Phi) + (1 - K_\mathbf{a})\sin(\Phi)]\\ J_\mathbf{a} &= -\cos(\theta_{in})T\sum_{\mathbf{b}\neq\mathbf{a}} n_\mathbf{b}\mathrm{Re}\left(V^*_\mathbf{ba}\right) + \mathcal{O}\left(\Gamma\gamma T^2\right)\\ K_\mathbf{a} &= \frac{\Gamma T}{2} + \mathcal{O}\left(\Gamma\gamma T^2\right)\end{aligned} \quad (2)$$

where $J_\mathbf{a}$ ($K_\mathbf{a}$) is the leading order, in ΓT, phase-shift (decoherence) of atom $\mathbf{a}$ resulting from resonant dipole-dipole interactions with all other atoms and whose full-time dependence in terms of the quantity γT is given in the supplementary materials (*51*). The parameter $\Phi = \Delta\phi(T) - 2\Delta\phi(3T/4) + 2\Delta\phi(T/4) - \Delta\phi(0)$ results from propagating noise-induced deviations

in the time-dependent relative phase between the atoms and the laser $\phi \rightarrow \phi + \Delta\phi(t)$ evaluated at each of the four rotation pulses.

On timescales comparable to T, the probe laser exhibits white frequency noise, where the RMS difference in phases over a time interval Δt is approximately $\sqrt{\langle \Delta\phi^2(\Delta t)\rangle}/\Delta t \approx 90\ \mathrm{mrads}^{-1}$ (*52*), contributing a zero-mean, stochastic signal on the order of $\Delta\tilde{D}_{\mathbf{A}} \approx \sqrt{\Phi^2}\tilde{N}_{\mathbf{A}} \approx \tilde{N}_{\mathbf{A}} \times$ 110 mrad to individual measurements of $\tilde{D}_{\mathbf{A}}$ (*51*). Noise in the observed data is consistent with $\sqrt{\Phi^2} \approx 230$ mrad and 300 mrad for the $q = 0$ and $q = 1$ transitions, respectively. We attribute the excess, transition-dependent noise to 3.2-nT (RMS) fluctuations in the magnetic field on timescales comparable to T (*9*, *51*, *54*). Owing to differences in the spatial profiles of $\Delta\bar{D}_{\mathbf{A}} \propto n_{\mathbf{a}}$ and $\bar{D}_{\mathbf{A}} \propto n_{\mathbf{a}} J_{\mathbf{a}}$, we are able to remove population differences caused by fluctuations in the probe-laser phase by applying corrections $\tilde{D}_{\mathbf{A}} \rightarrow \tilde{D}_{\mathbf{A}} - \mathcal{P}^{\mathrm{opt}}_{\Delta D}\Delta\bar{D}_{\mathbf{A}}$ to the presented data, where the coefficients $\mathcal{P}^{\mathrm{opt}}_{\Delta D}$ are obtained from least squares fits minimizing the quantity $\sum_{\mathbf{A}}\left(\mathcal{P}_{\Delta D}\Delta\bar{D}_{\mathbf{A}} + \mathcal{P}_D\bar{D}_{\mathbf{A}} - \tilde{D}_{\mathbf{A}}\right)^2/\mathrm{Var}\left(\tilde{D}_{\mathbf{A}}\right)$ over the parameters $\mathcal{P}_{\Delta D}$ and $\mathcal{P}_D$ (*51*).

Figure 2 shows the sum $N_{\mathbf{A}}$ and rectified difference $D^{\star}_{\mathbf{A}} = D_{\mathbf{A}}/\cos(\theta_{\mathrm{in}})$ signals averaged over the subset of data with maximal shift sensitivities: $\psi = 29.5°$ and $q = 0$. The ensemble-averaged cooperative Lamb shifts are plotted against $\cos\theta_{\mathrm{in}}$ for $\psi = 29.5°$ in Fig. 3A. A χ^2 analysis comparing the measured shifts with the model at each set of $(\cos\theta_{\mathrm{in}}, q, \psi)$ gives $\chi^2/(22\mathrm{d.o.f.}) \approx 1.18$. The shift's sensitivity to changes in the initial spin imbalance $\delta^{\star} = (2\pi T)^{-1}\sum_{\mathbf{A}} D^{\star}_{\mathbf{A}}/\sum_{\mathbf{A}} N_{\mathbf{A}}$ is plotted against ψ in Fig. 3B.

Whereas $V_{\mathbf{ba}}$ asymptotically decays with increasing separation as $1/kr_{\mathbf{ba}}$, the contained phase factors $e^{-\mathrm{i}(kr_{\mathbf{ba}} + \mathbf{k}_0\cdot\mathbf{r}_{\mathbf{ba}})}$ average to zero for incommensurate $ka_{\mathrm{lat}} \approx 7\pi/6$ and $\psi = 0°$, resulting in effectively nearest-neighbor interactions scaling with the local filling fractions—i.e., $J_{\mathbf{a}} \propto n_{\mathbf{a}}$. However, the subwavelength lattice spacing $ka_{\mathrm{lat}} < 2\pi$ guarantees the unique existence of the Bragg resonance at $\psi = \arccos(\pi/ka_{\mathrm{lat}}) \approx 30.8°$, satisfying $\mathbf{k}_0 \cdot a_{\mathrm{lat}}\hat{\mathbf{y}} = \pi$ such that the radiated fields add constructively along $\mathbf{k}_{\mathrm{R}} = \mathbf{k}_0 - 2(\mathbf{k}_0\cdot\hat{\mathbf{y}})\,\hat{\mathbf{y}}$. Numerically, we find that the ensemble-averaged interaction strengths are maximized, and scale with the system size as $N_{\mathrm{tot}}^{1/3}$, at angular detunings from exact Bragg resonance set by the diffraction limit $\pi/2k(w_{\mathbf{x}}^2 + w_{\mathbf{y}}^2)^{1/2} \approx 1.8°$ (*10*, *51*, *55*).

Fitting the observed shifts to $\tilde{\delta} = \tilde{\delta}^{\star}\cos\theta_{\mathrm{in}} + \tilde{\delta}^0$, we evaluate the mean clock shift owing to resonant dipole-dipole interactions at $\cos\theta_{\mathrm{in}} = 0$ to be $\tilde{\delta}^0/\nu = -1.3(8) \times 10^{-19}$, which demonstrates that systematic effects can be made negligible relative to the lowest reported total systematic uncertainties for optical atomic clocks (*56*–*58*). Figure 4 displays the fractional frequency instability of the $\tilde{\delta}^0/\nu$ evaluation, which exhibits a $1.8 \times 10^{-17}/\sqrt{\mathrm{Hz}}$ short-term white noise floor.

Discussion and outlook

We have performed measurements of, and successfully modeled, cooperative Lamb shifts in a 3D optical lattice clock. Control over the spatial orientations of the probe light and excited dipole moments allow for a notable modification of the magnitude of these effects—from levels relevant to state-of-the-art atomic clocks to more than an order of magnitude below. Technical dephasing caused by Raman scattering of optical lattice photons prevented the study of dynamics beyond $\Gamma t \ll 1$.

It is interesting to consider a regime where collective interactions are significantly stronger than technical dephasing rates. This may be achieved by probing transitions with stronger intrinsic dipole moments, either directly or by optically dressing the clock states (*59*). Under such conditions, where coherent manipulations on timescales much shorter than dynamics associated with free evolution may still be possible, collective light-matter interactions are expected to lead to spin squeezing (*60*) and other exotic states of quantum matter (*45*). The desire to produce metrologically useful entanglement motivates future work interrogating the $5s5p\ ^3P_0 \leftrightarrow 5s4d\ ^3D_1$ transition at 2.6 μm satisfying $ka_{\mathrm{lat}} \approx 1$ (*60*, *61*), where linearly polarized light incident along $\hat{\mathbf{k}}_0 \cdot \hat{\mathbf{B}} \ll 1$ can produce coherent dynamics that are dominant over collective dissipation. Beyond defining a precise platform for the study of effective photon-photon interactions (*44*, *45*), such engineered arrays of narrow-band quantum emitters provide a path to unexplored photonic devices based on controlled collective atom-photon dynamics.

REFERENCES AND NOTES

1. S. L. Braunstein, C. M. Caves, *Phys. Rev. Lett.* **72**, 3439–3443 (1994).
2. V. Giovannetti, S. Lloyd, L. Maccone, *Phys. Rev. Lett.* **96**, 010401 (2006).
3. C. M. Caves, *Phys. Rev. Lett.* **45**, 75–79 (1980).
4. W. M. Itano *et al*., *Phys. Rev. A* **47**, 3554–3570 (1993).
5. M. O. Scully, J. P. Dowling, *Phys. Rev. A* **48**, 3186–3190 (1993).
6. T. Akatsuka, M. Takamoto, H. Katori, *Phys. Rev. A* **81**, 023402 (2010).
7. S. L. Campbell *et al*., *Science* **358**, 90–94 (2017).
8. G. E. Marti *et al*., *Phys. Rev. Lett.* **120**, 103201 (2018).
9. E. Oelker *et al*., *Nat. Photonics* **13**, 714–719 (2019).
10. D. E. Chang, J. Ye, M. D. Lukin, *Phys. Rev. A* **69**, 023810 (2004).
11. S. Krämer, L. Ostermann, H. Ritsch, *Europhys. Lett.* **114**, 14003 (2016).
12. A. Cidrim *et al*., *Phys. Rev. Lett.* **127**, 013401 (2021).
13. J. D. Jackson, *Classical Electrodynamics* (Wiley, ed. 3, 1998).
14. P. de Vries, D. V. van Coevorden, A. Lagendijk, *Rev. Mod. Phys.* **70**, 447–466 (1998).
15. F. Andreoli, M. J. Gullans, A. A. High, A. Browaeys, D. E. Chang, *Phys. Rev. X* **11**, 011026 (2021).
16. R. H. Lehmberg, *Phys. Rev. A* **2**, 883–888 (1970).
17. R. H. Dicke, *Phys. Rev.* **93**, 99–110 (1954).
18. V. M. Fain, *Sov. Phys. JETP* **9**, 562 (1959).
19. R. Friedberg, S. Hartmann, J. Manassah, *Phys. Rep.* **7**, 101–179 (1973).
20. N. Skribanowitz, I. P. Herman, J. C. MacGillivray, M. S. Feld, *Phys. Rev. Lett.* **30**, 309–312 (1973).
21. M. Gross, C. Fabre, P. Pillet, S. Haroche, *Phys. Rev. Lett.* **36**, 1035–1038 (1976).
22. M. Gross, P. Goy, C. Fabre, S. Haroche, J. M. Raimond, *Phys. Rev. Lett.* **43**, 343–346 (1979).
23. D. Pavolini, A. Crubellier, P. Pillet, L. Cabaret, S. Liberman, *Phys. Rev. Lett.* **54**, 1917–1920 (1985).
24. P. V. Zinov'ev *et al*., *Sov. Phys. JETP* **58**, 1129–1133 (1983).
25. O. P. Varnavskii, A. N. Kirkin, A. M. Leontovich, *Sov. Phys. JETP* **59**, 716 (1984).
26. R. G. DeVoe, R. G. Brewer, *Phys. Rev. Lett.* **76**, 2049–2052 (1996).
27. M. D. Barnes, P. S. Krstic, P. Kumar, A. Mehta, J. C. Wells, *Phys. Rev. B* **71**, 241303 (2005).
28. M. Scheibner *et al*., *Nat. Phys.* **3**, 106–110 (2007).
29. A. Goban *et al*., *Phys. Rev. Lett.* **115**, 063601 (2015).
30. B. H. McGuyer *et al*., *Nat. Phys.* **11**, 32–36 (2015).
31. M. Houde, A. Mathews, F. Rajabi, *Mon. Not. R. Astron. Soc.* **475**, 514–522 (2017).
32. W. R. Garrett, R. C. Hart, J. E. Wray, I. Datskou, M. G. Payne, *Phys. Rev. Lett.* **64**, 1717–1720 (1990).
33. R. Röhlsberger, K. Schlage, B. Sahoo, S. Couet, R. Rüffer, *Science* **328**, 1248–1251 (2010).
34. J. Keaveney *et al*., *Phys. Rev. Lett.* **108**, 173601 (2012).
35. A. F. van Loo *et al*., *Science* **342**, 1494–1496 (2013).
36. Z. Meir, O. Schwartz, E. Shahmoon, D. Oron, R. Ozeri, *Phys. Rev. Lett.* **113**, 193002 (2014).
37. J. Rui *et al*., *Nature* **583**, 369–374 (2020).
38. G. Ferioli, A. Glicenstein, L. Henriet, I. Ferrier-Barbut, A. Browaeys, *Phys. Rev. X* **11**, 021031 (2021).
39. R. B. Hutson *et al*., *Phys. Rev. Lett.* **123**, 123401 (2019).
40. H. Zoubi, H. Ritsch, *Phys. Rev. A* **83**, 063831 (2011).
41. S. D. Jenkins, J. Ruostekoski, *Phys. Rev. A* **86**, 031602 (2012).
42. D. E. Chang, L. Jiang, A. V. Gorshkov, H. J. Kimble, *New J. Phys.* **14**, 063003 (2012).
43. R. J. Bettles, S. A. Gardiner, C. S. Adams, *Phys. Rev. A* **92**, 063822 (2015).
44. A. Asenjo-Garcia, M. Moreno-Cardoner, A. Albrecht, H. J. Kimble, D. E. Chang, *Phys. Rev. X* **7**, 031024 (2017).
45. L. Henriet, J. S. Douglas, D. E. Chang, A. Albrecht, *Phys. Rev. A* **99**, 023802 (2019).
46. E. Sierra, S. J. Masson, A. Asenjo-Garcia, *Phys. Rev. Res.* **4**, 023207 (2022).
47. Y.-X. Zhang, K. Mølmer, *Phys. Rev. Lett.* **122**, 203605 (2019).
48. A. Glicenstein *et al*., *Phys. Rev. Lett.* **124**, 253602 (2020).
49. L. Sonderhouse *et al*., *Nat. Phys.* **16**, 1216–1221 (2020).
50. W. R. Milner, L. Yan, R. B. Hutson, C. Sanner, J. Ye, *Phys. Rev. A* **107**, 063313 (2023).
51. See the supplementary materials.
52. D. G. Matei *et al*., *Phys. Rev. Lett.* **118**, 263202 (2017).
53. J. A. Muniz, D. J. Young, J. R. K. Cline, J. K. Thompson, *Phys. Rev. Res.* **3**, 023152 (2021).
54. M. M. Boyd *et al*., *Phys. Rev. A* **76**, 022510 (2007).
55. J. T. Manassah, *Phys. Lett. A* **374**, 1985–1988 (2010).
56. W. F. McGrew *et al*., *Nature* **564**, 87–90 (2018).
57. S. M. Brewer *et al*., *Phys. Rev. Lett.* **123**, 033201 (2019).
58. T. Bothwell *et al*., *Metrologia* **56**, 065004 (2019).
59. R. Santra, E. Arimondo, T. Ido, C. H. Greene, J. Ye, *Phys. Rev. Lett.* **94**, 173002 (2005).
60. C. Qu, A. M. Rey, *Phys. Rev. A* **100**, 041602 (2019).
61. S. J. Masson, J. P. Covey, S. Will, A. Asenjo-Garcia, Dicke superradiance in ordered arrays of multilevel atoms. arXiv:2304.00093 [quant-ph] (2023).
62. D. A. Howe, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **47**, 1102–1110 (2000).
63. R. B. Hutson, W. R. Milner, L. Yan, J. Ye, C. Sanner, Observation of mHz-level cooperative Lamb shifts in an optical atomic clock, dataset, Dryad (2023); https://doi.org/10.5061/dryad.wpzgmsbtj.

ACKNOWLEDGMENTS

We thank D. Kedar for technical assistance and D. E. Chang, H. Ritsch, and M. D. Lukin for useful discussions. We also thank S. L. Campbell, N. Darkwah Oppong, A. M. Rey, and D. Wellnitz for careful reading of the manuscript and for providing insightful comments. **Funding:** Funding for this work is provided by National Science Foundation (NSF) QLCI OMA-2016244, the US Department of Energy Center of Quantum System Accelerator, a V. Bush Fellowship, the National Institute of Standards and Technology (NIST), and NSF Phys-1734006. **Author contributions:** All authors contributed to carrying out the experiments, interpreting the results, and writing the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data and code used to generate the figures are archived at Dryad (*63*). **License information:**

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh4477
Supplementary Text
Figs. S1 to S4
Tables S1 to S3

Submitted 7 March 2023; accepted 15 December 2023
10.1126/science.adh4477

THIN FILMS

Super-tetragonal $Sr_4Al_2O_7$ as a sacrificial layer for high-integrity freestanding oxide membranes

Jinfeng Zhang[1]†, Ting Lin[2]†, Ao Wang[1]†, Xiaochao Wang[3]†, Qingyu He[4], Huan Ye[1], Jingdi Lu[1], Qing Wang[1], Zhengguo Liang[1], Feng Jin[1], Shengru Chen[2], Minghui Fan[1], Er-Jia Guo[2], Qinghua Zhang[2], Lin Gu[5], Zhenlin Luo[4], Liang Si[3,6]*, Wenbin Wu[1,7,8]*, Lingfei Wang[1]*

Identifying a suitable water-soluble sacrificial layer is crucial to fabricating large-scale freestanding oxide membranes, which offer attractive functionalities and integrations with advanced semiconductor technologies. Here, we introduce a water-soluble sacrificial layer, "super-tetragonal" $Sr_4Al_2O_7$ (SAO_T). The low-symmetric crystal structure enables a superior capability to sustain epitaxial strain, allowing for broad tunability in lattice constants. The resultant structural coherency and defect-free interface in perovskite ABO_3/SAO_T heterostructures effectively restrain crack formation during the water release of freestanding oxide membranes. For a variety of nonferroelectric oxide membranes, the crack-free areas can span up to a millimeter in scale. This compelling feature, combined with the inherent high water solubility, makes SAO_T a versatile and feasible sacrificial layer for producing high-quality freestanding oxide membranes, thereby boosting their potential for innovative device applications.

Transition metal oxide-based heterostructures are characterized by a wide array of emergent interfacial phenomena, stimulated by the coupling of spin, charge, orbital, and lattice degrees of freedom at the heterointerfaces (*1*, *2*). Examples include two-dimensional electron or hole gas (*3*), interfacial superconductivity (*4*), improper ferroelectricity (*5*), and magnetic or polar skyrmions (*6*, *7*). Although these interfacial phenomena hold rich physics and functionalities (*8*–*10*), the strong covalent bonds at film-substrate interfaces largely limit their integration with other low-dimensional material systems and thus the potential device applications (*11*, *12*). In recent years, freestanding oxide membrane exfoliating and transferring technologies have developed rapidly (*13*–*17*). Among these advancements, the water-assisted exfoliation of freestanding oxide membranes using cubic $Sr_3Al_2O_6$ (SAO_C) epitaxial sacrificial layers has emerged as one of the most prominent and feasible approaches (*14*). Since its discovery in 2016, SAO_C has boosted research on integrating ABO_3 perovskite oxide heterostructures with van der Waals materials and advanced semiconductor technologies, signifying great potential for next-generation electronic or spintronic devices (*18*–*20*). Moreover, SAO_C provides a step forward in exploiting functionalities that exclusively exist in the freestanding membrane form, including ferroelastic domain–mediated superelasticity (*21*, *22*), ferroelectricity in the monolayer limit (*23*), correlated electronic phase under extreme tensile strain (*24*, *25*), novel lateral twisting and boundary states (*26*, *27*), and switchable polar skyrmions (*28*).

Despite these promising advancements, the crystallinity and integrity of the freestanding oxide membranes remain unsatisfactory compared with typical van der Waals materials such as graphene and transition metal dichalcogenides (*29*, *30*). Particularly for the nonferroelectric (non-FE) oxides, the water-assisted release processes are often accompanied by degraded crystalline coherence length and high-density crack formation (*31*–*34*). Millimeter-sized crack-free membranes were rarely achieved (*24*, *25*). Such brittle fractures in released oxide membranes can be attributed to two main factors: (i) the intrinsic structural characteristics, including strong ionic or covalent bonds and lack of slip system; and (ii) extrinsic defect formation due to unavoidable relaxation of misfit strain (*34*–*37*). Because of the strong electron correlation nature, such unwilling structural changes also cause considerable degradation of physical properties in freestanding oxide membranes, limiting their potential in next-generation electronic device applications. To address this challenge, several new sacrificial layer materials have been developed recently, aiming to reduce the interfacial lattice mismatch and crack density. But the improvement is still limited by the discrete lattice constants, poor solubility, or nongeneric etchant (*13*, *37*–*39*).

In this study, we systematically explored the growth phase diagram of SAO_C films and discovered a previously unknown $Sr_4Al_2O_7$ p[illegible] (denoted as SAO_T). The biaxial-strained S[illegible] film has a tetragonal structural symmetry and Sr-rich stoichiometry, distinct from the well-recognized cubic SAO_C phase. Such a low-symmetry crystal structure of SAO_T enables superior flexibility under epitaxial strain and thus wide-range tunability of in-plane lattice constants. The resultant coherent growth of high-quality ABO_3/SAO_T epitaxial heterostructures considerably improves the crystallinity and integrity of water-released freestanding oxide membranes. For the representative non-FE nickelates, manganites, titanates, ruthenates, and stannates with a broad lattice constant range (3.85 to 4.04 Å), the crack-free areas of the membranes released from SAO_T can span up to a few millimeters in scale. The corresponding functionalities are comparable to the epitaxial counterparts. Moreover, the distinct atomic structure of SAO_T leads to an inherent high water solubility, thus ensuring an effective water-assisted exfoliation process. These compelling advantages make SAO_T film a versatile and viable water-soluble sacrificial layer for fabricating a broad array of high-quality freestanding oxide membranes, offering fertile grounds for the development of innovative electronic devices.

Growth window of SAO_T films

The strontium aluminate (SAO, including SAO_C and SAO_T) thin films and ABO_3/SAO multilayer heterostructures were epitaxially grown by pulsed laser deposition (PLD). We first grew the SAO films by laser ablation of a polycrystalline stoichiometric SAO_C target (molar ratio Sr:Al:O = 3:2:6) on (001)-oriented $(LaAlO_3)_{0.3}$-$(SrAl_{0.5}Ta_{0.5}O_3)_{0.7}$ [LSAT(001)] single crystalline substrates (*40*). The epitaxial quality and stoichiometry of the SAO films predominantly depend on two parameters: oxygen partial pressure (P_{O2}) and laser fluence (F_L). During the growth of a series of SAO films, we altered the P_{O2} from 10^{-4} to 20 Pa and adjusted the F_L from 1.0 to 2.5 J/cm^2. We show a full set of x-ray diffraction (XRD) 2θ-ω scans of SAO/LSAT(001) films near the LSAT(002) diffraction (fig. S1), along with two representative curves (Fig. 1A). Most curves exhibit a clear film diffraction peak near 45.8°, in line with previously reported SAO_C(008) diffractions (*14*). We summarize the out-of-plane d-spacing of SAO_C(008) and peak intensity in figs. S2 and S3. Based on these parameters and the sharpness of Laue fringes, we can evaluate the epitaxial quality of SAO films and construct the growth phase diagram (Fig. 1B). Using the stoichiometric SAO_C target, SAO_C films can grow well within a dome-like F_L-P_{O2} range, consistent with the large variety of SAO_C growth conditions reported in the literature (*13*). Film deposition beyond the F_L-P_{O2} boundaries mostly leads to off-stoichiometry and poor crystallinity.

[1]Hefei National Research Center for Physical Sciences at Microscale, University of Science and Technology of China, Hefei 230026, China. [2]Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China. [3]School of Physics, Northwest University, Xi'an 710127, China. [4]National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei 230026, China. [5]Beijing National Center for Electron Microscopy and Laboratory of Advanced Materials, Department of Materials Science and Engineering, Tsinghua University, Beijing 100084, China. [6]Institut für Festkörperphysik, TU Wien, 1040 Vienna, Austria. [7]Institutes of Physical Science and Information Technology, Anhui University, Hefei 230601, China. [8]Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing 210093, China.
*Corresponding author. Email: liang.si@ifp.tuwien.ac.at (L.S.); wuwb@ustc.edu.cn (W.W.); wanglf@ustc.edu.cn (L.W.)
†These authors contributed equally to this work.

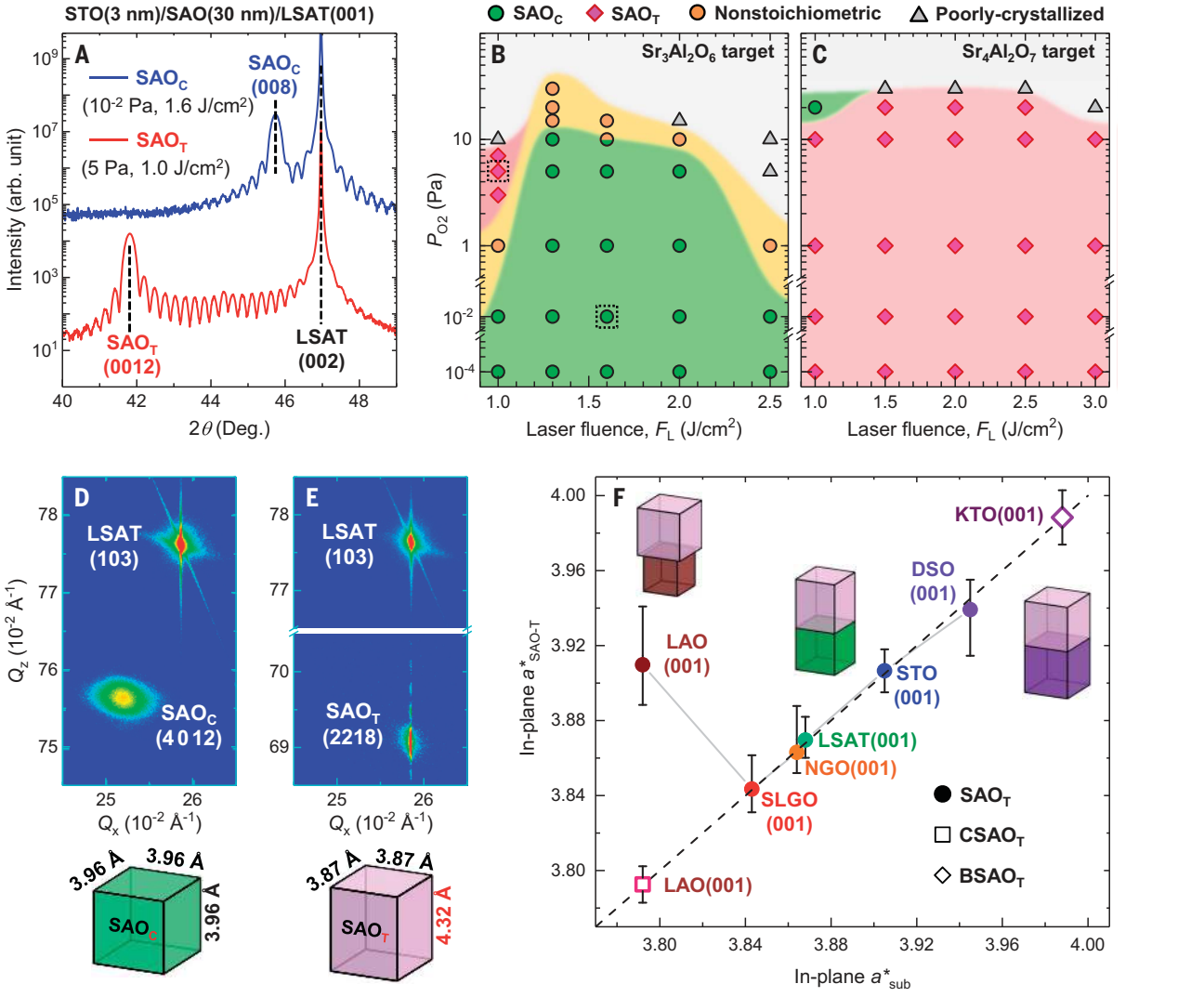


Fig. 1. Growth of SAO epitaxial films. (**A**) XRD 2θ-ω linear scans measured from 30-nm-thick SAO films grown on (001)-oriented $(LaAlO_3)_{0.3}$-$(SrAl_{0.5}Ta_{0.5}O_3)_{0.7}$ [LSAT (001)] substrates. The cubic and super-tetragonal–like SAO phases are denoted as SAO_C and SAO_T, respectively. (**B** and **C**) Laser fluence–oxygen partial pressure (F_L-P_{O2}) phase diagram of SAO film growth using (B) $Sr_3Al_2O_6$ and (C) $Sr_4Al_2O_7$ targets. Using a $Sr_3Al_2O_6$ target, SAO_T film can be grown in a narrow F_L-P_{O2} window, whereas film deposition using a $Sr_4Al_2O_7$ target enables a much broader F_L-P_{O2} window for growing high-quality SAO_T film. The growth conditions for the samples in (A) are marked by dashed boxes. (**D** and **E**) Reciprocal space mappings from (D) SAO_C/LSAT(001) and (E) SAO_T/LSAT(001) films. The lower panels of (D) and (E) depict schematics of pseudocubic SAO_C and SAO_T unit cells, with lattice constants labeled. According to the crystal structure analyses (Fig. 2), the diffractions of SAO_T in (A) and (E) should be indexed as (0012) and (2218), respectively. (**F**) Pseudocubic in-plane lattice constants ($a^*_{SAO\text{-}T}$) from the RSMs of 30-nm-thick SAO_T films grown on various substrates, including $LaAlO_3$(001) [LAO(001)], $SrLaGaO_4$(001) [SLGO(001)], $NdGaO_3$(001) [NGO(001)], LSAT(001), $SrTiO_3$(001) [STO(001)], and $DyScO_3$(001) [DSO(001)]. The $a^*_{SAO\text{-}T}$ values are plotted as a function of the in-plane lattice constants (a^*_{sub}). All the lattice constants are converted into pseudocubic notation. The $a^*_{SAO\text{-}T}$ values for Ba-doped and Ca-doped SAO_T films [$BSAO_T$/$KTaO_3$(001) and $CSAO_T$/LAO(001) films, respectively] are also included. The error bars represent the uncertainty of $a^*_{SAO\text{-}T}$ due to the broadening of diffraction spots in RSMs. The $a^*_{SAO\text{-}T}$ values for most of the SAO_T films [except for the SAO_T/LAO(001) film] align with the dashed line $a^*_{SAO\text{-}T} = a^*_{sub}$, suggesting a coherent strain state and broad strain-tuning range of $a^*_{SAO\text{-}T}$.

However, within a narrow window near F_L = 1.0 J/cm^2 and P_{O2} = 5 Pa, sharp Laue fringes re-emerge, and the SAO diffraction unexpectedly shifts to a much lower Bragg angle of ~41.8° (Fig. 1A and fig. S4), signifying the emergence of a new structural phase, which we denoted as SAO_T. Subsequent structure characterizations further reveal that the epitaxial quality of the SAO_T film is comparable to that of the optimized SAO_C films (fig. S5). Using energy-dispersive spectroscopy measurements, we further confirmed that the Sr:Al molar ratio of SAO_T films is ~2.05, close to the nominal value of a $Sr_4Al_2O_7$ compound (table S1 and fig. S6). The large deviation in chemical stoichiometry between the SAO_T film and SAO_C (Sr:Al ~ 1.51) target should relate to the kinetic processes during the laser ablation and deposition in such a narrow F_L-P_{O2} window [see section 1 of (*40*)]. Accordingly, we refined the film deposition using a $Sr_4Al_2O_7$ target and substantially expanded the growth window of high-quality SAO_T film (Fig. 1C and fig. S7). The wide P_{O2} range (10^{-4} to 20 Pa) and F_L range (1.0 to 3.0 J/cm^2) should align with most PLD system capabilities.

We further characterized the epitaxial strain states of the SAO_C and SAO_T films through reciprocal space mappings (RSMs). For the SAO_C/LSAT(001) film (Fig. 1D), the SAO_C(4012) and LSAT(103) diffractions have unequal in-plane and out-of-plane reciprocal space vectors (Q_x and Q_z). Both the in-plane and out-of-plane lattice constants in pseudocubic perovskite

notation derived from the RSMs are 3.96 Å ($a^*_{SAO\text{-}C}$). The $a^*_{SAO\text{-}C}$ value is the same as that of bulk SAO_C, signifying a fully relaxed epitaxial strain. For the SAO_T/LSAT(001) film (Fig. 1E), however, the elongated SAO_T(2218) diffraction shares the same Q_x with that of the LSAT(103) diffraction, demonstrating a coherent strain state. The in-plane and out-of-plane lattice constants in pseudocubic perovskite notation ($a^*_{SAO\text{-}T}$ and $c^*_{SAO\text{-}T}$) are 3.87 and 4.32 Å, respectively. We expect such a biaxially and compressively strained SAO_T/LSAT(001) film to show tetragonal symmetry (figs. S8 to S10). The tetragonality (c/a ratio) is up to ~1.12. Drawing an analogy to the compressive strain–induced isosymmetric phase transition in $BiFeO_3$ films (*41*), we suggest that the SAO_T phase could be a "super-tetragonal" structural polymorph, featuring prominent atomic arrangement changes from the parent SAO_C.

The ability to sustain epitaxial strain seems to be the most prominent feature of the SAO_T phase. The coherent strain state can be maintained in the SAO_T/LSAT(001) films up to a thickness of 100 nm (fig. S11). The resultant structural coherency further ensures a sharp and defect-free SAO_T/LSAT(001) interface (fig. S12). In addition to the LSAT(001), high-quality SAO_T film can be epitaxially grown on and coherently strained to a variety of (001)-oriented ABO_3 perovskite substrates (fig. S13). As summarized in Fig. 1F, the in-plane lattice constant $a^*_{SAO\text{-}T}$ can be continuously adjusted over a wide range, from 3.84 Å [on $SrLaGaO_4$(001)] to 3.95 Å [on $DyScO_3$(001)]. Further, by introducing Ba and Ca doping to the SAO_T, we successfully grew coherently strained $Ba_2Sr_2Al_2O_7$/$KTaO_3$(001) and $Ca_2Sr_2Al_2O_7$/$LaAlO_3$(001) films (fig. S14). Consequently, the strain-tuning range of $a^*_{SAO\text{-}T}$ is expanded to between 3.79 and 3.99 Å. The structural coherency even persists in the (110)-oriented SAO_T films grown on $SrLaGaO_4$(100) substrate (fig. S15). Such superior strain adaptability, rarely observed in the SAO_C counterparts, should be an inherent property of SAO_T, stemming from its distinctive crystal structure.

Crystal structure of SAO_T

After identifying the compelling structural flexibility of SAO_T, we next probed its structural differences with SAO_C at the atomic scale using cross-sectional scanning transmission electron microscopy (STEM). The STEM image of the epitaxial SAO_C/LSAT(001) film measured in high-angle annular dark-field (HAADF) mode displays a rhombus-like contrast along the LSAT[010] zone axis (Fig. 2, A and B), stemming from alternating "B-site" cations (Sr and Al) and regularly ordered oxygen and cation vacancies (*14*). As we schematically depict (Fig. 2C), such a cation ordering quadruples the lattice constant of SAO_C ($a_{SAO\text{-}C}$) compared with that of cubic ABO_3 perovskite. In sharp contrast, the HAADF-STEM images of the SAO_T film along LSAT[010] zone axis (Fig. 2, D and E) display a perovskite-like atomic contrast and an intensity modulation along the out-of-plane [001] axis. The "A-site" atomic columns exhibit higher intensity alternatively for every three perovskite-like unit cells (see fig. S16 for details). More notably, the HAADF-STEM image captured along the LSAT[1-10] zone axis (Fig. 2F) displays a complex cation ordering: A-sites are fully occupied and display a similar three-unit-cell intensity modulation, while the B-sites are alternatively occupied and show a blurry in-plane intensity modulation.

We determined the atomic structure of the SAO_T phase on the basis of the STEM images and density functional theory (DFT) calculations. To the best of our knowledge, none of the reported strontium aluminates can match the STEM and XRD results. We searched numerous possible structure candidates and eventually found that the SAO_T could share an atomic structure similar to that of the orthorhombic $Ba_4Al_2O_7$ compound (*42*). DFT-level structure relaxations further confirm that the orthorhombic $Sr_4Al_2O_7$ unit cell is thermodynamically stable, with simulated lattice constants $a_{SAO\text{-}T}$ = 10.798 Å, $b_{SAO\text{-}T}$ = 11.238 Å, and $c_{SAO\text{-}T}$ = 25.732 Å [see section 2 of (*40*)]. The atomic arrangements of the simulated SAO_T ($Sr_4Al_2O_7$) unit cell viewed along both SAO_T[100] and [010] axes match perfectly with the HAADF-STEM image along LSAT[1-10] zone axis (Fig. 2F

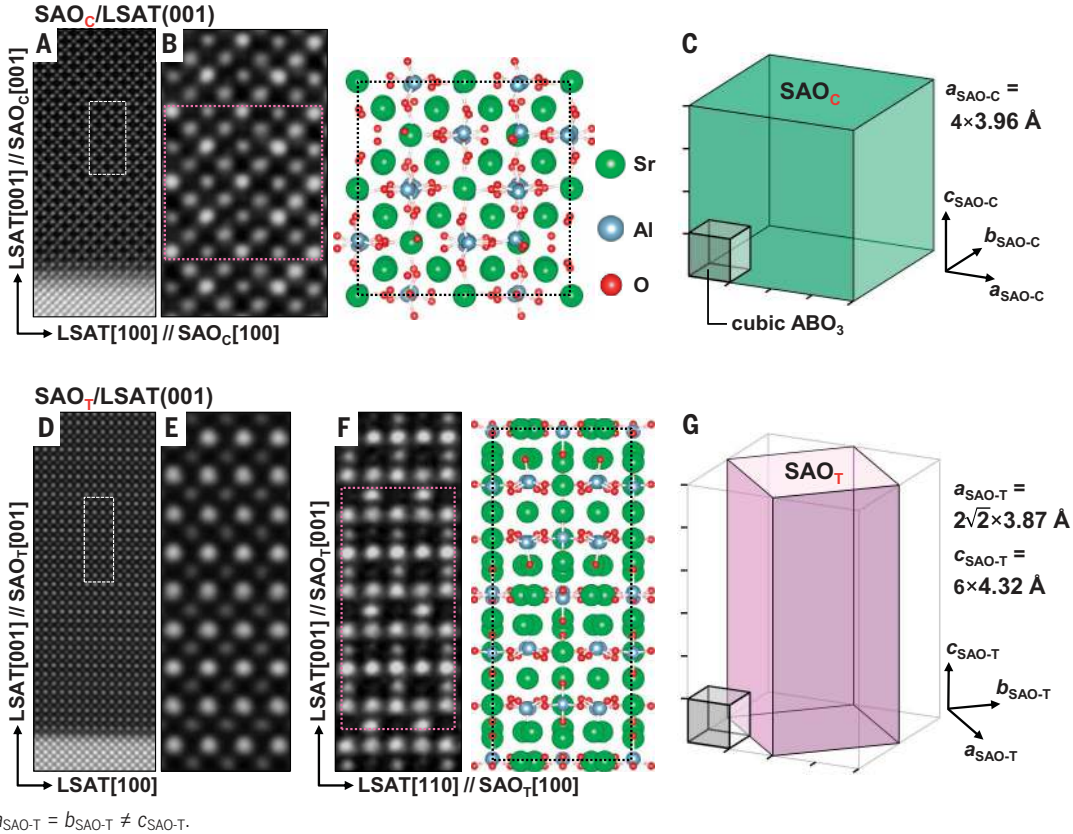


Fig. 2. Crystal structures of SAO_C and SAO_T films. (**A**) HAADF-STEM images captured near the SAO_C/LSAT(001) interface, viewed along the LSAT[010] axis. (**B**) Zoom-in HAADF-STEM image from the area marked by a white dashed box in (A) and DFT-relaxed atomic structure of SAO_C unit cell. (**C**) Schematic illustration of the relative dimensions of the SAO_C unit cell and a cubic ABO_3 perovskite unit cell. (**D**) HAADF-STEM image captured near the SAO_T/LSAT(001) interface. (**E**) Zoom-in HAADF-STEM image from the area marked by a white dashed box in (D). Both (D) and (E) are viewed along the LSAT[010] zone axis. (**F**) Zoom-in HAADF-STEM image measured from SAO_T film and DFT-relaxed crystal structure, both viewed along SAO_T[010] axis (parallel to LSAT[1-10]). (**G**) Schematic illustration of the relative dimensions of the SAO_T unit cell and a cubic ABO_3 perovskite unit cell. The lattice constants of SAO_C and SAO_T are labeled in (C) and (G). For the cubic SAO_C unit cell, $a_{SAO\text{-}C} = b_{SAO\text{-}C} = c_{SAO\text{-}C}$. And for the tetragonal SAO_T unit-cell, $a_{SAO\text{-}T} = b_{SAO\text{-}T} \neq c_{SAO\text{-}T}$.

and fig. S17). Hence, the epitaxial relationship of coherently grown SAO_T/LSAT(001) films should be SAO_T[100]//LSAT[110] and SAO_T[001]//LSAT[001] (fig. S18). Assuming that the SAO_T unit cell is $2\sqrt{2} \times 2\sqrt{2} \times 6$ times as large as the cubic perovskite unit cell (Fig. 2G), we can obtain the reduced lattice constants as $a^*_{SAO\text{-}T}$ = $b^*_{SAO\text{-}T}$ = 3.896 Å and $c^*_{SAO\text{-}T}$ = 4.288 Å. These simulated values are very close to those derived from the XRD results (table S2), further supporting the validity of our proposed structure. For the SAO_T films grown on cubic substrates [e.g., LSAT(001) or STO(001)], the biaxial strain could convert the lattice symmetry from orthorhombic to tetragonal. For the films grown on orthorhombic substrates [e.g., $NdGaO_3$(001)], the original orthorhombic symmetry is preserved (fig. S10).

The atomic structure of SAO_T also plays a deterministic role in enabling the coherent epitaxial strain at ABO_3/SAO_T interfaces. We first performed DFT calculations on the relative energy changes (ΔE) of SAO_C and SAO_T unit cells under manually imposed biaxial and anisotropic strain (*40*). For both biaxial and anisotropic strain configurations (Fig. 3, A and B, and fig. S19), the ΔE calculated from SAO_T is consistently smaller than that of SAO_C. Specifically, the low-symmetry SAO_T unit cell is more flexible in terms of accommodating the misfit strain imposed by either cubic or orthorhombic substrates, whereas the cubic SAO_C unit cell could be more rigid against strain-induced lattice distortion. Moreover, we also constructed SAO_C/STO(001) and SAO_T/STO(001) heterostructures as model systems (fig. S20) and evaluated their interfacial bonding strength. The DFT-calculated bonding energy at the SAO_T/STO(001) interface (1.97 eV) is more than twofold that at SAO_C/STO(001) interfaces (0.82 eV) and even comparable with that at the $LaAlO_3$/STO interface (2.34 eV) (Fig. 3C). Such a strong interfacial bonding strength and the inherent structural flexibility make SAO_T a versatile structure template for the coherent growth of various oxide films. From the perspective of water-soluble sacrificial layers, it could be the key to minimizing the interfacial misfit strain and improving the quality of exfoliated freestanding oxide membranes.

Freestanding oxide membranes released from SAO_T

We examined the potential of SAO_T as a water-soluble sacrificial layer. An "optimal" water-soluble sacrificial layer must satisfy three key requirements. It must enable the successful growth of target oxide films and maintain the high crystallinity and integrity in the released freestanding membranes. Additionally, the representative functionalities of the exfoliated freestanding oxide membranes should be comparable to the epitaxial counterparts. And lastly, it should dissolve easily in water, allowing efficient membrane exfoliation within reasonable durations. To evaluate these criteria for the SAO_T film, we grew several typical perovskite oxide films on both SAO_T and SAO_C epitaxial films and then conducted comparative studies on the integrity, crystallinity, functionalities, and exfoliation speed of the freestanding membranes.

We first performed comparative characterizations on the integrity and crystallinity of perovskite oxide membranes released from both SAO_C and SAO_T. As summarized in table S3, these oxides include $NdNiO_3$ (NNO), $LaNiO_3$ (LNO), $La_{0.7}Ca_{0.3}MnO_3$ (LCMO), $SrTiO_3$ (STO), $SrRuO_3$ (SRO), $SrSnO_3$ (SSO), and $BaTiO_3$ (BTO), with a broad range of bulk lattice constants (a_p, in pseudocubic notation) from 3.81 to 4.04 Å. To minimize the lattice mismatch between SAO and target oxides, we choose to grow the ABO_3/SAO bilayers on either LSAT(001) or STO(001) substrates. As depicted in Fig. 4A, we used standard polydimethylsiloxane (PDMS)–assisted release and transfer procedures of freestanding oxide membranes from both SAO_C and SAO_T (*40*). For the FE BTO membranes, the inherent superelasticity accommodates stress and deformations generated during the lift-off process. Hence, the choice of water-soluble sacrificial layers (SAO_C or SAO_T) becomes less critical for the crystallinity and integrity (fig. S21) (*21*, *22*, *43*). Nevertheless, for the other non-FE oxides, the membranes released from SAO_C and SAO_T show pronounced differences. According to the optical microscopic images (Fig. 4B), the freestanding oxide membranes released from

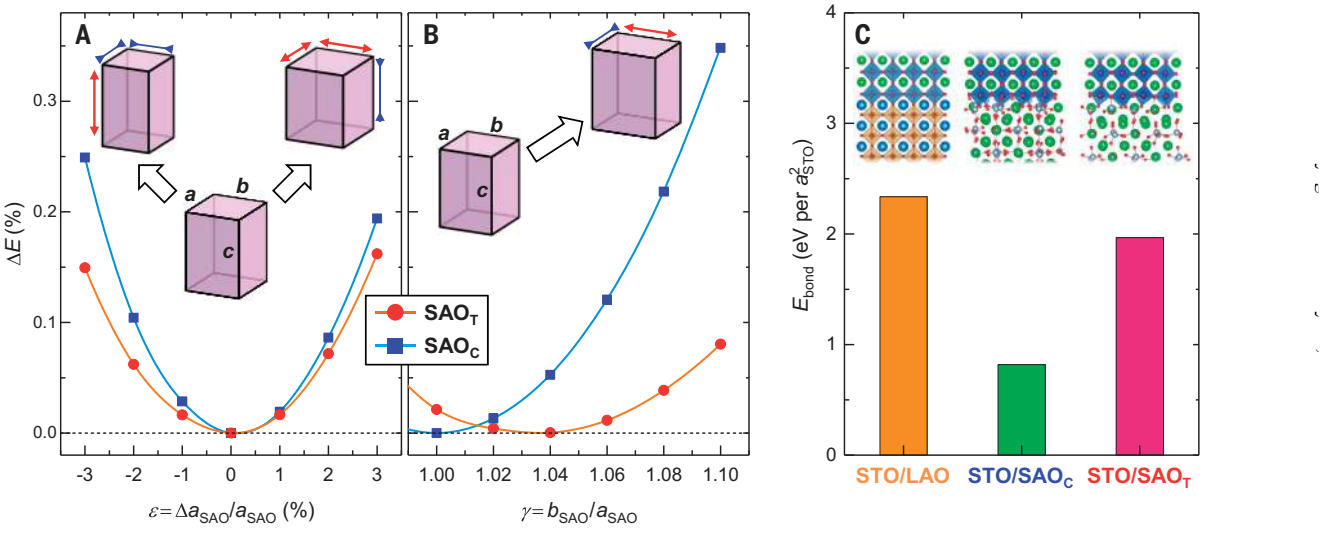


Fig. 3. Strain-related density functional theory calculations on SAO_C and SAO_T unit cells. (**A**) DFT-calculated energy changes (ΔE) of SAO_C and SAO_T unit cells under biaxial strain ε, determined from the relative in-plane lattice constant change ($\Delta a_{SAO}/a_{SAO}$). For direct comparison between SAO_C and SAO_T unit cells, ΔE values are normalized by the total energy of the relaxed unit cells. Both ΔE-ε curves display parabolic trends and local minima at ε = 0. Notably, the ΔE of SAO_T unit cell is consistently smaller than that of the SAO_C unit cell, particularly under compressive strain (ε < 0). (**B**) DFT-calculated ΔE of SAO_C and SAO_T unit cells under anisotropic strain γ, determined by the orthorhombicity (b_{SAO}/a_{SAO}). The ΔE-γ curve of SAO_C unit cell shows a local minimum at γ = 1, whereas the curve of SAO_T unit cell shows a local minimum at γ = 1.04, consistent with its inherent orthorhombic symmetry. Consistently, the ΔE of the SAO_T unit cell shows a much weaker dependence on the anisotropic strain γ. (**C**) DFT-calculated interfacial bonding energy (E_{bond}) of the LAO/STO(001), SAO_C/STO(001), and SAO_T/STO(001) heterostructures. The E_{bond} values are normalized by the number of STO unit cells bonded at the heterointerface. The insets of (C) are schematics of the three interface structures used for DFT calculation.

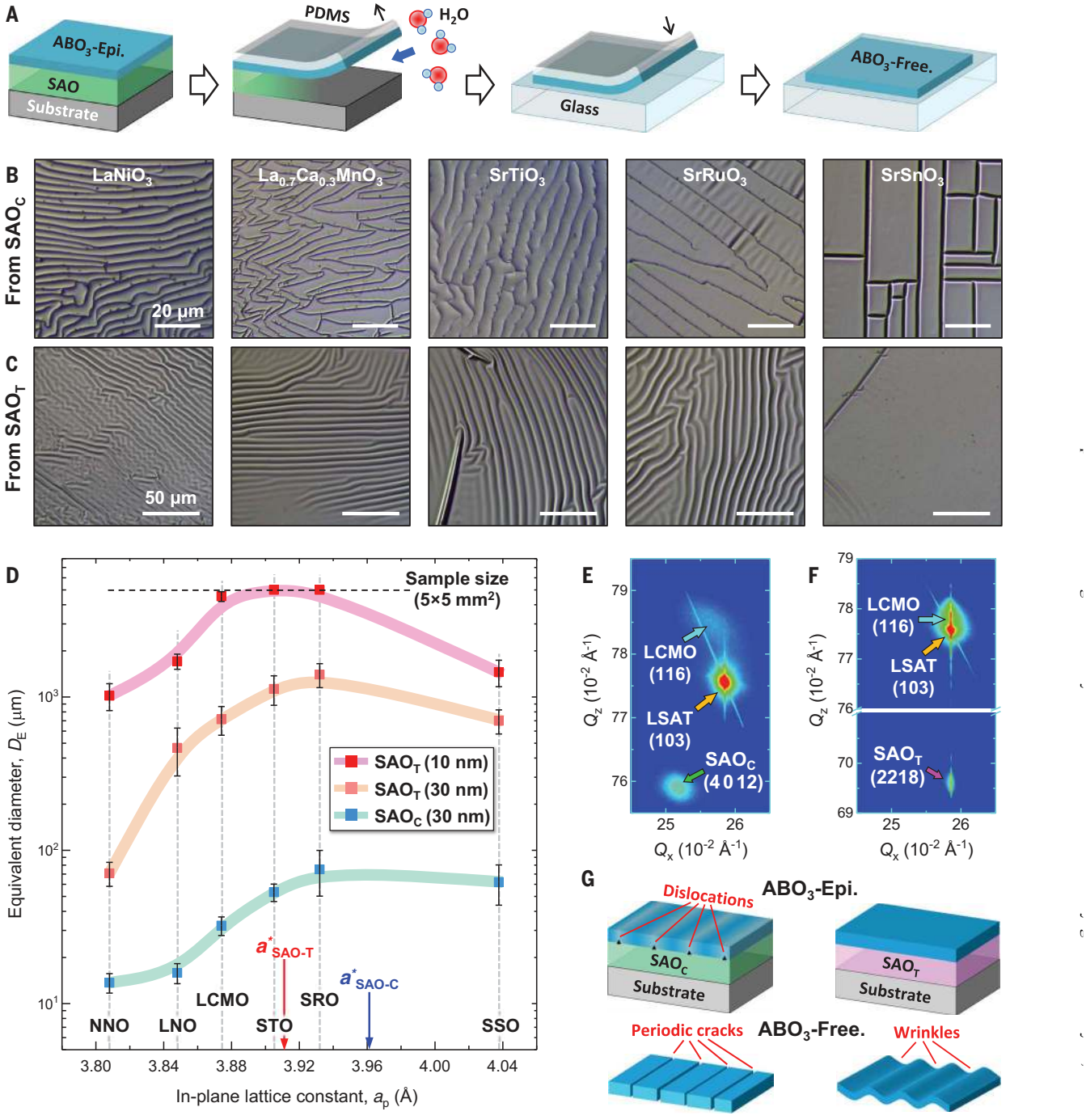


Fig. 4. Freestanding oxide membranes released from SAO_C and SAO_T. (**A**) Schematic illustration of PDMS and water-assisted exfoliation of freestanding oxide membranes from SAO_C and SAO_T sacrificial layers. (**B** and **C**) Optical microscopic images of 35-nm-thick $LaNiO_3$ (LNO), $La_{0.7}Ca_{0.3}MnO_3$ (LCMO), $SrTiO_3$ (STO), $SrRuO_3$ (SRO), and $SrSnO_3$ (SSO) films peeled from (B) SAO_C and (C) SAO_T layers. All of the freestanding oxide membranes exfoliated from SAO_C show periodic and high-density cracks. By contrast, the membranes exfoliated from SAO_T show large-scale crack-free but wrinkled morphology. (**D**) Summary of averaged equivalent diameter $D_E = A_{free}^{1/2}$, where the A_{free} is the uncracked area of the freestanding membranes. The averaged D_E is plotted as a function of the in-plane lattice constant in pseudocubic perovskite notation (a_p), and the error bars represent the standard deviations of D_E from five membranes. The D_E values for $NdNiO_3$ (NNO) membranes are also included in (D). The in-plane lattice constants of bulk SAO_T and SAO_C in pseudocubic perovskite notations ($a^*_{SAO\text{-}T}$ and $a^*_{SAO\text{-}C}$) are marked in (D). The LCMO, STO, and SRO membranes released from 10-nm-thick SAO_T are almost crack-free. Thus, the D_E value approaches the sample size (5 mm, marked by a horizontal dashed line). (**E** and **F**) RSMs around LCMO(116) diffractions measured from LCMO epitaxial film grown on (E) SAO_C/LSAT(001) and (F) SAO_T/LSAT(001). In (E), the diffused LCMO (116) diffraction spans between the LSAT(116) and SAO_C(4012), indicating a partial strain relaxation. In (F), the in-plane reciprocal space vector (Q_x) values of LCMO(116) and SAO_T(4018) align with that of the LSAT(103), indicating a coherent strain state of the LCMO/SAO_T/LSAT(001) heterostructure. (**G**) Schematics elucidating the crack and wrinkle formations in freestanding oxide membranes. The perovskite oxide epitaxial films and freestanding membranes are marked as "ABO_3-Epi." and "ABO_3-Free.," respectively.

30-nm SAO_C exhibit high-density and periodic cracks, which are qualitatively similar to the morphologies shown in the literature (*24*, *37*). In contrast, the oxide membranes released from 30-nm SAO_T (Fig. 4C) show crack-free regions spanning up to several hundred micrometers in scale. For the membranes released from 10-nm SAO_T, the crack-free region can be expanded further, to a few millimeters in scale (figs. S22 to S28), which could be attributed to the improved structural coherency and slower release speed [see section 4 of (*40*)]. Notably, the optical microscopic images reveal micrometer-scale and periodic wrinkling morphologies in the crack-free regions, which were commonly observed in the FE oxide membranes with superelasticity. Accordingly, the freestanding oxide membranes released from SAO_T should possess superior integrity and flexibility, even withstanding the large lattice deformation in the wrinkled microstructures.

To quantify the SAO_T-induced integrity improvements in oxide membranes, we calculated the average equivalent diameters (D_E) of the uncracked areas in the freestanding membranes. The D_E versus a_p curves are summarized in Fig. 4D. Within a broad a_p range of 3.81 to 4.04 Å, the D_E values of membranes released from 10 nm (30 nm) SAO_T reach a few millimeters (hundreds of micrometers) in scale, which are orders of magnitude higher than the corresponding D_E values in the SAO_C case. According to the aforementioned structural differences between SAO_C and SAO_T, we speculate that the strain coherency at ABO_3/SAO interfaces may play a crucial role in determining the crystallinity and integrity of these freestanding membranes. Taking LCMO/SAO/LSAT(001) as a model system, we verified this hypothesis by detailed strain analyses. For LCMO film grown on SAO_C/LSAT(001), RSM near LSAT(103) diffraction (Fig. 4E) shows a partial strain relaxation at the LCMO/SAO_C interfaces. The weak and broad LCMO(116) diffractions of the exfoliated freestanding LCMO membrane (fig. S22) further signify an unsatisfactory crystallinity. In contrast, RSM characterizations (Fig. 4F) from the LCMO/SAO_T/LSAT(001) bilayer demonstrate a coherent strain state and high epitaxial quality. The strong and sharp LCMO(116) Bragg diffraction of the freestanding LCMO membranes confirms a persistent high crystallinity even after water-assisted exfoliation (fig. S22).

The strong correlation between the high integrity of oxide membranes and the coherent strain state at the ABO_3/SAO interface can be understood by a simple scenario (Fig. 4G). For the ABO_3/SAO_C epitaxial heterostructures, the robust cubic lattice of SAO_C inhibits the epitaxial strain propagation from the substrate to ABO_3 epitaxial films. The unavoidable lattice mismatch between ABO_3 and SAO_C must be accommodated by the formation of periodic dislocations (*34*, *36*). During the water-assisted exfoliation process, these defects inevitably rupture the film lattice, leading to the formation of periodic cracks. The correlation between lattice mismatch and membrane integrity is further implied by the steep slope-like line shape of the D_E-a_p curve. For the ABO_3/SAO_T heterostructures, on the contrary, the inherent structural flexibility of SAO_T and strong interfacial bonding enable a coherent strain state. The resultant sharp and dislocation-free ABO_3/SAO_T interface should effectively hinder crack formation during exfoliation. In line with this picture, wrinkle formation is primarily driven by the release of compressive strain during membrane exfoliation (fig. S29).

We characterized the evolutions of physical properties in various oxide membranes released from SAO_C and SAO_T. For the ferromagnetic metal LCMO/SAO_T/LSAT(001) film and corresponding membrane, the temperature-dependent magnetization (*M*-*T*) and resistivity (ρ-*T*) curves reveal a sharp paramagnetic insulator (PMI) to ferromagnetic metal (FMM) transition (Fig. 5, A and B). The Curie temperature (T_C), saturated magnetization, and residual resistivity are comparable to those of the LCMO/LSAT(001) epitaxial films, consistent with the observed high crystallinity and integrity. In contrast, for both the LCMO/SAO_C/LSAT(001) epitaxial film and the corresponding LCMO membrane, the PMI-to-FMM transition becomes more slanted. These degradations in ferromagnetism and metallicity can be attributed to the residual tensile strain at the LCMO/SAO_C interface and to the high-density cracks formed during exfoliation (Fig. 2) (*24*, *44*). The itinerant ferromagnet SRO films also show a similar trend. The *M*-*T* and ρ-*T* curves of the membranes released from SAO_T reveal a sharp FMM transition at T_C ~ 150 K. The residual resistance ratio (RRR) value is up to 4.83, comparable with PLD-grown SRO/STO(001) epitaxial films (Fig. 5, C to H) (*6*, *45*, *46*). The SRO membrane also exhibits a

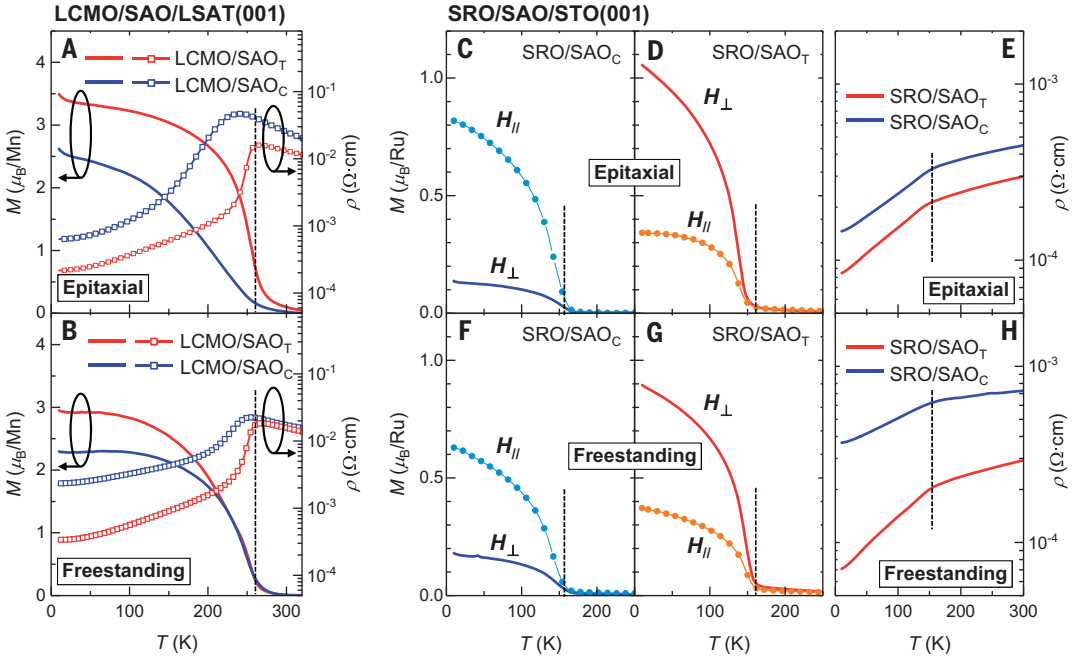


Fig. 5. Physical properties of LCMO and SRO epitaxial films and freestanding membranes. (**A**) Temperature-dependent resistivity (ρ-*T*) and magnetization (*M*-*T*) curves measured from the LCMO(35 nm)/SAO_C/LSAT(001) and LCMO(35 nm)/SAO_T/LSAT(001) epitaxial films. (**B**) ρ-*T* and *M*-*T* curves measured from the freestanding LCMO membranes exfoliated from SAO_C and SAO_T. The *M*-*T* curves were measured with an in-plane magnetic field $\mu_0 H$ = 1000 Oe. (**C** to **E**) *M*-*T* [(C) and (D)] and ρ-*T* (E) curves measured from the SRO(35 nm)/SAO_C/STO(001) and SRO(35 nm)/SAO_T/STO(001) epitaxial films. (**F** to **H**) *M*-*T* [(F) and (G)] and ρ-*T* (H) curves measured from the SRO freestanding membranes released from SAO_C and SAO_T. All the *M*-*T* curves were measured with a magnetic field of 1000 Oe, applied both parallel ($H_{//}$) or perpendicular ($H_\perp$) to the film plane. The dashed lines are guidelines for the Curie temperature of LCMO and SRO.

strong perpendicular magnetic anisotropy (MA) dominated by the intrinsic magnetocrystalline anisotropy (*47*), which also signifies a high crystallinity. In contrast, the SRO membrane exfoliated from SAO_C shows a demagnetization-dominated in-plane MA and a reduced RRR (down to 1.93), which are consistent with the degradation of crystallinity and crack formation. Following the same scenario, we also characterized the electrical transport of LNO and NNO membranes (figs. S30 and S31). Consistently, their representative transport properties in membranes released from SAO_T are well maintained or even improved. In brief, SAO_T can universally ensure both the high integrity and epitaxial film–like functionalities of oxide membranes.

Finally, we examined the exfoliation efficiency of oxide membranes from SAO_T. To perform an equitable comparison between SAO_C and SAO_T, we opted to exfoliate the 50-nm-thick BTO membranes. Due to intrinsic superelasticity, the exfoliation speed of BTO membranes should not be influenced by crack formation and associated extrinsic water penetration. According to the in situ monitoring by optical microscope (movies S1 and S2), the water-assisted exfoliation speed of BTO membranes released from SAO_T is approximately one order of magnitude faster than that released from SAO_C. The trend of faster exfoliation is universally applicable for all the other oxide membranes we grew (fig. S32), signifying a higher water solubility of SAO_T than SAO_C. The dissolution speed of SAO_T film also highly depends on the film thickness and Ca or Ba doping (fig. S33), which provides independent parameters for simultaneously optimizing the exfoliation efficiency and quality. The high water solubility also has a structural origin. As depicted in fig. S34, the Al-O networks in SAO_T comprise discrete AlO_4^{5-} and $Al_3O_{10}^{11-}$ groups, which hydrolyze in water more readily than the $Al_6O_{18}^{18-}$ rings in SAO_C (*14*, *42*, *48*). Despite such a high water solubility, the SAO_T film exhibits exceptional stability against ambient moisture over 40 days when incorporating an ultrathin STO protective layer (fig. S35). Consistently, this long-term stability of SAO_T could be attributed to the high epitaxial quality of the ABO_3/SAO_T heterostructure (*49*).

Conclusions

We identified the super-tetragonal SAO_T as a promising water-soluble sacrificial layer with several compelling advantages. First, SAO_T showcases remarkable structural flexibility to adapt the epitaxial strain imposed by various perovskite substrates, providing wide-range tunability of in-plane lattice constant. Such inherent structural flexibility further ensures high-quality epitaxy of a broad spectrum of ABO_3/SAO_T heterostructures with coherent strain states and dislocation-free interfaces, which restrain crack formation during water-assisted exfoliation. For various non-FE oxide membranes with lattice constants ranging from 3.85 to 4.04 Å, the crack-free areas span up to a few millimeters in scale. Moreover, the strain tunability of SAO_T persists with Ba or Ca doping and (110)-oriented epitaxial growth, broadening its potential in developing novel freestanding oxides beyond traditional perovskites. Next, the SAO_T has a wide and stable growth window, accessible for standard PLD techniques and compatible with the growth of most perovskite oxides. Lastly, its high water solubility streamlines the membrane exfoliation process. With these attributes, the SAO_T sacrificial layer offers a versatile and feasible experimental approach to producing large-scale, crack-free freestanding oxide membranes, the crystallinity and functionalities of which are comparable to those of the epitaxial films. The discovery of SAO_T introduces a pivotal complement to the widely used SAO_C sacrificial layer, which may substantially promote the potential of freestanding oxide membranes for innovative, low-dimensional, and flexible device applications (*18*–*20*).

REFERENCES AND NOTES

1. H. Y. Hwang *et al*., *Nat. Mater.* **11**, 103–113 (2012).
2. P. Yu, Y. H. Chu, R. Ramesh, *Mater. Today* **15**, 320–327 (2012).
3. A. Ohtomo, H. Y. Hwang, *Nature* **427**, 423–426 (2004).
4. A. D. Caviglia *et al*., *Nature* **456**, 624–627 (2008).
5. E. Bousquet *et al*., *Nature* **452**, 732–736 (2008).
6. L. Wang *et al*., *Nat. Mater.* **17**, 1087–1094 (2018).
7. S. Das *et al*., *Nature* **568**, 368–372 (2019).
8. J. Mannhart, D. G. Schlom, *Science* **327**, 1607–1611 (2010).
9. S. Manipatruni *et al*., *Nature* **565**, 35–42 (2019).
10. V. Garcia, M. Bibes, *Nat. Commun.* **5**, 4289 (2014).
11. S. R. Bakaul *et al*., *Nat. Commun.* **7**, 10547 (2016).
12. H. Lu *et al*., *Nat. Commun.* **5**, 5518 (2014).
13. F. M. Chiabrera *et al*., *Ann. Phys.* **534**, 2200084 (2022).
14. D. Lu *et al*., *Nat. Mater.* **15**, 1255–1260 (2016).
15. Y. H. Chu, *NPJ Quantum Mater.* **2**, 67 (2017).
16. H. S. Kum *et al*., *Nature* **578**, 75–81 (2020).
17. A. Sambri *et al*., *Adv. Funct. Mater.* **30**, 1909964 (2020).
18. A. J. Yang *et al*., *Nat. Electron.* **5**, 233–240 (2022).
19. B. S. Y. Kim, Y. Hikita, T. Yajima, H. Y. Hwang, *Nat. Commun.* **10**, 5312 (2019).
20. D. Lu, S. Crossley, R. Xu, Y. Hikita, H. Y. Hwang, *Nano Lett.* **19**, 3999–4003 (2019).
21. G. Dong *et al*., *Science* **366**, 475–479 (2019).
22. B. Peng *et al*., *Sci. Adv.* **6**, eaba5847 (2020).
23. D. Ji *et al*., *Nature* **570**, 87–90 (2019).
24. S. S. Hong *et al*., *Science* **368**, 71–76 (2020).
25. R. Xu *et al*., *Nat. Commun.* **11**, 3141 (2020).
26. P.-C. Wu *et al*., *Nat. Commun.* **13**, 2565 (2022).
27. S. Chen *et al*., *Adv. Mater.* **35**, e2206961 (2023).
28. L. Han *et al*., *Nature* **603**, 63–67 (2022).
29. Y. Liu, Y. Huang, X. Duan, *Nature* **567**, 323–333 (2019).
30. G. Kim *et al*., *Adv. Mater.* **35**, e2203373 (2023).
31. Y. Li *et al*., *Adv. Mater.* **34**, e2106826 (2022).
32. B. Zhang, C. Yun, J. L. MacManus-Driscoll, *Nano-Micro Lett.* **13**, 39 (2021).
33. Q. Wang *et al*., *Crystals* **10**, 733 (2020).
34. D. J. Baek, D. Lu, Y. Hikita, H. Y. Hwang, L. F. Kourkoutis, *APL Mater.* **5**, 096108 (2017).
35. S. S. Hong *et al*., *Sci. Adv.* **3**, eaao5173 (2017).
36. D. Lu *et al*., *Phys. Status Solidi Rapid Res. Lett.* **12**, 1700339 (2018).
37. P. Singh *et al*., *ACS Appl. Electron. Mater.* **1**, 1269–1274 (2019).
38. Y. Bourlier *et al*., *ACS Appl. Mater. Interfaces* **12**, 8466–8474 (2020).
39. H. Peng *et al*., *Adv. Funct. Mater.* **32**, 2111907 (2022).
40. Materials and methods are available as supplementary materials.
41. A. J. Hatt, N. A. Spaldin, C. Ederer, *Phys. Rev. B* **81**, 054109 (2010).
42. V. Kahlenberg, *Mineral. Mag.* **65**, 533–541 (2001).
43. C. Jin *et al*., *Adv. Sci. (Weinh.)* **8**, e2102178 (2021).
44. G. Gao, S. Jin, W. Wu, *Appl. Phys. Lett.* **90**, 012509 (2007).
45. H. G. Lee *et al*., *Adv. Mater.* **32**, e1905815 (2020).
46. J. Lu *et al*., *Adv. Mater.* **33**, e2102525 (2021).
47. G. Koster *et al*., *Rev. Mod. Phys.* **84**, 253–298 (2012).
48. V. Kahlenberg, B. Lazić, S. V. Krivovichev, *J. Solid State Chem.* **178**, 1429–1439 (2005).
49. D. Li *et al*., *Nano Lett.* **21**, 4454–4460 (2021).

ACKNOWLEDGMENTS

The sample fabrication was partially carried out at the University of Science and Technology of China (USTC) Center for Micro and Nanoscale Research and Fabrication. Magnetic and electric characterizations were partially carried out at the Instruments Center for Physical Science, USTC, with the assistance of J. Zhao, X. Liu, and M. Ge. Z.Lu. acknowledges support from BL02U2 at the Shanghai Synchrotron Radiation Facility for the synchrotron x-ray diffraction characterizations. L.S. acknowledges support from the Vienna Scientific Cluster for the first-principles calculations. L.W. acknowledges the assistance of Anhui Epitaxy Technology Co. Ltd. with pulsed laser deposition and the Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, with target preparations. All authors appreciate the invaluable discussions and comments from T. W. Noh, D. Li, D. Lu, Y. Nie, and M. Yang. **Funding:** This work was supported by the National Key Research and Development Program of China (grants 2020YFA0309100, 2023YFA1406404, 2022YFA1603902, and 2019YFA0308500), the National Natural Science Foundation of China (grants 12074365, 12374094, 12304153, U2032218, 11974326, 12175235, 52072400, 52025025, and 51991344), the CAS Project for Young Scientists in Basic Research (YSBR-084), the Fundamental Research Funds for the Central Universities (grants WK9990000102 and WK2030000035), the Anhui Provincial Natural Science Foundation (grant 2308085MA15), the Hefei Science Center CAS Foundation (2021HSC-CIP017 and 2016HSC-IU06), the Beijing Natural Science Foundation (Z190010), and the Technical Talent Promotion Plan (TS2021002). **Author contributions:** L.W. and W.W. conceived of the idea and supervised the project. J.Z., T.L., A.W., and X.W. contributed equally to this work. J.Z., L.S., E.-J.G., L.W., and W.W. prepared the manuscript. J.Z., A.W., H.Y., Q.W., and S.C. prepared the epitaxial films and freestanding membranes. J.Z., A.W., Z.Li., F.J., and J.L. conducted structure and physical property characterizations. T.L., Q.Z., and L.G. conducted the STEM measurements. X.W. and L.S. conducted the DFT calculations. Q.H. and Z.Lu. performed the 3D-RSM measurements. M.F. performed the energy-dispersive spectroscopy characterizations. E.-J.G. supervised the second harmonic generation experiments and analyzed the data. All authors were involved in the data analyses and manuscript preparation. **Competing interests:** L.W., J.Z., and A.W. are inventors on a Chinese patent application on the $Sr_4Al_2O_7$ growth recipe (2023115077984) submitted by the University of Science and Technology of China. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:**

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adi6620
Materials and Methods
Supplementary Text
Figs. S1 to S35
Tables S1 to S3
References (*50*–*75*)
Movies S1 and S2

Submitted 18 May 2023; resubmitted 25 August 2023
Accepted 13 December 2023
10.1126/science.adi6620

ORGANIC CHEMISTRY

Catalytic enantioselective reductive Eschenmoser-Claisen rearrangements

Guoting Zhang[1], Matthew D. Wodrich[2], Nicolai Cramer[1]*

An important challenge in enantioselective catalysis is developing strategies for the precise synthesis of neighboring congested all-carbon quaternary stereocenters. The well-defined transition states of [3,3]-sigmatropic rearrangements and their underlying stereospecificity render them powerful tools for the synthesis of such arrays. However, this type of pericyclic reaction remains notoriously difficult to catalyze, especially in an enantioselective fashion. Herein, we describe an enantioselective reductive Eschenmoser-Claisen rearrangement catalyzed by chiral 1,3,2-diazaphospholene-hydrides. This developed transformation enables full control of the two newly formed acyclic stereogenic centers, leading to amides with vicinal all-carbon quaternary-tertiary or quaternary-quaternary carbon atoms.

The biological and physical properties of organic molecules are intimately linked to their stereochemical configurations. Enantioselective catalysis has emerged as a practical technology for stereoselective efficient synthesis (*1*). Quaternary all-carbon stereocenters are a commonly occurring pattern of many natural products and biologically active agents and bear particular synthetic challenges (*2*). Compared with assembling an isolated quaternary carbon stereogenic center, arrays of adjacent chiral carbon atoms, including quaternary-tertiary centers (*3*) and vicinal quaternary centers (*4*–*6*), invariably increase the synthetic difficulty. Claisen-[3,3]-sigmatropic rearrangements have proven a valuable method for the diastereoselective construction of such congested vicinal stereogenic centers due to the predictable chair-like transition state (Fig. 1A) (*7*, *8*). High-fidelity transfer of configurations in substrate-embedded stereogenic centers contributes to the popularity of these rearrangements (*9*). However, the search for catalytic enantioselective variants of this transformation has been a challenging objective for several decades.

The effort to construct vicinal stereogenic centers containing quaternary carbons by catalytic asymmetric Claisen rearrangements was successful only on a case-by-case basis (*10*, *11*), thus remaining a formidable and largely unmet challenge (Fig. 1B). In addition to

[1]Laboratory of Asymmetric Catalysis and Synthesis, Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.
[2]Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.
*Corresponding author. Email: nicolai.cramer@epfl.ch

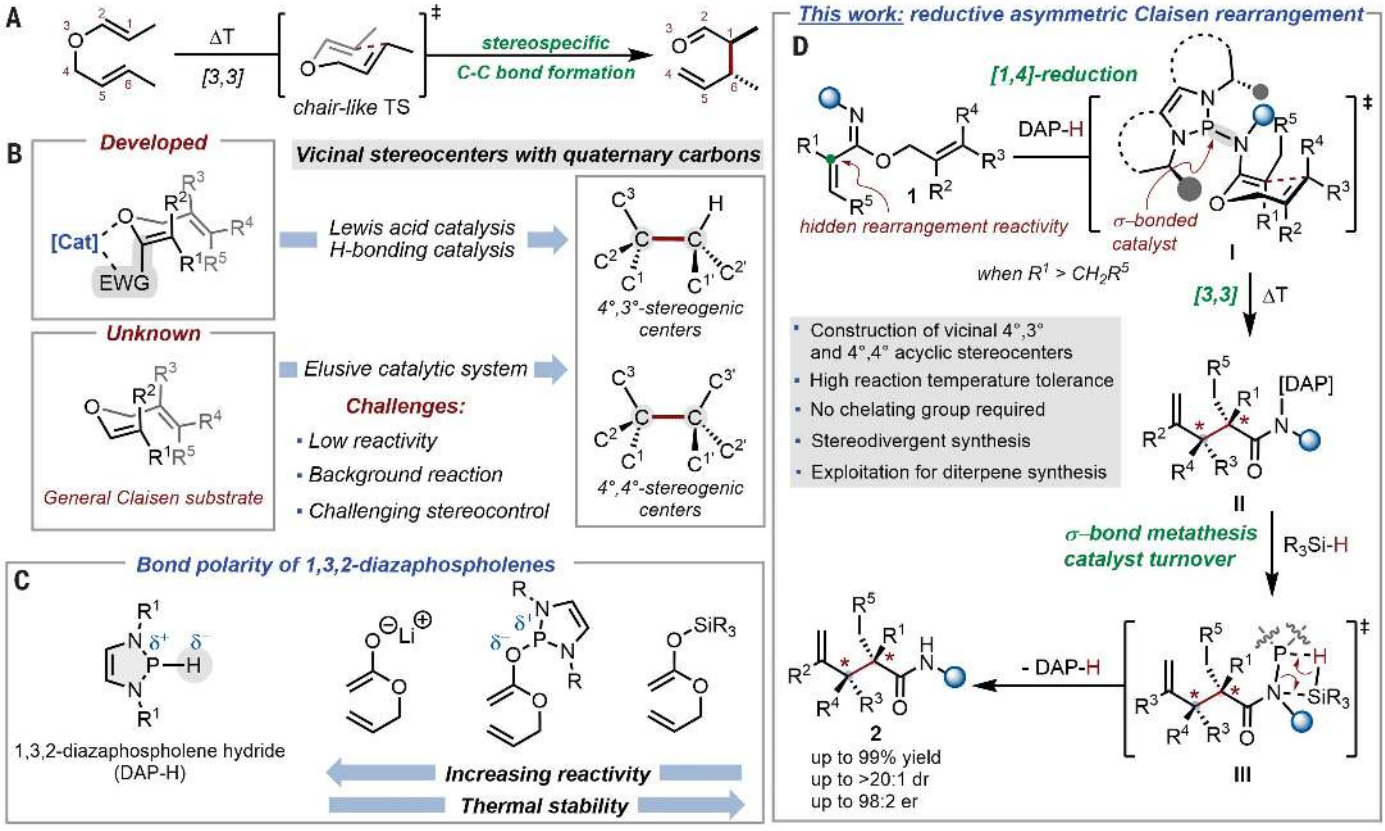
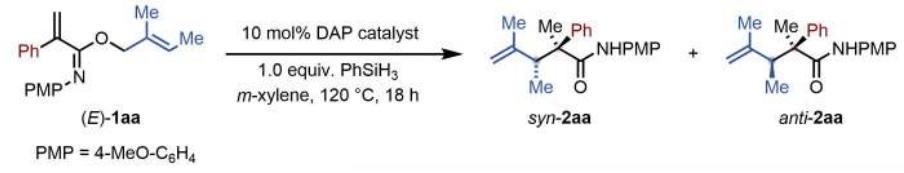
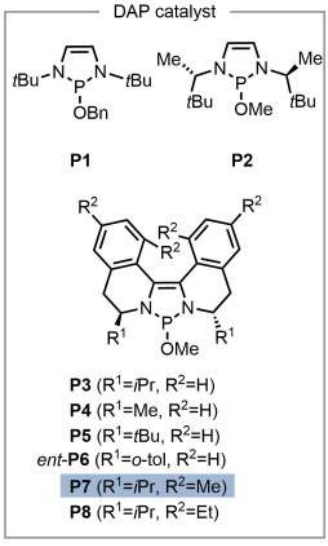


Fig. 1. Catalytic enantioselective Claisen rearrangements to access vicinal stereocenters containing chiral quaternary carbons. (**A**) Classical uncatalyzed thermal stereoselective Claisen rearrangement. (**B**) Existing approaches and unmet challenges for catalytic enantioselective Claisen rearrangements to construct vicinal carbon stereogenic centers with quaternary carbon atoms. (**C**) Bond polarity of DAP. (**D**) Present transformation: DAP-catalyzed enantio- and diastereoselective reductive Eschenmoser-Claisen rearrangement process.





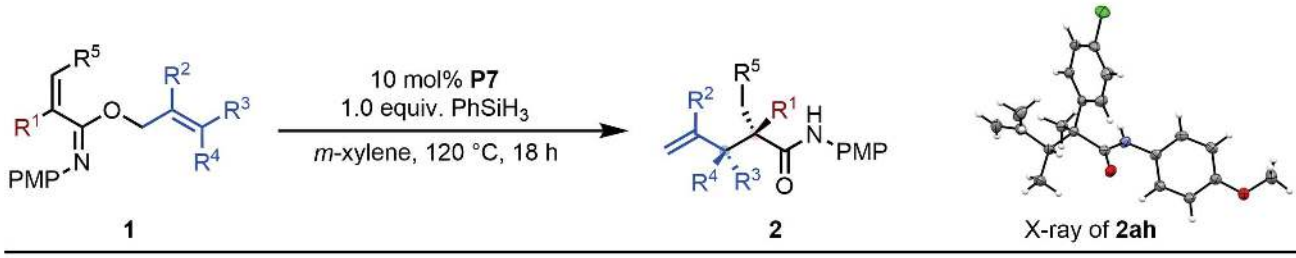
Entry	DAP catalyst	Yield (%)	*syn*:*anti*	er (*syn*-**2aa**)
1	**P1**	98	83:17	-
2	**P2**	58	80:20	25:75
3	**P3**	70	85:15	87:13
4	**P4**	99	81:19	75:25
5	**P5**	<5	-	-
6	*ent*-**P6**	76	75:25	19:81
7	**P7**	92	92:8	97.5:2.5
8	**P8**	42	86:14	91:9
	Change from entry 7			
9	without **P7** or $PhSiH_3$	0	-	-
10	with 0.33 equiv. $PhSiH_3$	91	89:11	97.5:2.5
11	HBPin instead of $PhSiH_3$	85	90:10	96:4
12	Ph_2SiH_2 instead of $PhSiH_3$	<5	-	-
13	at 60 °C	13	82:18	70:30
14	at 80 °C	43	92:8	97.5:2.5
15	at 130 °C	96	88:12	97.5:2.5

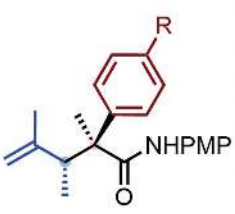
Fig. 2. Development and optimization of DAP-catalyzed reductive enantioselective Eschenmoser-Claisen rearrangement.

enantio induction with stoichiometric additives (*12*, *13*), pioneering catalytic strategies generally involved the bidentate interactions of substrate with the catalyst, such as a chiral Lewis acid (*14*–*18*) or a hydrogen-bond donor (*19*, *20*). The catalytic performance is tightly linked to a bidentate coordination ability, often engineered into the substrate through an activating carbonyl group in addition to the embedded oxygen atom enabling the stereocontrol but largely limiting scope and hence the synthetic applicability. A stepwise Ru(II)-catalyzed formal asymmetric Claisen rearrangement operating by π-allyl intermediates with a limitation to vicinal tertiary stereogenic centers was reported (*21*). The core problems arising with unbiased substrates lacking chelating functionality include poor reactivity, strong uncatalyzed background reactions, and weak influences on stereoselectivity. Principally, increasing the reaction temperature helps to overcome the higher activation barrier induced by the forced proximity of two sterically crowded olefins. However, the flexibility and relative weakness of the abovementioned noncovalent bonding interactions lead to severe losses of stereocontrol under harsh conditions and increasing competition from strong background reactivity. Naturally, a temporarily covalently bound catalytic system with thermal stability would more likely overcome sluggish rearrangements at higher reaction temperature while maintaining stereocontrol. For example, the chiral auxiliary strategy represents a feasible approach for stereoselective Claisen rearrangements (*9*). However, its major drawback is the lack of turnover for the bound chiral units and the additional steps required for installation and removal. Preliminary advances on covalently bound catalytic systems have been achieved by chiral *N*-heterocyclic carbene catalysis (*22*) and gold-catalyzed tandem Claisen rearrangements (*23*, *24*), although the very specific turnover pathways render them narrowly applicable for the specific substrate structures and unsuitable for constructing vicinal stereogenic carbon centers.

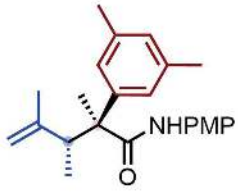
We reasoned that appropriate covalently bound robust chiral catalyst enabling well-defined and rigid transition states at higher reaction temperatures paired with a suitable catalyst turnover strategy would be of pivotal importance for a general enantioselective Claisen rearrangement. Our recent investigation of tandem pericyclic rearrangement reactions triggered by 1,3,2-diazaphospolene (DAP)–catalyzed conjugate reduction of α,β-unsaturated carboxylic acid derivatives (*25*, *26*) demonstrated the considerable potential of DAP catalysis to address this gap in enantioselective catalysis. To date, chiral DAP catalysts have only been used for enantioselective reductions in which the hydride transfer was the enantio-determining step (*27*–*29*). Because of the characteristic σ-aromaticity (*30*, *31*), one of the key exploitable characteristics of *P*-hydrido-1,3,2-diazaphospholenes (DAP-H) is their behavior as highly nucleophilic molecular hydrides with low basicity, making them highly competent for the reduction of various polarized unsaturated bonds (*32*–*35*). Analogous to the anion-accelerated Claisen rearrangement strategy (*36*), the polarized nature of the DAP-bound enolate bond also facilitates the sigmatropic rearrangement. The DAP-enolates have the advantage of not suffering from the inherent thermal lability of typical metal ester enolates, being stable at a broader range of temperatures. Conversely, σ-bond metathesis of the heteroatom-substituted DAP species with a terminal reductant enables the facile regeneration of DAP-H (*37*). On the basis of these considerations, we designed a catalytic reductive enantioselective Claisen rearrangement reaction by using a chiral DAP hydride (Fig. 1D).

We opted to investigate the Eschenmoser-Claisen rearrangement using allyl acryloyl imidates **1** as suitable acrylic ester surrogates. The Eschenmoser variant provides valuable amides instead of carboxylic acid or aldehyde products of the classical Claisen reaction and benefits from a superior tunability and selectivity control through its additional variable nitrogen substituent. The transformation is initiated by the formation of an active DAP-H species from the DAP-alkoxide precatalyst by σ-bond metathesis with phenylsilane. The envisioned transformation starts with a selective conjugate reduction of acryloyl imidates by the DAP-H moiety, generating *N*-DAP–bound *N*,*O*-ketene acetal **I** ideally possessing a defined double bond geometry. This armed intermediate **I** is set for the targeted [3,3]-sigmatropic rearrangement, specifically an Eschenmoser-Claisen rearrangement. The covalently bound chiral catalyst is anticipated to efficiently control the facial selectivity through a favorable chair-like transition state. Upon rearrangement, *N*-DAP–substituted amide **II** was generated with high precision of the vicinal array of stereogenic centers. The catalytic cycle is closed by σ-bond metathesis with silane, releasing DAP-H and *N*-silylated amide **III** that rapidly undergoes hydrolysis upon contact with silica gel. Overall, varieties of amides with acyclic contiguous stereogenic quaternary carbons (4°,3° and 4°,4°) are efficiently obtained with high stereoselectivities. The strategy of the reductively induced process suppresses any background rearrangement by only revealing the reactive [3,3]-pattern upon covalent installment of the chiral control element. The robust and rigid transition state of this process is not only capable of maintaining high selectivities at high reaction

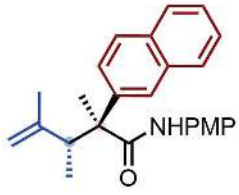




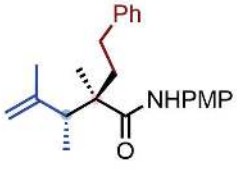
2aa (R=H), 92 %, 12:1 dr, 97.5:2.5 er
2ab (R=OMe), 77 %, 8:1 dr, 98:2 er
2ac (R=Me), 94 %, 9:1 dr, 97:3 er
2ad (R=CF_3), 87 %, 12:1 dr, 96:4 er
2ae (R=CN), 86 %, 12:1 dr, 97:3 er
2af (R=F), 84 %, 11:1 dr, 98:2 er
2ag (R=Br), 90 %, 14:1 dr, 98:2 er
2ah (R=Cl), 92 %, 11:1 dr, 96:4 er



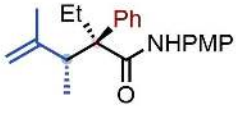
2ai, 78 %, 8:1 dr, 95:5 er



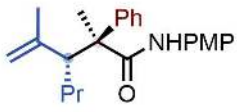
2aj, 93 %, >20:1 dr, 97.5:2.5 er



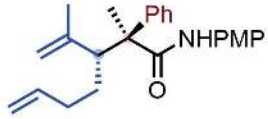
2ak, 69 %, 3:1 dr, 95:5 er



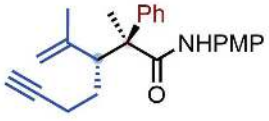
2al, 68 %, 8:1 dr, 97:3 er



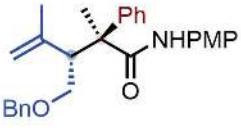
2am, 90 %, 8:1 dr, 97:3 er



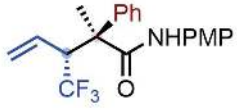
2an, 93 %, 8:1 dr, 97:3 er



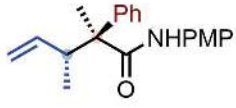
2ao, 98 %, 8:1 dr, 96:4 er



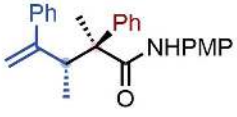
2ap, 89 %, 12:1 dr, 96:4 er



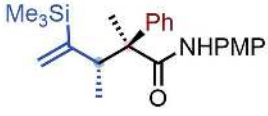
2aq, 75 %, 2:1 dr, 90:10 er



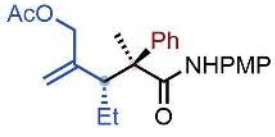
2ar, 79 %, 3:1 dr, 96:4 er



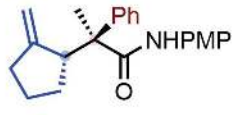
2as, 76 %, 5:1 dr, 97:3 er



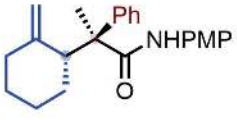
2at, 94 %, 5:1 dr, 97:3 er



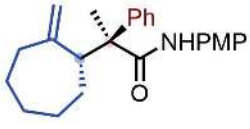
2au, 86 %, 8:1 dr, 98:2 er



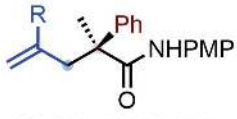
2av, 98 %, 11:1 dr, 96:4 er



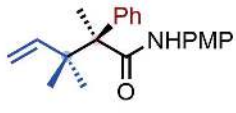
2aw, 99 %, 8:1 dr, 97.5:2.5 er



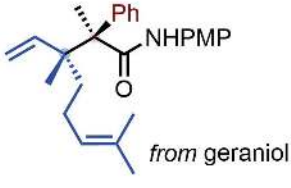
2ax, 96 %, 14:1 dr, 97:3 er



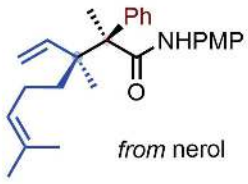
2ay (R=H), 79 %, 91:9 er
2az (R=Me), 82 %, 92.5:7.5 er



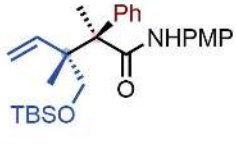
2ba, 80 %, 89:11 er



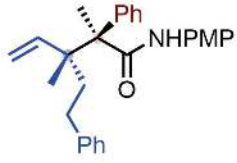
2bb, 54 %, 2.5:1 dr, 95:5 er



2bc, 58 %, 2.5:1 dr, 95:5 er



2bd, 48 %, 4:1 dr, 89:11 er



2be, 56 %, 2:1 dr, 96:4 er

Fig. 3. Scope of DAP-catalyzed enantioselective reductive Eschenmoser-Claisen rearrangement.

temperatures, but also eliminates the dependence on a chelating group.

Reaction development

We started our investigations using acryloyl imidate **1aa** as a model substrate (Fig. 2). Achiral DAP-catalyst **P1** gave corresponding rearranged amide **2aa** in virtually quantitative yield and with 83:17 diastereomeric ratio using phenylsilane as the terminal reductant (Fig. 2, entry 1). Chiral catalyst **P2** derived from a readily accessible diimine delivers a solid proof of concept providing **2aa** with a 25:75 enantiomeric ratio (entry 2). Further screening of C_2-symmetric *bis*-dihydroisoquinoline–derived DAP catalysts **P3** to **P6** with different side arms (entries 3 to 6) revealed that **P3** featuring isopropyl groups gives the best overall result for **2aa** (70%, 85:15 dr, 87:13 er). Additional *o-, p*-dimethyl substituents on the aromatic backbone (**P7**) improved stereoselectivities (92:8 dr and 97.5:2.5 er) as well as the yield (92%) of **2aa**. By contrast, slightly bulkier analog **P8** led to inferior results in all three key metrics (entry 8). Omittance of the DAP catalyst or phenylsilane completely stalled the reaction, indicating the absence of any background reactivity (entry 9). Using 0.33 equivalents of $PhSiH_3$ delivers similar catalytic performance, indicating that all three hydrides of the silane are engaged in the reaction (entry 10). The evaluation of other terminal reductants showed that HBpin is also competent, yielding **2aa** with slightly lower stereoselectivities (entry 11), whereas diphenylsilane was not a suitable reductant (entry 12). The effect of the reaction temperature on the outcome of the transformation revealed some unusual behavior. The onset of the Eschenmoser-Claisen rearrangement occurs around 60°C and, as expected, the reaction rate increases from 60°C to 120°C (entries 13 to 15 and 7). However, an increase of temperature also significantly improves both the diastereo- and enantiocontrol with a large stable plateau between 80° and 120°C slightly dropping beyond 130°C.

Substrate scope

With the aforementioned conditions, we investigated the scope of the reductive enantioselective Eschenmoser-Claisen rearrangement (Fig. 3). Regarding the R^1 substituent on the double bond of acryloyl imidates, a range of aromatic groups bearing either electron-donating or electron-withdrawing functionalities all provided *syn*-amides **2aa** to **2aj** in good to excellent yields and enantio- and diastereoselectivities. Aryl groups with increased electron density slightly decreased the reactivity (4OMe, **2ab** and 3,5-dimethyl, **2aj**). The halide functionalities, fluoro- (**2af**), bromo- (**2ag**), and chloro- (**2ah**), are fully compatible under DAP catalysis. The absolute configuration of obtained amides was established through single crystal x-ray diffraction analysis of product **2ah**. A 2-naphthyl substituted substrate underwent the reductive rearrangement smoothly, giving amide **2ai** in 93% yield with 97.5:2.5 er and >20:1 dr. In addition to aryl groups, R^1 can be aliphatic, e.g., phenylethyl delivering amide **2ak** with 95:5 er and 3:1 dr. Imidate substrate **1al** having an internal alkene (R^5 = Me) allows the formation of product (**2al**) with a quaternary stereogenic center beyond the methyl substitution while maintaining the excellent selectivity characteristics of the rearrangement.

We then broadly investigated the substitution pattern of various allyl units (**2am** to **2aq**). Major structural variation of R^3 is possible without compromising the reaction selectivity and efficiency, providing access to the corresponding amides bearing functionalities such as alkyl (Pr, **2am,** Et, **2au**), alkenyl (**2an**), alkynyl (**2ao**), benzyl ether (**2ap**), and a CF_3 group (**2aq**) at the tertiary stereogenic center. The substituent R^2 was found to influence the diastereoselectivity of the rearrangement. The lack of this substituent (R^2 = H) results in diminished diastereoselectivities (**2aq** and **2ar**) but little influence on the enantioselectivity. In addition to Me, various functionalities on R^2, including phenyl (**2as**), trimethylsilyl (**2at**), and CH_2OAc (**2au**), conserved high overall selectivities. The TMS group of **2at** is cleavable and the resulting allylic acetate of **2au** renders this product suitable for Pd-catalyzed allylic alkylations. Linking R^2 and R^3 in cyclic substructures delivers desired amides with 5- to 7-membered rings with exocyclic double bonds (**2av** to **2ax**) in excellent yields and stereoselectivities. Substrates lacking substituents R^3 and R^4 reacted smoothly to amides (**2ay** and **2az**) with a single quaternary stereocenter. Substrates containing substituents on R^3 as well as on R^4 still underwent the [3,3]-sigmatropic rearrangement and generated products (**2ba** to **2be**) containing two vicinal quaternary stereogenic centers with excellent enantioselectivity. The assembly and selectivity control of such vicinal quaternary stereogenic centers, especially in acyclic systems, are extremely difficult to achieve by other methods.

Synthetic application

Because substrates derived from the different double-bond isomers of geraniol and nerol provided access to the opposite diastereomeric products **2bb** and **2bc**, we reasoned that this transformation is capable of providing a stereodivergent pathway to selectively access all four possible stereoisomers of the rearranged products (*38*). The absolute and relative configurations are controlled by a well-defined transition-state. The absolute configurations of **2aa** are set by the DAP catalyst **P7** and *ent*-**P7**, and the relative stereochemistry is embedded by adjusting the olefin geometry of the allyl unit (Fig. 4A). In this respect, (*Z*)-**1aa** mostly maintains the selectivity performance with slightly higher diastereoselectivities and minimally reduced enantioselectivity. Such a direct catalytic stereodivergent pathway providing access to enantioenriched building blocks bearing vicinal quaternary stereocenters is of great relevance for natural product-oriented syntheses. To illustrate this utility, we developed a formal synthesis of the alkaloid (+)-aphanorphine (Fig. 4B). Toward this goal, allyl 2-(3-methoxyphenyl) acryloyl imidate **3** smoothly underwent rearrangement under our standard Eschenmoser-Claisen conditions, providing amide **4** in 81% yield and 90:10 er, upgraded by recrystallization to 98:2 er. The amide was hydrolyzed to the corresponding carboxylic acid **5** (*39*) and subsequently cyclized under an oxidative Pd-catalyzed intramolecular oxyarylation to tricyclic lactone **6**. The aminolysis of **6** afforded amide **7** in 74% yield. In turn, reduction of **7** by $LiAlH_4$ gave amine **8**, which is a known intermediate for the synthesis of (+)-aphanorphine **9** (*40*). Considering the rare capability of synthesizing all stereo analogs of vicinal quaternary carbon stereocenters, the enantioselective rearrangement could provide a general stereodivergent strategy for the synthesis of clerodane diterpenes (Fig. 4C). Clerodane diterpenes form a very large and diverse class secondary metabolites exhibiting broad biological activities (*41*). Structurally, clerodane diterpenes are classified in four types, TT, CT, CC, and TC, on the basis of their relative configuration of decalin junction (*cis*/*trans*) and relative stereochemistry of the methyl-substituted stereogenic centers of C-8 and C-9 (*cis*/*trans*). To showcase the potential, we targeted Δ^4-3-octalone as a late key intermediate for TC-type strigillanoic acid B. We started the synthesis from an acryloyl imidate **10**, which selectively provided corresponding amide **11** in 89% yield and with 7:1 dr and 96.5:3.5 er. The amide was reduced through **11** to a primary alcohol, which was subsequently converted to its methyl ether. Palladium-catalyzed methoxylation of the aryl chloride gave product **12**. The silyl group was cleaved by a hydroboration-oxidation-Peterson olefination sequence giving **13** in 85% yield. A selective iron-catalyzed seleno-cyclization of **13** with *N*-phenylselenophthalimide provided compound **14**. Reductive cleavage of the phenylseleno group generated tetraline **15**. In turn, Birch reduction followed by acid-assisted isomerization selectively delivered Δ^4-3-octalone **16** in 82% yield. Enone **16** is a highly versatile intermediate for the synthesis of both *cis*- and *trans*-decalins such as *ent*-strigillanoic acid B (type TC, **17**) and 15,16-epoxy-*cis*-cleroda-3,13 (16),14-triene (type CC, **18**) (*42*). Given the capability of the rearrangement to control the

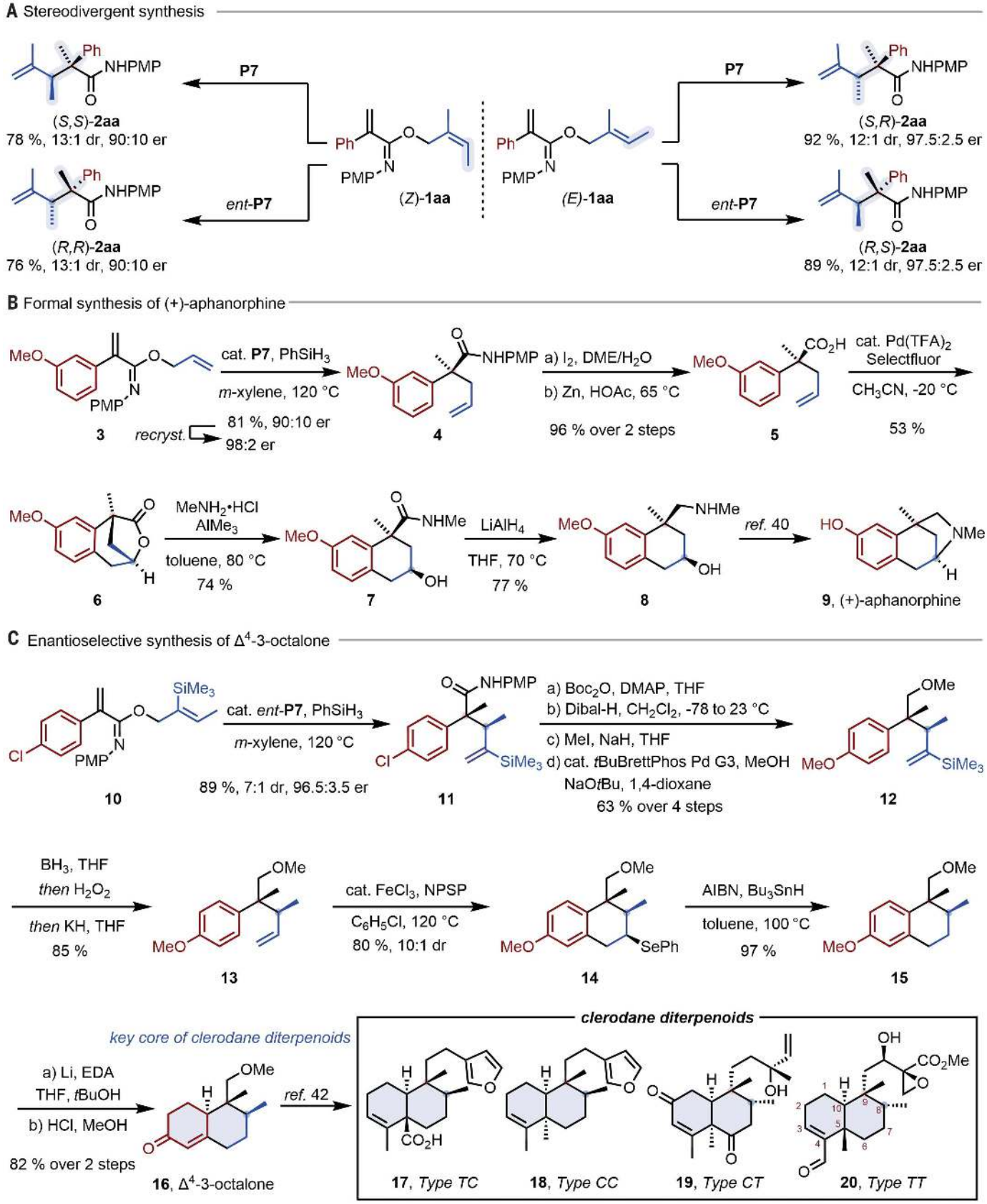


Fig. 4. Synthetic applications. (**A**) Selective stereodivergent syntheses of the four possible stereoisomers of the vicinal stereogenic centers. (**B**) Formal total synthesis of (+)-aphanorphine. (**C**) Enantioselective synthesis of Δ^4-3-octalone as an exemplary key core for the synthesis of clerodane diterpenes. NPSP, *N*-phenylselenophthalimide; EDA, ethylenediamine.

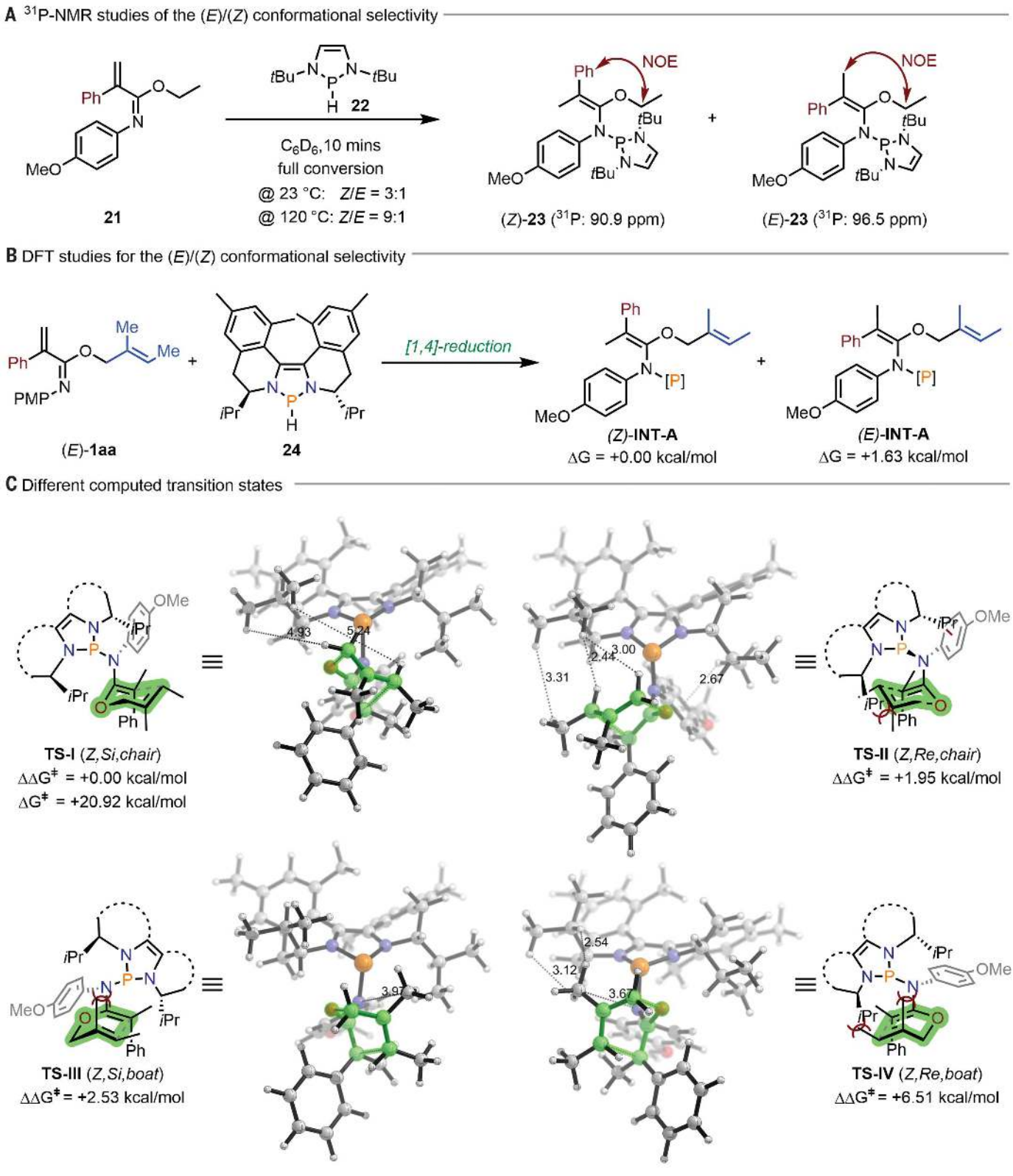


Fig. 5. Mechanistic insights of the stereoinduction step of the [3,3]-sigmatropic rearrangement. (**A**) ^{31}P-NMR studies of the (*E*)/(*Z*) conformational selectivity of conjugate reduction. (**B**) DFT studies for the (*E*)/(*Z*) conformational selectivity. (**C**) Different computed transition states of the [3,3]-sigmatropic rearrangement.

selective formation of each diastereomer on C-8 and C-9, our process holds great potential to access two other types of clerodane diterpenes, for example, 6β-acetoxy-2-oxokolavenool (type CT, **19**) and clerocidin (type TT, **20**).

Mechanistic studies

To gain insight into the stereoinduction mechanism of the rearrangement step of the transformation, we conducted mechanistic studies (Fig. 5). A key factor in the stereocontrol of the [3,3]-sigmatropic rearrangement is the double bond geometry of the *N*,*O*-ketene acetal intermediate **I** generated through the conjugate reduction of the acryloyl imidate substrate by the chiral DAP-H species. To study the (*Z*):(*E*) ratio of the reduction step decoupled from the rearrangement process by ^{31}P-nuclear magnetic resonance (^{31}P-NMR) imaging, we chose ethyl acryloyl imidate **21** as a reduceable but not rearrangeable surrogate. Treating **21** with stoichiometric amounts of DAP-H **22** rapidly produced the two conformational isomers (*Z*)-**23** and (*E*)-**23** with a ratio of 3:1 at 23°C. The (*Z*):(*E*) ratio increased to 9:1 by either heating the mixture to 120°C or conducting the reduction at a higher temperature, indicating a reversible dynamic behavior. On the basis of the characteristic NOE difference for the two isomers, the major isomer could be identified as a (*Z*) conformation (see the supplementary materials for more details). This finding is consistent with results by density functional theory (DFT) computations (Fig. 5B; also see the full computational details in supplementary materials). The computations attribute to the (*Z*) configuration of intermediate **INT-A** generated from **1aa** and **P7**-derived DAP-H **24** a 1.63 kcal/mol lower energy than its conformational isomer (*E*)-**INT-A**, corresponding well to the 9:1 ratio experimentally observed. This equilibration to a higher (*Z*) ratio at elevated temperature explains the very unusual increase in diastereoselectivity that we observed during the optimization studies. The [3,3]-sigmatropic rearrangement itself can potentially involve several conformations linked to different stereochemical outcomes. According to the DFT computations, through the (*Z*)-**INT-A** and a chair-like transition state, the lowest-energy transition state is **TS-I** (*Z, Si, chair*), where a *Si*-face attack of the allyl units to *N*,*O*-ketene acetal produces the major stereoisomer (*S*,*R*)-**2aa**. The activation energy from (*Z*)-**INT-A** to reach **TS-I** was computed to be 20.92 kcal/mol, fully consistent with an onset of the rearrangement at 60° to 80°C. By contrast, **TS-II** (*Z, Re,* chair), which leads to enantiomeric product (*R*,*S*)-**2aa**, undergoes *Re*-face rearrangement and has a transition state barrier that is 1.95 kcal/mol higher in energy compared with **TS-I** due to the steric repulsion of the allyl group with the isopropyl side arm of catalyst **P7**. In the optimized structure of **TS-II**, an unfavorable steric interaction between an isopropyl side arm of the DAP catalyst and the para-methoxyphenyl (PMP) group of the substrate destabilizes this minor transition state. By comparison, the two possible boat-like transition states were found to have higher energy barriers: 2.53 kcal/mol for **TS-III** and 6.51 kcal/mol for **TS-IV**. The steric clash between the methyl group located at R^2 and *N*-substituent significantly destabilizes the boat-like transition states, consistent with the experimentally observed influence of R^2 for the diastereoselectivity of the process. Overall, the predicted diastereo- and enantioselectivities for **2aa** are 87:13 dr and 93.5:6.5 er (see fig. S10 for topological details of all transition states), which match well with the experimental values of 92:8 dr and 97.5:2.5 er.

Conclusions

We have discovered that chiral DAP hydrides catalyze highly enantioselective reductive Eschenmoser-Claisen rearrangements by covalent bonding to induce well-defined favorable transition states even at high reaction temperatures. As a result, sterically demanding arrays of vicinal quaternary stereogenic centers can be synthesized without the requirement for additional chelating activating groups. This prototype process illustrates the large application potential of chiral DAP catalysts for further classes of enantioselective transformations.

REFERENCES AND NOTES

1. E. N. Jacobsen, A. Pfaltz, H. Yamamoto, *Comprehensive Asymmetric Catalysis* (Springer, 1999).
2. K. W. Quasdorf, L. E. Overman, *Nature* **516**, 181–191 (2014).
3. D. Pierrot, I. Marek, *Angew. Chem. Int. Ed.* **59**, 36–49 (2020).
4. M. Büschleb *et al*., *Angew. Chem. Int. Ed.* **55**, 4156–4186 (2016).
5. R. Long, J. Huang, J. Gong, Z. Yang, *Nat. Prod. Rep.* **32**, 1584–1601 (2015).
6. F. Zhou *et al*., *Chem. Sci.* **11**, 9341–9365 (2020).
7. M. Hiersemann, U. Nubbemeyer, *The Claisen Rearrangement: Methods and Applications* (Wiley, 2007).
8. A. M. Martín Castro, *Chem. Rev.* **104**, 2939–3002 (2004).
9. H. Ito, T. Taguchi, *Chem. Soc. Rev.* **28**, 43–50 (1999).
10. Y. Liu, X. Liu, X. Feng, *Chem. Sci.* **13**, 12290–12308 (2022).
11. J. Miró *et al*., *J. Am. Chem. Soc.* **142**, 6390–6399 (2020).
12. T. P. Yoon, D. W. C. MacMillan, *J. Am. Chem. Soc.* **123**, 2911–2912 (2001).
13. E. J. Corey, D. H. Lee, *J. Am. Chem. Soc.* **113**, 4026–4028 (1991).
14. L. Abraham, R. Czerwonka, M. Hiersemann, *Angew. Chem. Int. Ed.* **40**, 4700–4703 (2001).
15. J. Tan, C.-H. Cheon, H. Yamamoto, *Angew. Chem. Int. Ed.* **51**, 8264–8267 (2012).
16. E. C. Linton, M. C. Kozlowski, *J. Am. Chem. Soc.* **130**, 16162–16163 (2008).
17. Y. Liu *et al*., *Angew. Chem. Int. Ed.* **53**, 11579–11582 (2014).
18. H. Zheng *et al*., *Nat. Commun.* **9**, 1968 (2018).
19. C. Uyeda, E. N. Jacobsen, *J. Am. Chem. Soc.* **130**, 9228–9229 (2008).
20. C. Uyeda, A. R. Rötheli, E. N. Jacobsen, *Angew. Chem. Int. Ed.* **49**, 9753–9756 (2010).
21. M. E. Geherty, R. D. Dura, S. G. Nelson, *J. Am. Chem. Soc.* **132**, 11875–11877 (2010).
22. J. Kaeobamrung, J. Mahatthananchai, P. Zheng, J. W. Bode, *J. Am. Chem. Soc.* **132**, 8810–8812 (2010).
23. H. Wu, W. Zi, G. Li, H. Lu, F. D. Toste, *Angew. Chem. Int. Ed.* **54**, 8529–8532 (2015).
24. H. Kim, J. Jang, S. Shin, *J. Am. Chem. Soc.* **142**, 20788–20795 (2020).
25. J. H. Reed, P. A. Donets, S. Miaskiewicz, N. Cramer, *Angew. Chem. Int. Ed.* **58**, 8893–8897 (2019).
26. G. Zhang, N. Cramer, *Angew. Chem. Int. Ed.* **62**, e202301076 (2023).
27. M. R. Adams, C.-H. Tien, R. McDonald, A. W. H. Speed, *Angew. Chem. Int. Ed.* **56**, 16660–16663 (2017).
28. T. Lundrigan *et al*., *J. Am. Chem. Soc.* **141**, 14083–14088 (2019).
29. S. Miaskiewicz, J. H. Reed, P. A. Donets, C. C. Oliveira, N. Cramer, *Angew. Chem. Int. Ed.* **57**, 4039–4042 (2018).
30. S. Burck, D. Gudat, M. Nieger, W.-W. Du Mont, *J. Am. Chem. Soc.* **128**, 3946–3955 (2006).
31. D. Gudat, A. Haghverdi, M. Nieger, *Angew. Chem. Int. Ed.* **39**, 3084–3086 (2000).
32. D. M. C. Ould, R. L. Melen, *Chemistry* **26**, 9835–9845 (2020).
33. J. H. Reed, J. Klett, C. Steven, N. Cramer, *Org. Chem. Front.* **7**, 3521–3529 (2020).
34. A. W. H. Speed, *Chem. Soc. Rev.* **49**, 8335–8353 (2020).
35. J. Zhang, J.-D. Yang, J.-P. Cheng, *Natl. Sci. Rev.* **8**, nwaa253 (2020).
36. S. E. Denmark, M. A. Harmata, *J. Am. Chem. Soc.* **104**, 4972–4974 (1982).
37. C. C. Chong, H. Hirao, R. Kinjo, *Angew. Chem. Int. Ed.* **54**, 190–194 (2015).
38. S. Krautwald, E. M. Carreira, *J. Am. Chem. Soc.* **139**, 5627–5639 (2017).
39. P. Metz, *Tetrahedron* **49**, 6367–6374 (1993).
40. S. Takano, K. Inomata, T. Sato, K. Ogasawara, *J. Chem. Soc. Chem. Commun.* **21**, 1591–1592 (1989).
41. R. Li, S. L. Morris-Natschke, K.-H. Lee, *Nat. Prod. Rep.* **33**, 1166–1226 (2016).
42. T. Tokoroyama *et al*., *Tetrahedron* **44**, 6607–6622 (1988).

ACKNOWLEDGMENTS

We thank R. Scopelliti and F. Fadaei Tirani for x-ray crystallographic analysis of compounds **2ah**, **1as**, and **45** and R. Laplaza for computational assistance. The Laboratory for Computational Molecular Design at Ecole Polytechnique Fédérale de Lausanne (EPFL) is acknowledged for providing computational resources. **Funding:** This work was supported by EPFL and the Swiss National Science Foundation (grant 200020_212165). **Author contributions:** N.C. conceived and supervised the project. G.Z. performed the experiments. M.D.W. performed the DFT computations. All authors designed the experiments, analyzed the data, and drafted the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The x-ray crystallographic coordinates for the structures reported in this study have been deposited at the Cambridge Crystallographic Data Centre (CCDC) under deposition numbers 2242580, 2284115, and 2285000 and can be obtained free of charge from https://www.ccdc.cam.ac.uk/structures/. All other molecular characterization data are included in the supplementary materials. Requests for materials or additional information should be addressed to the corresponding author.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adl3369
Materials and Methods
Figs. S1 to S12
Tables S1 and S2
NMR Spectra
References (*43–82*)

Submitted 12 October 2023; accepted 19 December 2023
10.1126/science.adl3369

GAMMA-RAY ASTRONOMY

Acceleration and transport of relativistic electrons in the jets of the microquasar SS 433

H.E.S.S. Collaboration*†

SS 433 is a microquasar, a stellar binary system that launches collimated relativistic jets. We observed SS 433 in gamma rays using the High Energy Stereoscopic System (H.E.S.S.) and found an energy-dependent shift in the apparent position of the gamma-ray emission from the parsec-scale jets. These observations trace the energetic electron population and indicate that inverse Compton scattering is the emission mechanism of the gamma rays. Our modeling of the energy-dependent gamma-ray morphology constrains the location of particle acceleration and requires an abrupt deceleration of the jet flow. We infer the presence of shocks on either side of the binary system, at distances of 25 to 30 parsecs, and that self-collimation of the precessing jets forms the shocks, which then efficiently accelerate electrons.

SS 433 (also cataloged as V1343 Aql) is a binary system comprising a compact object, likely a black hole (*1*–*3*), and a type A supergiant star (*4*). Material from the supergiant is accreted onto the black hole, causing the latter to launch a pair of jets moving in opposite directions at approximately a quarter of the speed of light (*c*) (*5*–*7*). The jets are orientated almost perpendicular to our line of sight from Earth (*8*) and precess with a half-opening angle of 20° and a period of 162 days (*9*–*12*). Adopting a distance measurement of 5.5 kpc (*7*), optical and radio observations have shown that the precessing jets extend to distances of ~10^{-3} pc (*13*) and ~0.1 pc (*7*) from the black hole, respectively. X-ray emission reappears 25 pc from the binary (Fig. 1), indicating collimated flows (the outer jets) on larger scales, which emit x-ray photons through nonthermal processes (*14*–*17*).

The outer jets terminate ~100 pc from the black hole (*14*), where they deform the surrounding radio nebula (known as W 50 or SNR G039.7-02.0), which is thought to be the supernova remnant associated with the formation of the compact object in SS 433 (*18*). The morphology of W 50 indicates that the opening angle of the outer jets is considerably smaller than the 20° precession angle of the inner jets (*19*); the origin of this discrepancy is unknown (*20*). The lack of apparent change in the measured positions of radio filaments in the jet termination regions over a 33-year period provides an upper limit on their velocity of <0.023*c* (*21*, *22*), although it is unclear whether the radio filaments trace the jets' flow. Bright x-ray knots emitting synchrotron radiation have been observed in the outer jets, but the temporal baseline and angular resolution were insufficient to determine their velocity (*23*). The dynamics of the outer jets and their termination process are poorly understood.

Several attempts have been made to use x-ray observations to probe the nonthermal emission mechanisms and internal dynamics of the eastern (*16*, *24*) and western (*17*) outer jets. However, observations of the x-ray synchrotron emission alone cannot resolve variations in the distribution of accelerated particles. The intensity of synchrotron emission is approximately proportional to the number density of accelerated electrons and the energy density of the magnetic field; the latter is poorly constrained. X-ray–emitting electrons can also up-scatter low-energy photons to the gamma-ray regime through the inverse Compton scattering process. This process directly traces the population of high-energy electrons because the diffuse low-energy photon distribution in the Galaxy is expected to be smooth on the spatial scale of the outer jets (*25*, *26*). Previous observations of tera–electron volt gamma rays emitted by the outer jets of SS 433 (*27*) indicate that the same energetic electrons responsible for the x-ray emission also produce gamma rays through inverse Compton scattering (*28*, *29*). However, the angular resolution was insufficient to determine the emission regions and therefore the source of the energetic particles.

H.E.S.S. observations of SS 433

We imaged the outer jets of SS 433 at tera–electron volt energies using the High Energy Stereoscopic System (H.E.S.S.) array of imaging atmospheric Cherenkov telescopes. The observations totaled more than 200 hours of exposure time and were analyzed by using previously described methods, which were optimized for faint sources and the highest energies (*30*). The extended source HESS J1908+063 (also known as MGRO J1908+06) contaminates part of the SS 433 jet so was modeled then subtracted from the data (figs. S2 and S3). The resulting gamma-ray image (Fig. 1A) sh[ows] two regions of gamma-ray emission at [the] previously known x-ray positions of the eastern and western jets, with peak statistical significances of 7.8σ and 6.8σ, respectively. No significant (>5σ) emission was detected from the central binary or the eastern termination region (Fig. 1A), which is as we expected because the x-ray emission from those regions predominantly has a thermal emission mechanism (*14*, *31*). Fermi J1913+0515 (which has coordinates right ascension 288.28°±0.04, declination 5.27°±0.04) is a giga–electron volt gamma-ray source known to pulsate with a period consistent with the jet precession (*32*), indicating a potential connection with the SS 433 system. We detected no significant tera–electron volt emission from this source (*26*). The measured spectral energy distributions of each of the jets are shown in Fig. 1, B and C.

To investigate the energy dependence of the gamma-ray emission, we split the full H.E.S.S. energy range into three bands (0.8 to 2.5 TeV, 2.5 to 10 TeV, and >10 TeV), which were selected to have approximately the same gamma-ray excess counts over the background in each band. The significance maps for each band are shown in Fig. 2. We detected significant (>5σ) gamma-ray emission along both jets for the two highest-energy bands. In the lowest-energy band, we found lower-significance evidence of emission at 4.4σ and 4.7σ for the eastern and western jets, respectively. Gamma-ray emission >10 TeV appears only at the base of the outer jets (visible in x-rays) for both the eastern and western jets. By contrast, lower-energy gamma rays have their peak surface brightnesses at locations further along each jet, except for the lowest-energy band on the eastern side. In the latter case, no significant emission was detected inside the x-ray jet region, and evidence for emission appears close to the outer jet base (Fig. 2A). In the western jet, the best-fitting positions of the gamma-ray emission in each energy band have distances from the central binary (table S4) that differ from each other by 0.97σ and 2.6σ when comparing adjacent energy bands and by 5.3σ when comparing the lowest- and highest-energy band. The equivalent values for the eastern jet are 2.6σ, 3.3σ, and 0.1σ. Our significance calculations include both systematic and statistical sources of uncertainty and a trials factor correction (*26*).

Location of the particle acceleration

We interpret the offsets between the emission in different energy bands as indicating that transport of particles in the outer jets is dominated by advection (the bulk jet flow), not diffusion (random scattering of the particles by magnetic field fluctuations). The energy-dependent morphology then reflects an energy-dependent particle energy loss timescale. We infer that the emission arises from relativistic

*Corresponding author: Laura Olivera-Nieto (laura.olivera-nieto@mpi-hd.mpg.de); Brian Reville (brian.reville@mpi-hd.mpg.de); Jim Hinton (jim.hinton@mpi-hd.mpg.de); Michelle Tsirou (michelle.tsirou@desy.de)
†H.E.S.S. Collaboration authors and affiliations are listed in the supplementary materials.

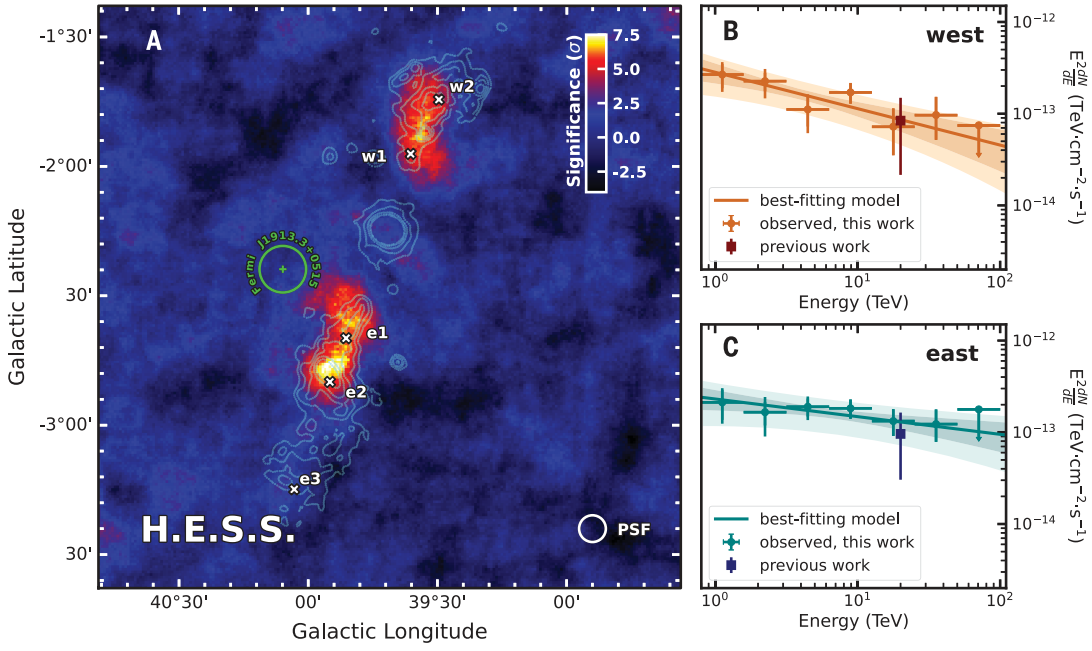


Fig. 1. Gamma-ray observations of SS 433. (**A**) Significance map of the H.E.S.S. observations at energies >0.8 TeV (color bar). Cyan contours show the previously published x-ray emission (*14*, *15*). Labeled white crosses indicate locations of x-ray regions discussed in the text. Significance is for the H.E.S.S. excess counts above the background before accounting for statistical trials and after subtraction of the extended source HESS J1908+063 (subtraction shown in fig. S2). The map has been smoothed with a top-hat function of radius 0.1°. The white circle indicates the 68% containment region of the H.E.S.S. point-spread function (PSF). The green cross indicates the position of the possibly related source Fermi J1913+0515, and the green circle is its uncertainty. (**B**) Orange circles indicate our measured spectral energy distribution of the gamma-ray emission from the western jet. The red square is a measurement from previous observations (*27*). Error bars indicate the combined statistical (1σ) and systematic uncertainties; downward arrows indicate upper limits at 95% confidence. The solid line is the best-fitting power-law function, with dark and light shaded regions indicating the statistical and systematical uncertainties, respectively. (**C**) Same as (B) but for the eastern jet. The regions from which the spectra shown in (B) and (C) were extracted are shown in fig. S2B.

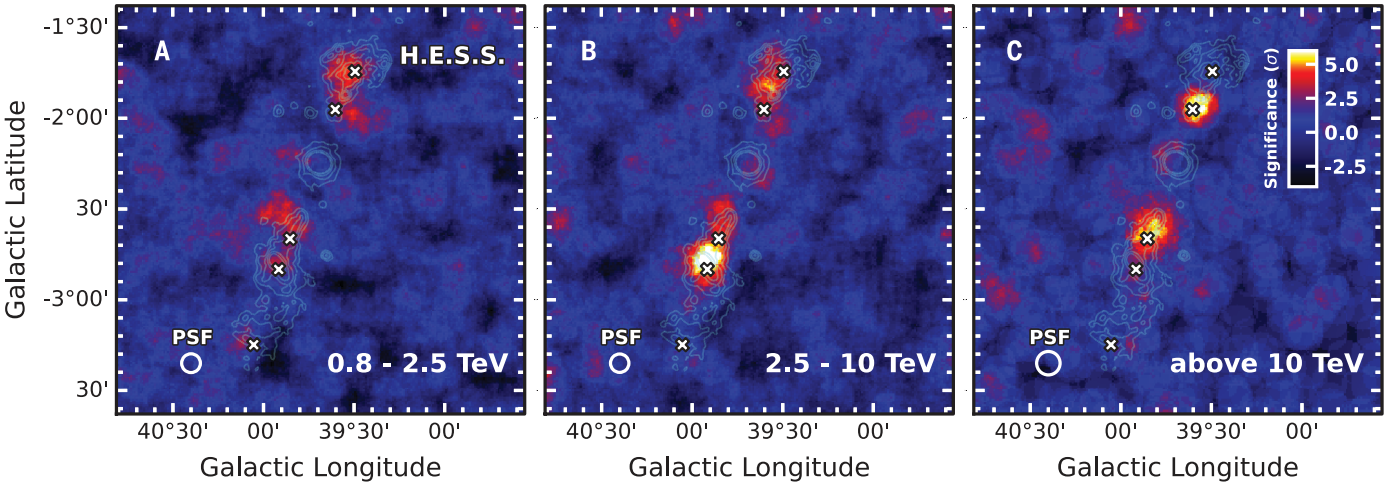


Fig. 2. Gamma-ray observations in different energy bands. Same as Fig. 1A, but split into three gamma-ray energy bands of (**A**) 0.8 to 2.5 TeV, (**B**) 2.5 to 10 TeV, and (**C**) >10 TeV.

electrons, not hadrons, because the loss timescale for hadronic processes depends only very weakly on particle energy (*33*). The dominant energy-loss mechanism for high-energy electrons is likely synchrotron cooling. We conclude that the observed gamma-ray emission is the result of inverse Compton scattering (*33*, *34*) of photons by high-energy electrons. Iron and other heavy nuclei are known to be present in the jet (*35*), so they might also be accelerated in the same region, but our observations cannot be used to constrain their presence (supplementary text).

The shorter cooling time of higher-energy electrons limits the distance from the acceleration site within which they can radiate because they are transported away through either diffusion or advection. The absence of emission above 10 TeV at the location of the x-ray knots e2 and w2 (Fig. 1A) indicates that they cannot be sites of particle acceleration to tera–electron volt energies, contradicting previous interpretations (*27*, *36*). Instead, the concentration of emission above 10 TeV at the base of the x-ray emission from the outer jets indicates that this region is the site of particle acceleration to very high energies. We interpret the energy-dependent position of the gamma-ray emission in the jets of SS 433 as a consequence of the cooling and transport of particles that are accelerated at the base of the outer jets. A schematic diagram of our proposed interpretation is shown in Fig. 3.

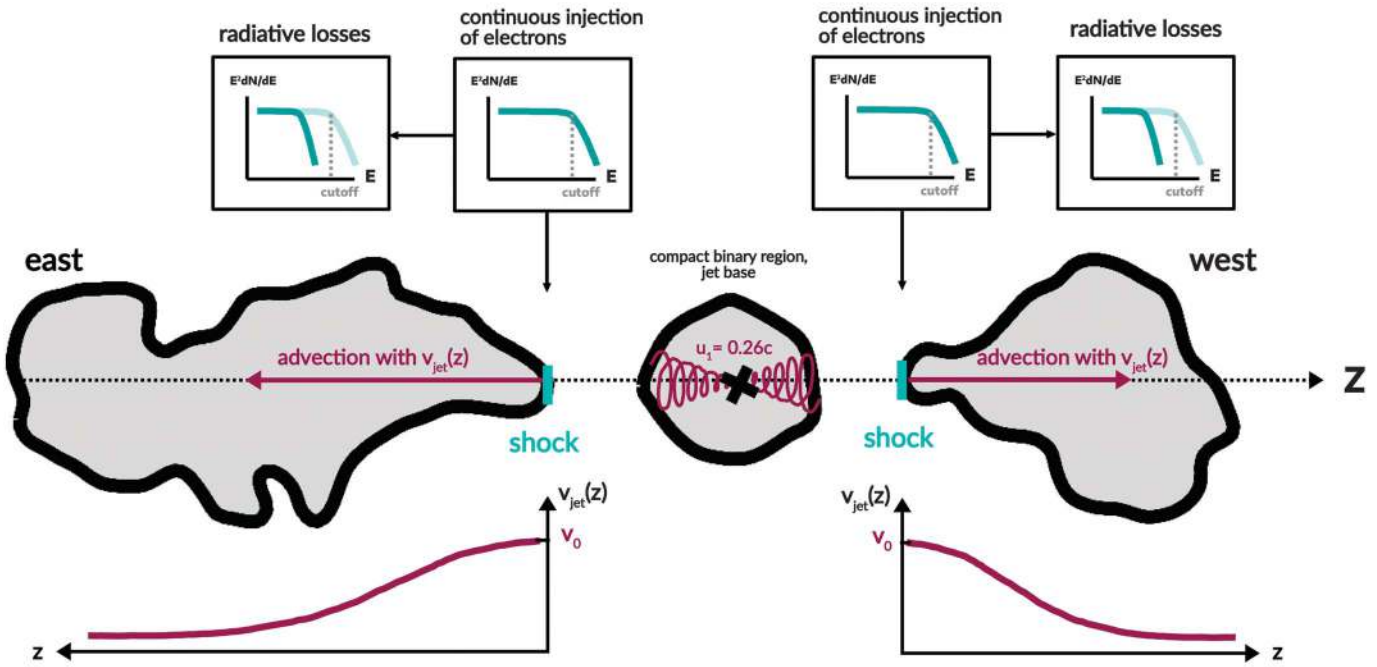


Fig. 3. Schematic diagram of our model. Thick black lines roughly outline the x-ray emission (gray shading) from the central region and the outer jets on the plane of the sky (rotated from the orientation in Fig. 1). The precessing jet is launched with velocity $u_1 \approx 0.26c$ (purple spirals) and travels outward until it encounters a shock discontinuity (cyan bars) at the base of the outer jets. Our 1D model injects electrons continuously at the outer jet base, with an energy spectrum derived from fitting the multiwavelength observations of the outer jets (table S5). The injected electrons lose energy because of radiative losses, which affects their spectrum (indicated in the insets across top). Particles are transported along the jet axis coordinate z through advection (purple arrows) and diffusion (not depicted), with the jet flow at velocity $v_{\rm jet}(z)$. The velocity at the base of the jets behind the shock v_0 is determined by fitting this model to the H.E.S.S. data. We assume the jet flow decelerates as it moves away from the jet base, indicated with the purple curves below the diagram. We also considered the alternative case of a constant-velocity jet, which leads to similar conclusions (supplementary text).

Modeling the outer jet dynamics

Previous studies have shown that the jets are launched from the black hole with initial velocities of $u_1 \approx 0.26c$ (*5*–*7*). We combined the distances between the gamma-ray excess regions in different energy bands with the electron cooling timescales (*26*) to determine the velocity v_0 of the outer jets at their base, ~25 pc away from the central binary. This calculation requires us to assume a spatial dependence of the deceleration of the jets as a function of the distance from the central binary. We used the observed opening angle of the jets in x-ray images to determine the deceleration profile by assuming that the jet flow is incompressible (fig. S14). We also considered a jet propagating with constant velocity, under different energy loss assumptions, which leads to consistent values of v_0 (supplementary text). Our observations cannot distinguish between the different jet propagation scenarios considered.

We modeled the energy-dependent morphology of the gamma-ray emission using a one-dimensional (1D) Monte Carlo simulation that includes radiation and cooling of particles as they are transported along the jet (*26*). The model injects electrons at the base of the outer jet with an energy spectrum assumed to be of the form $dN/dE \propto E^{-\Gamma_e} \cdot \exp\left(-\frac{E}{E_{\rm cut}}\right)$ (Fig. 3), where N is the number of electrons; E is their energy; and Γ_e and $E_{\rm cut}$ are the spectral index and cutoff energy, respectively. We determined the best-fitting parameters of the injected electron spectrum and the average local magnetic field strength from the multiwavelength spectral energy distribution of each outer jet (fig. S10) separately. The value of $E_{\rm cut}$ is not constrained by the data; we found only a lower limit of >200 TeV at 68% confidence level (CL). The model assumes that the injection is continuous for 10,000 years, this timescale being constrained by the combination of existing giga–electron volt gamma-ray flux upper limits (*37*) and the measured tera–electron volt gamma-ray flux (fig. S11). This electron injection timescale is consistent with previous dynamical estimates for the age of the W 50/SS 433 complex, which range between 10,000 and 100,000 years (*19*, *20*). Our simulation evolves the electron population numerically in discrete time steps. In each step, electrons are advected with the local jet velocity then diffuse along the jet axis (neglecting transverse diffusion) and cool radiatively. This leads to an energy- and spatially dependent electron distribution, from which we calculated 1D profiles of the resulting nonthermal emission. The resulting spatial distribution in the gamma-ray range only weakly depends on the parameters of the injected particle distribution (*26*).

Using the H.E.S.S. data, we derived spatial profiles of the gamma-ray flux along the axis joining both outer jets through the central binary in the same three energy bands used in Fig. 2. We fitted the resulting model emission profile to the data with free parameters v_0 and the diffusion coefficient (the latter assumed to be spatially uniform). The injected electron spectrum parameters are fixed to the values obtained from fitting the multiwavelength spectral energy distributions described above. We assumed the same starting velocity for both the eastern and western jet. The best-fitting value is $v_0 = (0.083 \pm 0.026_{\rm stat} \pm 0.010_{\rm syst})c$. The systematic uncertainty is derived from the choice of parameters for the injected electron spectrum (*26*). The gamma-ray spatial profiles and the best-fitting model are shown in Fig. 4.

Interpretation as a standing shock

Our modeling shows that the data are consistent with the presence of a particle accelerator, likely a shock, at the base of the SS 433 outer jets and that it is capable of accelerating particles to very high energies. Our lower limit on

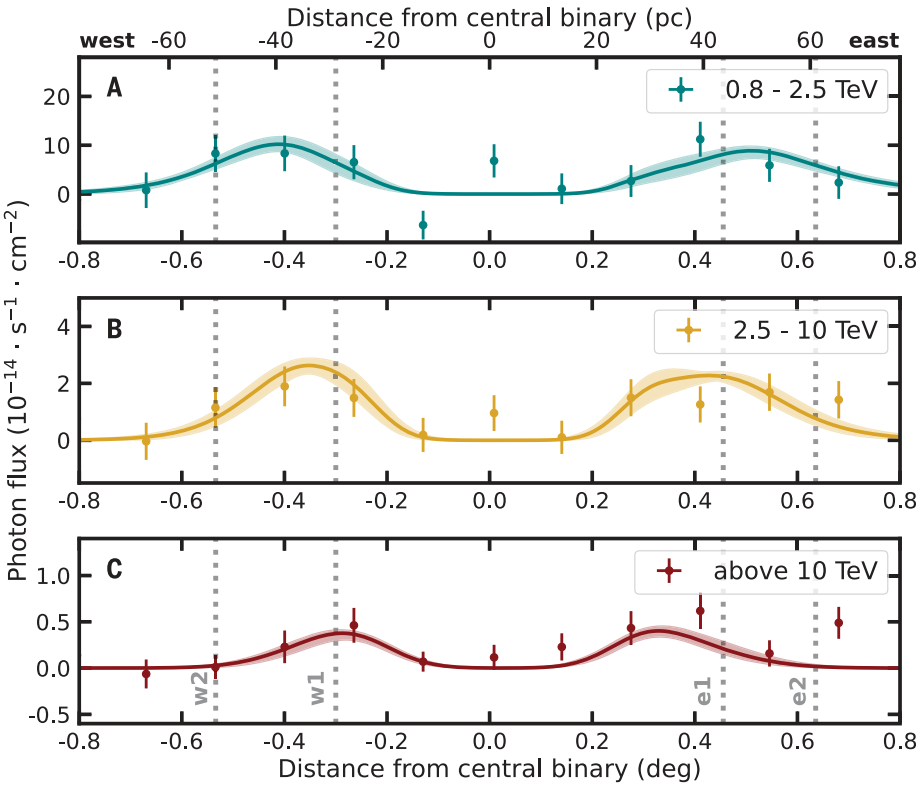


Fig. 4. Gamma-ray flux profiles along the jets compared with the model prediction. Data points are the measured gamma-ray fluxes in spatial bins of 0.14° along the axis joining both jets through the central binary. (**A**, **B**, and **C**) Results for the same three energy bands as in Fig. 2. Error bars indicate the combined statistical (1σ) and systematic uncertainties. Solid lines indicate the prediction of our best-fitting 1D model. The shaded areas indicate the combined statistical uncertainty of the best-fitting parameters. Dashed gray vertical lines, labeled in (C), indicate the positions of the x-ray regions e1, e2, w1, and w2 (Fig. 1A). The top axis indicates physical scale, assuming a distance to the system of 5.5 kpc (*7*).

E_{cut} indicates the acceleration of electrons to energies >200 TeV (68% CL). At the inferred magnetic field strength of ~20 G (table S5), to keep up with cooling the acceleration rate must be close to the theoretical maximum, assuming diffusive shock acceleration (*26*, *38*). Therefore, the jet flow cannot have decelerated much from its inferred launch velocity of 0.26*c* before reaching the shock because if it had, the particle acceleration could not compete with radiative losses at electron energies above several hundred tera–electron volts (fig. S13).

The velocity we infer at the base of the outer jets v_0 is a fraction ($\chi = 0.319 \pm 0.10_{stat} \pm 0.039_{syst}$) of the jet launch velocity. This is compatible with the velocity ratio expected for a subrelativistic shock, which is $\chi \approx 0.25$ (*39*, *40*). A shock at this location is consistent with the spatial coincidence between the position of the highest-energy gamma-ray emission and the location where the x-ray emission reappears (*16*, *17*). This region has previously been interpreted as the acceleration site, but without involving shocks (*16*). We conclude that if the advection in the jet flow is taken into account, the observations are consistent with the shock acceleration scenario. Our observations also constrain the velocity of the shock, which would have needed to advance a small distance (≪10 pc) in the lifetime of the tera–electron volt gamma-ray–emitting electrons (fig. S13).

There is no single model that has yet reproduced all the observational features of SS 433 (*19*, *41*). Although simulations can account for the observed difference in opening angle between the inner and outer jets owing to the action of the ambient medium (*20*, *42*, *43*), this process would take place near the binary and does not result in the observed sharp transitions or shocks at 25 to 30 pc. The mirrored reappearance of the jets at this distance implies that the determination of this radius has a physical cause, although there is no further observational evidence to indicate that this location is unusual. Radio observations of the jet launch region reveal the presence of an equatorial outflow perpendicular to the axis of the jets (*44*), which has previously been proposed to form a quasi-spherical shock at distances of tens of parsecs from the binary (*45*). However, the x-ray shell that would be produced by such a shock has not been detected.

The proximity of SS 433 to Earth permits the investigation of shock physics and associated nonthermal processes in mildly relativistic jets. These insights might also apply to other microquasars (*46*) and to the larger and more distant jets launched from the centers of other galaxies, in which jet substructure cannot be resolved at high energies (*47*). Our results imply that shocks can form within relativistic jets and accelerate particles at close to the theoretical maximum energy (*33*, *48*). Thus, microquasars could contribute to the measured Galactic cosmic-ray flux at peta–electron volt energies, whereas extragalactic jets could reach the exa–electron volt regime of ultrahigh-energy cosmic rays (supplementary text).

REFERENCES AND NOTES

1. E. Seifina, L. Titarchuk, *Astrophys. J.* **722**, 586–604 (2010).
2. A. M. Cherepashchuk, K. A. Postnov, A. A. Belinski, *Mon. Not. R. Astron. Soc.* **485**, 2638–2641 (2019).
3. A. M. Cherepashchuk, A. A. Belinski, A. V. Dodin, K. A. Postnov, *Mon. Not. R. Astron. Soc.* **507**, L19–L23 (2021).
4. T. C. Hillwig, D. R. Gies, *Astrophys. J.* **676**, L37–L40 (2008).
5. B. Margon, H. C. Ford, S. A. Grandi, R. P. S. Stone, *Astrophys. J.* **233**, L63 (1979).
6. S. S. Eikenberry *et al.*, *Astrophys. J.* **561**, 1027–1033 (2001).
7. K. M. Blundell, M. G. Bowler, *Astrophys. J.* **616**, L159–L162 (2004).
8. D. H. Roberts *et al.*, *Astrophys. J.* **719**, 1918–1925 (2010).
9. A. C. Fabian, M. J. Rees, *Mon. Not. R. Astron. Soc.* **187**, 13P–16P (1979).
10. M. Milgrom, *Astron. Astrophys.* **76**, L3 (1979).
11. B. Margon *et al.*, *Astrophys. J.* **230**, L41 (1979).
12. S. Fabrika, *Astrophys. Space Phys. Res.* **12**, 1 (2004).
13. K. Davidson, R. McCray, *Astrophys. J.* **241**, 1082 (1980).
14. W. Brinkmann, B. Aschenbach, N. Kawai, *Astron. Astrophys.* **312**, 306 (1996).
15. S. Safi-Harb, H. Ögelman, *Astrophys. J.* **483**, 868–881 (1997).
16. S. Safi-Harb *et al.*, *Astrophys. J.* **935**, 163 (2022).
17. K. Kayama *et al.*, *Publ. Astron. Soc. Jpn.* **74**, 1143–1156 (2022).
18. R. Elston, S. Baum, *Astron. J.* **94**, 1633 (1987).
19. P. T. Goodall, F. Alouani-Bibi, K. M. Blundell, *Mon. Not. R. Astron. Soc.* **414**, 2838–2859 (2011).
20. M. G. Bowler, R. Keppens, *Astron. Astrophys.* **617**, A29 (2018).
21. P. T. Goodall, K. M. Blundell, S. J. Bell Burnell, *Mon. Not. R. Astron. Soc.* **414**, 2828–2837 (2011).
22. H. Sakemi, R. Omae, T. Ohmura, M. Machida, *Publ. Astron. Soc. Jpn.* **73**, 530–544 (2021).
23. M. G. Watson, R. Willingale, J. E. Grindlay, F. D. Seward, *Astrophys. J.* **273**, 688 (1983).
24. W. Brinkmann, G. W. Pratt, S. Rohr, N. Kawai, V. Burwitz, *Astron. Astrophys.* **463**, 611–619 (2007).
25. C. C. Popescu *et al.*, *Mon. Not. R. Astron. Soc.* **470**, 2539–2558 (2017).
26. Materials and methods are available as supplementary materials.
27. A. U. Abeysekara *et al.*, *Nature* **562**, 82–85 (2018).
28. P. Bordas, V. Bosch-Ramon, J. M. Paredes, *Int. J. Mod. Phys. D* **19**, 749–755 (2010).
29. T. Sudoh, Y. Inoue, D. Khangulyan, *Astrophys. J.* **889**, 146 (2020).
30. L. Olivera-Nieto, H. X. Ren, A. M. W. Mitchell, V. Marandon, J. A. Hinton, *Eur. Phys. J. C* **82**, 1118 (2022).
31. H. L. Marshall, C. R. Canizares, N. S. Schulz, *Astrophys. J.* **564**, 941–952 (2002).
32. J. Li *et al.*, *Nat. Astron.* **4**, 1177–1184 (2020).
33. F. A. Aharonian, *Very High Energy Cosmic Gamma Radiation: A Crucial Window on the Extreme Universe* (World Scientific Book, 2004).
34. G. R. Blumenthal, R. J. Gould, *Rev. Mod. Phys.* **42**, 237–270 (1970).
35. S. Migliari, R. Fender, M. Méndez, *Science* **297**, 1673–1676 (2002).
36. S. S. Kimura, K. Murase, P. Mészáros, *Astrophys. J.* **904**, 188 (2020).
37. K. Fang, E. Charles, R. D. Blandford, *Astrophys. J. Lett.* **889**, L5 (2020).

38. L. O. Drury, *Rep. Prog. Phys.* **46**, 973–1027 (1983).
39. L. D. Landau, E. M. Lifshitz, *Fluid Mechanics* (Pergamon Press, 1987).
40. J. G. Kirk, P. Duffy, *J. Phys. G Nucl. Part. Phys.* **25**, R163–R194 (1999).
41. T. Ohmura *et al*., *Astrophys. J.* **910**, 149 (2021).
42. R. Monceau-Baroux, O. Porth, Z. Meliani, R. Keppens, *Astron. Astrophys.* **561**, A30 (2014).
43. R. Monceau-Baroux, O. Porth, Z. Meliani, R. Keppens, *Astron. Astrophys.* **574**, A143 (2015).
44. K. M. Blundell, A. J. Mioduszewski, T. W. B. Muxlow, P. Podsiadlowski, M. P. Rupen, *Astrophys. J.* **562**, L79–L82 (2001).
45. A. Konigl, *Mon. Not. R. Astron. Soc.* **205**, 471–485 (1983).
46. D. Kantzas *et al*., *Mon. Not. R. Astron. Soc.* **500**, 2112–2126 (2021).
47. H.E.S.S. Collaboration *et al*., *Nature* **582**, 356 (2020).
48. A. M. Hillas, *Annu. Rev. Astron. Astrophys.* **22**, 425–444 (1984).
49. L. Olivera-Nieto, LauraOlivera/particle-transport-1D: v1.0. Zenodo (2023); https://doi.org/10.5281/zenodo.10229720.

ACKNOWLEDGMENTS

We thank J. Li and K. Fang for providing the reduced Fermi-LAT dataset and giga–electron volt flux points, respectively. We appreciate the excellent work of the technical support staff in Berlin, Zeuthen, Heidelberg, Palaiseau, Paris, Saclay, Tübingen, and Namibia in the construction and operation of the equipment. This work benefited from services provided by the H.E.S.S. Virtual Organisation, which is supported by the national resource providers of the EGI Federation. **Funding:** The support of the Namibian authorities and of the University of Namibia in facilitating the construction and operation of H.E.S.S. is gratefully acknowledged, as is the support by the German Ministry for Education and Research (BMBF); the Max Planck Society; the German Research Foundation (DFG); the Helmholtz Association; the Alexander von Humboldt Foundation; the French Ministry of Higher Education; Research and Innovation; the Centre National de la Recherche Scientifique (CNRS/IN2P3 and CNRS/INSU); the Commissariat à l'énergie atomique et aux énergies alternatives (CEA); the UK Science and Technology Facilities Council (STFC); the Irish Research Council (IRC) and the Science Foundation Ireland (SFI); the Knut and Alice Wallenberg Foundation; the Polish Ministry of Education and Science, agreement 2021/WK/06; the South African Department of Science and Technology and National Research Foundation; the University of Namibia; the National Commission on Research, Science & Technology of Namibia (NCRST); the Austrian Federal Ministry of Education, Science and Research and the Austrian Science Fund (FWF); the Australian Research Council (ARC); the Japan Society for the Promotion of Science; the University of Amsterdam; and the Science Committee of Armenia, grant 21AG-1C085. **Author contributions:** L. Olivera-Nieto performed the H.E.S.S. analysis, and M. Tsirou performed the cross-check analysis. N. Tsuji analyzed the x-ray data. B. Reville, L. Olivera-Nieto, and J. Hinton performed the interpretation and modeling. The manuscript was prepared by L. Olivera-Nieto, B. Reville, and M. Tsirou. S. Wagner is the collaboration spokesperson. All other H.E.S.S. collaboration authors contributed to the design, construction, and operation of H.E.S.S. or the development and maintenance of data handling, data reduction, or data analysis software. All authors meet the journal's authorship criteria and have reviewed, discussed, and commented on the results and the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The H.E.S.S. data are available at https://www.mpi-hd.mpg.de/HESS/pages/publications/auxiliary/2023_SS433/index.html, including the counts and background maps, the subsequently derived significance and flux maps (Figs. 1A and 2), the flux profiles (Fig. 3), and the raw data and derived flux points from the spectral measurements of the jets (Fig. 1, B and C). Our modeling code is available at https://github.com/LauraOlivera/particle-transport-1D and archived at Zenodo (*49*).

SUPPLEMENTARY MATERIALS

Submitted 20 April 2023; accepted 4 December 2023
10.1126/science.adi2048

ENVIRONMENTAL POLICY

Machine learning predicts which rivers, streams, and wetlands the Clean Water Act regulates

Simon Greenhill[1,2]*, Hannah Druckenmiller[3,4]†, Sherrie Wang[2,5,6]†, David A. Keiser[7,8,9], Manuela Girotto[10], Jason K. Moore[11], Nobuhiro Yamaguchi[12], Alberto Todeschini[12], Joseph S. Shapiro[1,9,13]*

We assess which waters the Clean Water Act protects and how Supreme Court and White House rules change this regulation. We train a deep learning model using aerial imagery and geophysical data to predict 150,000 jurisdictional determinations from the Army Corps of Engineers, each deciding regulation for one water resource. Under a 2006 Supreme Court ruling, the Clean Water Act protects two-thirds of US streams and more than half of wetlands; under a 2020 White House rule, it protects less than half of streams and a fourth of wetlands, implying deregulation of 690,000 stream miles, 35 million wetland acres, and 30% of waters around drinking-water sources. Our framework can support permitting, policy design, and use of machine learning in regulatory implementation problems.

The 1972 Clean Water Act (CWA), a critically important US environmental policy, represents the cornerstone of federal water quality regulation. Given the importance of healthy waterways for flood protection, clean drinking water, ecosystem health, and economic activity (*1*–*3*), the CWA and reforms to it have enormous potential ecological and economic consequences.

Four recent judicial and executive CWA rules have substantially rewritten CWA coverage—*Rapanos*, the Clean Water Rule (CWR), the Navigable Waters Protection Rule (NWPR), and *Sackett*. A third of the US Supreme Court's environmental cases since 1972 have addressed the CWA, far more than any other environmental policy (*4*). Supreme Court Justice Kennedy's 2006 *Rapanos* opinion found that the CWA protects water resources with a "significant nexus" to navigable waters, meaning a biological, chemical, or physical connection. Justice Scalia's *Rapanos* plurality opinion requiring a surface water connection was the basis for the Trump administration's NWPR, which excluded isolated wetlands and ephemeral streams and required a surface water connection to navigable waters. The Obama administration's CWR clarified jurisdiction under *Rapanos*, though did not seek to change jurisdiction substantially. The Biden administration implemented *Rapanos* with modest modifications (*5*, *6*). The Supreme Court's 2023 *Sackett* decision limits regulation, especially the significant nexus standard.

The CWA protects the "Waters of the United States," but does not define which waters this phrase describes. This makes it difficult to understand precisely which waters gain and lose protection under recent rules. The CWA originally protected navigable waters and their tributaries, under the Constitution's interstate commerce clause. The CWA targets pollution, though CWA Section 404's regulation of the discharge of dredged or fill material into jurisdictional waters affects land-use development.

The US Environmental Protection Agency (EPA) and Army Corps of Engineers (ACE) (*7*) summarize, "**EXISTING TOOLS CANNOT ACCURATELY MAP THE SCOPE OF CLEAN WATER ACT JURISDICTION**" (formatting in original). Accurate mapping has been infeasible because the CWA and rules interpreting it give sufficiently general guidance that ACE must evaluate the geophysical conditions of a water resource to determine whether the CWA regulates it.

Media reports and an amicus brief by the American Water Works Association, National Association of Wetland Managers, and others, for example, assert that NWPR eliminates CWA protection for at least 18% of streams and 51% of wetlands (*8*, *9*). Such statistics identify waters sharing specific characteristics that loosely approximate criteria for regulation in a CWA rule, then assume those waters are regulated (*10*). The EPA and ACE (*7*) call these statistics "highly unreliable" owing to their lack of data on which waters are regulated.

This paper provides the first national, geographically resolved estimate of legally binding CWA regulation. Waters of the United Sta… Machine Learning (WOTUS-ML), a deep lear… model that we build, predicts CWA jurisdictional determinations under *Rapanos*, CWR, and NWPR. WOTUS-ML also classifies water resources into regulatory and hydrological categories. The Biden administration's CWA Rule (*2*) called for "machine learning and artificial intelligence methods to develop a jurisdictional status predictive model," which this paper provides. Our training data come from ACE records of 150,680 Approved Jurisdictional Determinations (AJDs), legally binding case-by-case decisions that ACE engineers make, which represent possible water resources (though 15% of AJDs are uplands). Existing research has not analyzed all these AJDs.

Jurisdictional determinations work as follows (fig. S1). A developer (e.g., a factory builder) where jurisdictional waters may be present can ask ACE to provide an AJD, which can take months, owing to backlogs or ACE's desire to observe a site in multiple seasons. A developer wishing to minimize delay and uncertainty or believing the water is jurisdictional may alternately provide a Preliminary Jurisdictional Determination (PJD). If the water has a PJD or if an AJD concludes the water is jurisdictional, the CWA requires the developer to obtain a Section 404 Permit, which may mandate compensatory investments, change development plans, or involve interactions with the Endangered Species Act (*11*). Nonjurisdictional waters face no CWA regulation. Development of jurisdictional waters without a permit can incur penalties and require the site be returned to its original state. Because we observe much of the data that ACE engineers use, our setup somewhat re-creates the ACE engineer's decision problem.

The WOTUS-ML model

We use the AJDs to train WOTUS-ML (fig. S2). The WOTUS-ML architecture is the widely used ResNet-18 convolutional neural network (Fig. 1) (*12*, *13*). We predict whether a site is regulated and which of nine hydrological (*14*) and nine legal classifications of water types the site represents, e.g., whether it is an isolated wetland or ephemeral stream, according to either the leading scientific classification of wetland and stream types, or according to a CWA rule's language (*13*) (tables S1 and S2). We pool data on the *Rapanos*, CWR, and NWPR rules and include an input layer identifying which rule each AJD used. Because *Sackett* AJDs begin in late 2023, after this study's timeframe, we do not train or predict on *Sackett*, though we discuss methodologies relevant to it (*13*). We divide the ground-truth data into disjoint test, training, and validation sets (*13*) (fig. S3). Because ACE requests AJDs to list water resource centroids, we interpret WOTUS-ML as better suited to classify whether a

[1]Department of Agricultural and Resource Economics, University of California, Berkeley, Berkeley, CA 94720, USA. [2]Goldman School of Public Policy, University of California, Berkeley, Berkeley, CA 94720, USA. [3]Resources for the Future, Washington, DC 20036, USA. [4]Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA. [5]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [6]Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [7]Department of Resource Economics, University of Massachusetts, Amherst, Amherst, MA 010013, USA. [8]Center for Agricultural and Rural Development, Iowa State University, Ames, IA 50011, USA. [9]National Bureau of Economic Research, Cambridge, MA 02139, USA. [10]Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA 94720, USA. [11]US Department of Energy, Washington, DC 20585, USA. [12]School of Information, University of California, Berkeley, Berkeley, CA 94720, USA. [13]Department of Economics, University of California, Berkeley, Berkeley, CA 94720, USA.
*Corresponding author. Email: sgreenhill@berkeley.edu (S.G.); joseph.shapiro@berkeley.edu (J.S.S.)
†These authors contributed equally to this work.

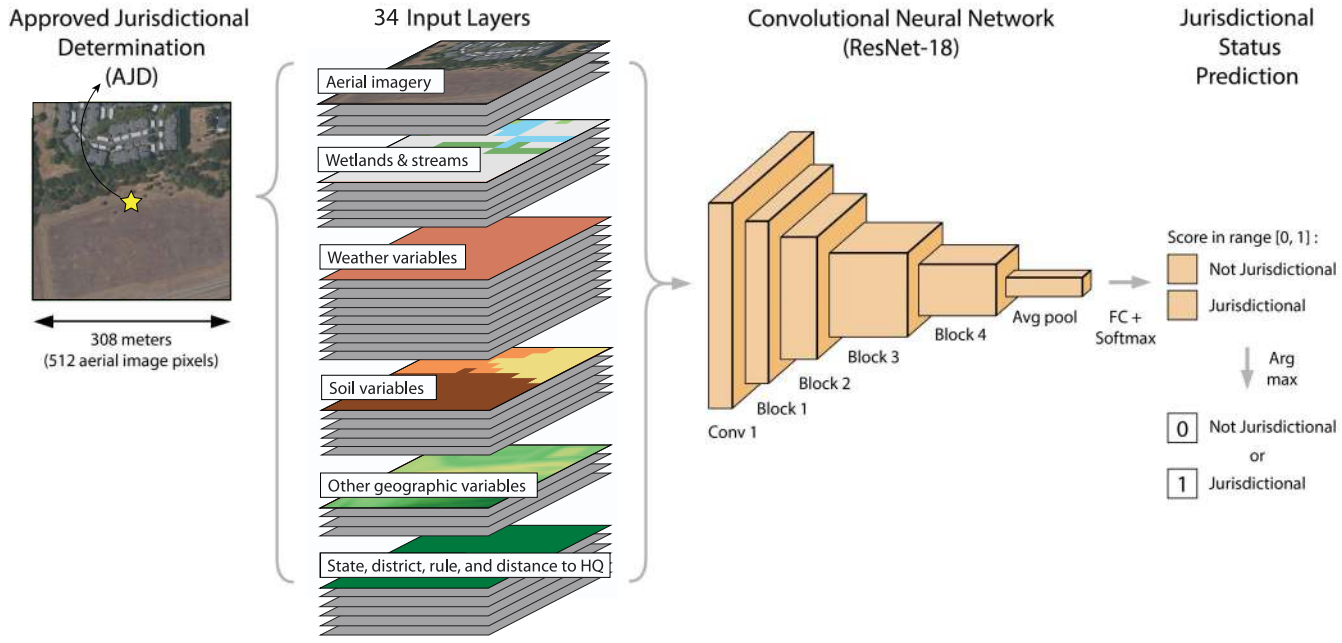


Fig. 1. WOTUS-ML model architecture uses a ResNet-18 with 34 input layers. The training data are images centered around an AJD. The model takes 34 input layers, described in fig. S5. The ResNet-18 architecture begins with a convolutional block (Conv 1), followed by four residual blocks (blocks 1 through 4), an average pooling operation, and finally a fully connected layer. The outputs of the fully connected layer are passed through a softmax function, producing a score in [0, 1]. We predict that a site is jurisdictional if the score exceeds 0.5.

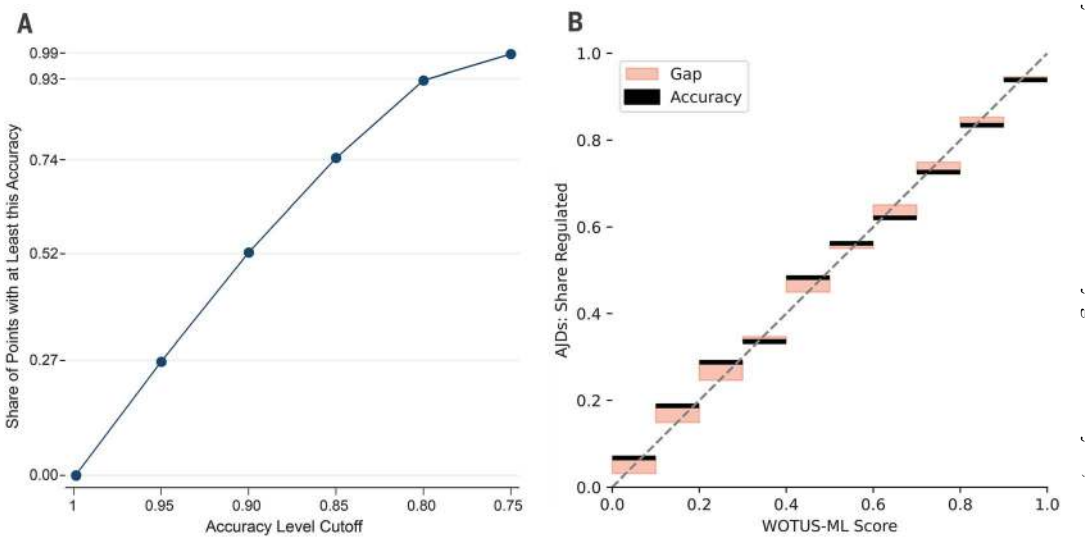


Fig. 2. WOTUS-ML scores allow unbiased estimates of regulatory probability. (**A**) WOTUS-ML estimates regulation with high accuracy for many sites. This figure finds threshold WOTUS-ML scores such that the average point beyond the threshold has at least a given accuracy in the AJD test set (0.95, 0.90, etc.). The vertical axis plots the share of points with WOTUS-ML scores beyond this cutoff. This indicates, for example, the share of points for which WOTUS-ML can predict AJD outcomes with 95% accuracy, with 90% accuracy, etc. Figure S9 provides details of calculations underlying (A). (**B**) WOTUS-ML scores reflect the probability of regulation. AJD test set is split into 10 equal-width bins containing AJDs with WOTUS-ML scores of 0.0 to 0.1, 0.1 to 0.2, etc. The black lines show the share of AJDs that are jurisdictional in each bin. The model's score is interpreted as the model's confidence that a given AJD is determined to be WOTUS. The red bar shows the gap between the model's average confidence and the share of jurisdictional AJDs in each bin. The dashed 45° line is the ideal jurisdictional share for each confidence level. If confidence and accuracy are equal, the model is calibrated and we can interpret the confidence score as a probability. If the red bar is below the diagonal, the model is too confident in its predictions, and vice versa.

resource is regulated than at delineating wetland boundaries.

The model receives as input an image of 34 layers, which include red, green, and blue (RGB) and near-infrared bands from aerial photographs from the National Agricultural Imagery Program (NAIP) (*15*); soil, groundwater, and elevation data from the Gridded National Soil Survey Geographic Database (*16*); hydrological information from the National Hydrography Dataset (NHD) (*17*, *18*); wetland coverage from the National Wetland Inventory (NWI) (*19*); ACE regulatory district and state boundaries (*20*); and related records (*13*) (table S3). Figure S4 maps several layers.

Figure S5 shows inputs for one site. We use images from at least several months before each AJD to avoid temporal leakage (*13*). Political boundaries might reflect political forces within a CWA rule that could change, though ACE engineers are not political appointees and we find stable patterns within ACE districts over time.

WOTUS-ML outputs scores ranging from zero to one that a given latitude and longitude pair is regulated, separately for *Rapanos*, NWPR, and CWR. Zero represents confidence the point is unregulated; one represents confidence the point is regulated (fig. S6). When evaluating predictive accuracy, we round scores to binary predictions—"regulated" versus "not regulated."

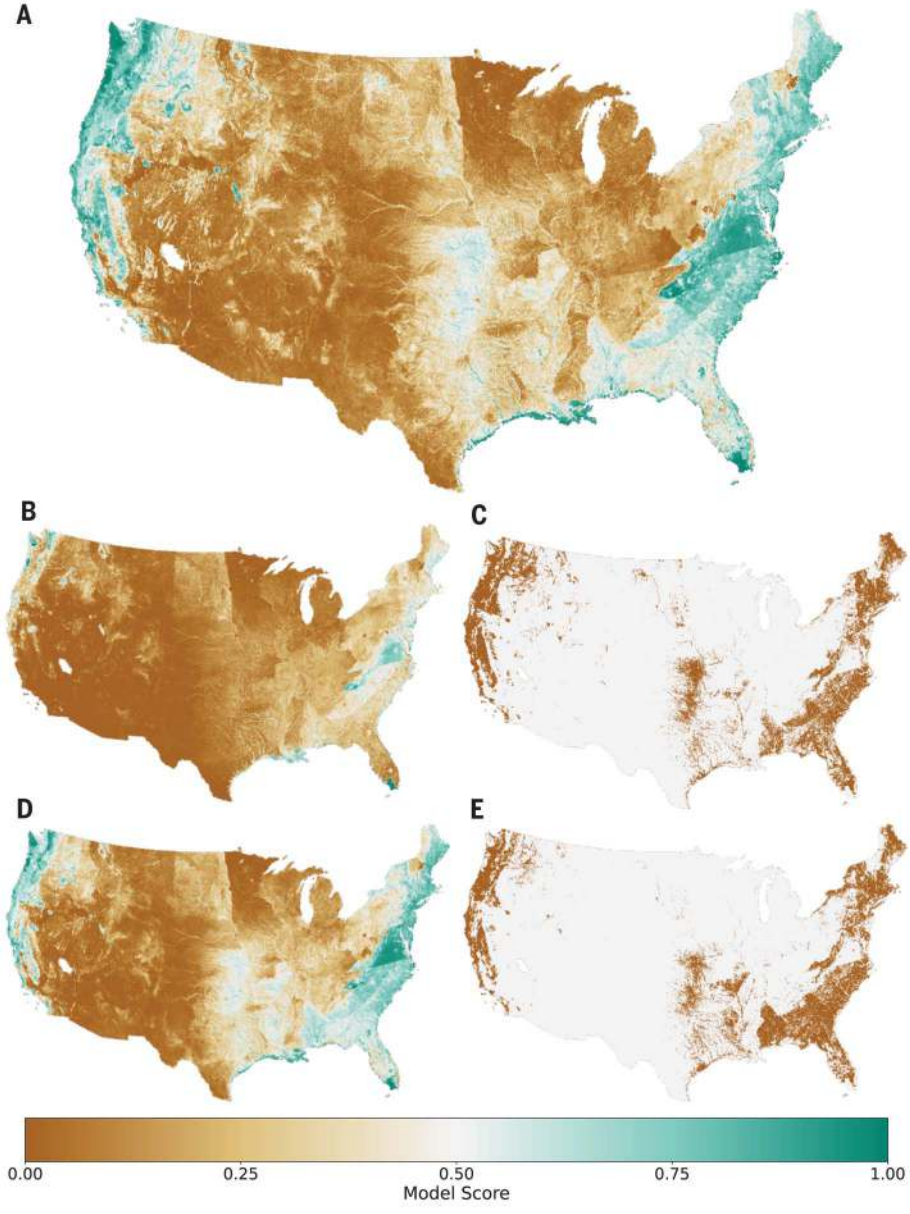


Fig. 3. Estimated probability of CWA regulation for 4 million prediction points across the USA. (**A**) Estimated probability of CWA regulation (WOTUS-ML score) under *Rapanos*. (**B**) Estimated regulatory probability under NWPR. (**C**) Estimated regulation changes from *Rapanos* to NWPR. (**D**) Estimated regulatory probability under CWR. (**E**) Estimated regulation changes from CWR to NWPR. A "regulation change" describes when the WOTUS-ML binary classification score (>50%, <50%) changed status. Brown represents deregulation, green represents new regulation. Map creates a 247 by 576 grid and displays the mean model score in each bin (~28 prediction points per bin).

WOTUS-ML also outputs scores for each of nine Cowardin codes and nine resource types.

Model accuracy and bias

WOTUS-ML correctly predicts outcomes for 79% of AJDs in a held-out test set (table S4). The area under the model's receiver operating characteristic curve (AUC) is 0.85 (fig. S7). Among test set AJDs, 35% are regulated, so model learning accounts for a 14 percentage point improvement in accuracy above a naïve baseline. We measure model learning as the difference between accuracy and max (share regulated, share unregulated). Type I and II errors vary by rule (table S5). WOTUS-ML has an accuracy below 100% for several reasons (*13*).

WOTUS-ML accuracy varies across settings (tables S4 and S6 and S7). Test set accuracy is 82% for AJDs without an ACE field visit and 74% for AJDs with a field visit. Field visits provide information unavailable to WOTUS-ML (*21*). WOTUS-ML has similar accuracy on wetlands and streams but greater accuracy for estuaries, which are always regulated. The model is extremely accurate in ACE districts such as St. Paul, which covers Minnesota and Wisconsin, and where regulation rates are low. WOTUS-ML has an accuracy around 80% for sites with characteristics typical of AJDs and for sites very different from typical AJDs, which supports its external validity nationally (*13*) (fig. S8). WOTUS-ML has 75% accuracy for identifying ephemeral resources, which is useful given the absence of national maps of such resources (table S8). We focus on WOTUS-ML's binary analysis of jurisdiction, which has greater accuracy than its resource type and Cowardin code predictions.

WOTUS-ML predicts a subset of sites with high accuracy (Fig. 2A and fig. S9). In 27% of sites with scores below 0.07 or above 0.95, WOTUS-ML has 95% accuracy. If a developer used WOTUS-ML for these sites, ACE would agree for 95% of sites. For 52% of sites, where the score is below 0.17 or above 0.83, WOTUS-ML has 90% accuracy. In such sites, WOTUS-ML could save resources. The mean Section 404 permit costs $5,000 to $39,000 (*2*). Although we are unaware of cost estimates for AJDs, if WOTUS-ML saved this amount in delay and uncertainty for each AJD where WOTUS-ML has 95% accuracy, it would save $209 million to $1.6 billion over our sample. If WOTUS-ML let developers avoid Section 404 permit costs for the 20% of PJDs where WOTUS-ML has 95% confidence that the PJD is not regulated, it could save $150 million to $1.2 billion annually. These illustrative numbers demonstrate potentially high returns to efficient adjudication (*21*).

WOTUS-ML scores are well-calibrated to probabilities, because they provide an unbiased estimate of the probability of regulation (Fig. 2B).

For example, when WOTUS-ML outputs a score between 0.3 and 0.4 for test-set AJDs, 34% of those AJDs are regulated. Hence, we refer to WOTUS-ML scores as predicted regulatory probabilities. Unbiasedness supports WOTUS-ML's use as a decision support tool.

Required accuracy may vary by purpose (*13*). For example, developers might most value a signal with 95% accuracy of whether a site is regulated, to decide whether to provide a PJD or request an AJD. By contrast, ACE might value knowledge of nonextremal WOTUS-ML scores, which could help focus ACE resources on ambiguous cases.

Opening the black box

Feature importance analysis clarifies how a complex model like WOTUS-ML functions, though has limitations (*13*). We use permutation tests to elucidate which input layers WOTUS-ML relies on for predictions (*22*). We randomly permute groups of layers across samples, breaking the link between that feature and its label. A feature's permutation importance equals the difference between the accuracy with features intact and the accuracy with that feature permuted.

When permuted across all samples nationally, climate data most affect model accuracy (fig. S10A), perhaps because precipitation and temperature predict streamflow and wetland prevalence. Stream and wetland vector data are second-most important, which is intuitive because they are the regulated entities. Next are state, district, and rule measures, reflecting the dependence of jurisdiction on nongeophysical features. Elevation, soil, and groundwater characteristics also matter. NAIP imagery is less important, perhaps because other layers are derived from remote-sensing data.

We also perform within-state permutation tests by shuffling layers solely within state samples (fig. S10B). This indicates which features help WOTUS-ML replicate AJDs within a state, which more closely resembles the work of ACE engineers. Within a state, stream and wetland data most account for model accuracy. ACE engineers consult these datasets in deciding AJDs (*21*).

WOTUS-ML predictions of regulatory probabilities and changes

Using WOTUS-ML, we predict regulatory probabilities for 4 million random prediction points across the US, plus a random sample of PJDs and traditional navigable waterways (*13*, *23*). Because WOTUS-ML accuracy is independent of the probability that a point is an AJD, our model evaluation using the AJD test set provides useful information about the model's performance on the prediction points (supplementary text section B.4 and fig. S8).

Rapanos regulates 22% of points in the US; NWPR regulates 8% (Table 1). Areas where a land-use model predicts that development will occur in the 2020s (*24*) have slightly higher levels of regulation than random prediction points. Because AJDs represent potential water resources, the mean AJD has 15 percentage points greater jurisdictional probability under *Rapanos* than the mean random point. Nearly half of PJDs are jurisdictional under *Rapanos*; 28% are jurisdictional under NWPR. Hence, developers may incorrectly assume that sites are jurisdictional and request PJDs rather than AJDs; or they may request PJDs for other reasons, including expediting processing under other federal regulations. These possibilities demonstrate the potential value of WOTUS-ML as a decision support tool. Two-thirds of NHD streams are regulated under *Rapanos*, but only 46% under NWPR (Table 1). NWPR thus deregulates 686,000 stream miles (table S9 provides state-level calculations), more than every river and stream in California, Florida, Illinois, New York, Ohio, Pennsylvania, and Texas combined. Compared to *Rapanos*, NWPR deregulates 30% of stream and wetland areas in subwatersheds that provide drinking water for the average American (table S10).

We also analyze differences across stream types (Table 1). Our estimate that 100% of traditional navigable rivers are regulated under either rule gives confidence in WOTUS-ML because all rules regulate traditional navigable waterways (*23*). *Rapanos* regulates 55% of intermittent and ephemeral streams, but NWPR regulates 30%. Thus, NWPR deregulates 45% of regulated intermittent and ephemeral streams, though NWPR only deregulates 21% of all national streams.

Rapanos regulates 52% of wetlands, whereas NWPR regulates 27% of all wetlands (Table 1). Thus, NWPR removes jurisdiction for just under half of regulated wetlands, or 25% of all wetlands. This 25% statistic is far below the amicus and media assertions that NWPR deregulates 51% of all wetlands (*9*). NWPR deregulates over a third of wetlands that are adjacent to or abutting a stream or river, and two-thirds of isolated wetlands. NWPR deregulates 35 million wetland acres (table S9). This represents 15% of wetland area in the continental US at the time of European settlement, or over a fourth of the wetlands that disappeared between the time of European settlement and today (*25*). This deregulated wetland area represents $12 billion to $23 billion in annual flood mitigation benefits, or $250 billion to $458 billion in present-value flood mitigation benefits, discounted at 5%. The deregulated wetland area represents $249 billion to $381 billion of land value. Additionally, this large wetland area provides additional important

Table 1. What does the Clean Water Act regulate? "ICLUS" (Integrated Climate and Land Use Scenarios) (*24*), "Rivers and streams," and "Wetlands" describe subsets of the four million prediction points. Table shows the share of points that the CWA regulates, measured as the share with a WOTUS-ML score above 50%. For navigable waters, we evaluate jurisdiction based on the mean WOTUS-ML score among points within each named river, because CWA regulates each water resource, and report the share of named rivers predicted as jurisdictional. "WOTUS-ML resource type" refers to points where WOTUS-ML predicts that the listed resource type is the most likely under either *Rapanos* or NWPR, according to the classification listed in table S2.

	Rapanos	**NWPR**
General groups of points		
All 4 million points	0.22	0.08
Urban growth areas (ICLUS)	0.30	0.09
AJD test set	0.37	0.16
Preliminary jurisdictional determinations (PJDs)	0.48	0.28
Rivers and streams		
All (NHD all)	0.67	0.46
Traditional navigable waters (NHD+law)	1.00	1.00
Perennial (NHD 46006)	0.83	0.67
Intermittent or ephemeral (NHD 46003, 46007)	0.55	0.30
Wetlands		
All (NWI palustrine)	0.52	0.27
Emergent (NWI)	0.47	0.20
Forested (NWI)	0.59	0.32
Adjacent or abutting (WOTUS-ML resource type)	0.88	0.57
Isolated (WOTUS-ML resource type)	0.39	0.14

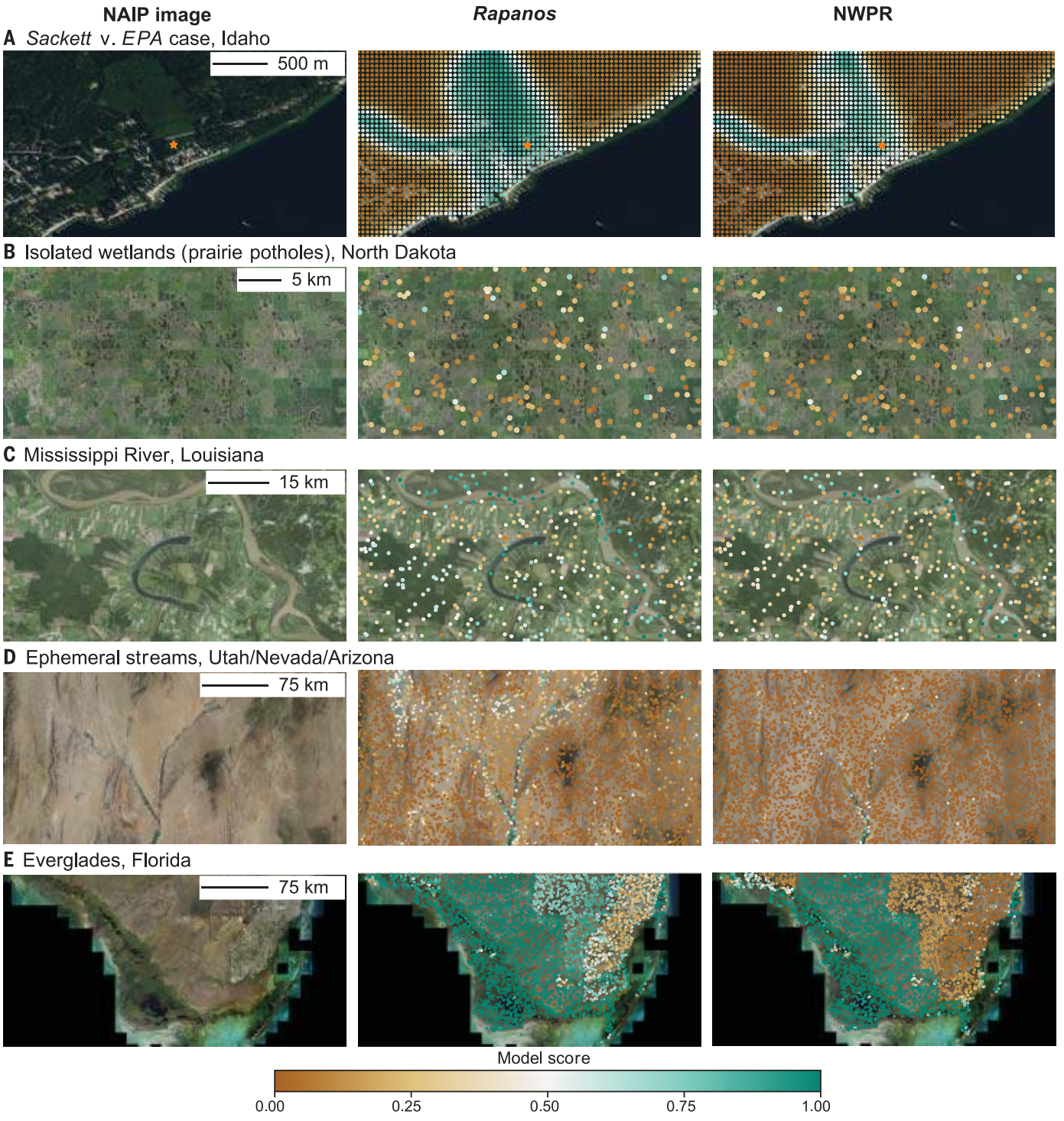


Fig. 4. Case studies of CWA regulation reveal local heterogeneity and differences across rules. (**A**) A densely sampled area around the Sackett property where jurisdiction is the subject of *Sackett* v. *EPA*, near Priest Lake, Idaho. The property is marked with an orange star. Points on this property have a mean model score of about 0.5, consistent with the ambiguity that produced litigation. Areas within and near wetlands have higher scores, though scores around the edge of the large centrally located wetland decrease under NWPR. Average model scores under *Rapanos* and NWPR are 0.34 and 0.22, respectively. (**B**) Isolated wetlands (prairie potholes), Benson County, North Dakota. Points near prairie potholes are more likely to be regulated under *Rapanos* and deregulated under NWPR. ACE categorizes 80% of AJDs in Benson County, ND, as isolated wetlands. Average model scores under *Rapanos* and NWPR are 0.22 and 0.17, respectively. (**C**) Farmland along the Mississippi River, Baton Rouge, Louisiana. WOTUS-ML predicts that most points around the Mississippi are regulated under *Rapanos* and NWPR. Farmland, which the CWA explicitly ignores, receives lower scores under both rules. Average model scores under

Rapanos and NWPR are 0.45 and 0.35, respectively. (**D**) Ephemeral streams north of Lake Mead, near Nevada, Utah, and Arizona borders. WOTUS-ML predicts ephemeral streams in the north-central part of the image lose regulation from *Rapanos* to NWPR. Lake Mead remains regulated under both rules. The southwestern part of the image includes Solar Energy Zones near Dry Lake, Nevada, where renewable energy development is occurring and requires AJDs. ACE categorizes 69% of AJDs in Utah and Nevada under NWPR as ephemeral streams. Average model scores under *Rapanos* and NWPR are 0.09 and 0.01, respectively. (**E**) Southern Florida, including the Everglades and Miami. WOTUS-ML predicts that *Rapanos* and NWPR regulate Everglades National Park and most other protected wildlife areas. Wetlands and developed areas along the northwestern and eastern image edges have much lower model scores under NWPR. Average model scores under *Rapanos* and NWPR are 0.85 and 0.64, respectively. We randomly choose foreground and background ordering of points in all panels.

species habitat protection, recreational opportunity, and other ecosystem values.

Rapanos jurisdiction reflects geophysical and political patterns. *Rapanos* regulates wetlands in the coastal South and mid-Atlantic and coastal streams and wetlands near the Pacific. *Rapanos* regulates less of the arid West, though *Rapanos* regulates some ephemeral streams. Parts of the Fall Line separating the Coastal Plain in the mid-Atlantic and South have discrete changes in jurisdiction. Major waterways are visible because their jurisdiction contrasts with lower jurisdiction for surrounding areas. ACE and state boundaries reveal differences (Fig. 3A). For example, the New England ACE district concludes that most AJDs are jurisdictional, whereas the St. Paul district concludes that few are.

Under NWPR, most predicted jurisdictional probabilities are below 20% (Fig. 3B and fig. S6, E and F). Major waterways remain regulated, though with narrower channels, potentially owing to decreased jurisdiction of nearby wetlands. The least jurisdiction is in the arid West, where ephemeral streams are common.

Between *Rapanos* and NWPR, regulation decreases most around isolated wetlands in the mid-Atlantic and Gulf coast (Fig. 3C). WOTUS-ML shows no substantial areas of increased regulation under NWPR (*13*). The arid west shows limited areas of decreased regulation.

CWR has broadly similar jurisdiction patterns to *Rapanos* (Fig. 3D and fig. S6), with less coverage of ephemeral streams in the arid West. CWR, like *Rapanos*, has higher coverage than NWPR across the country (Fig. 3E).

Patterns within the 38 ACE districts are somewhat persistent, which supports our methodology's medium-run validity. In all years, St. Paul is in the bottom 10 districts in share of jurisdictional AJDs and Norfolk is in the top 10. New England always has among the fewest AJDs of any district. One exception is Florida, where politics led to decentralization of the AJD process in 2020.

Case studies give confidence in the model's results, illuminate new patterns, and clarify what CWA rules regulate (Fig. 4). Regulation is heterogeneous within narrow and broad geographic areas. In case studies A through C and E, some wetlands are regulated but nearby forests and farms are not. In arid streams around the Arizona, Nevada, and Utah borders, *Rapanos* regulates some ephemeral streams and NWPR deregulates most (Fig. 4D). Each case study has points near 0.50, where regulation is uncertain, and points closer to 0 and 1, where regulation is more certain.

Discussion

To inform prominent debates about which water resources should be regulated, it is important to know which resources are regulated. WOTUS-ML provides such evidence for recent CWA rules. It reveals enormous sets of natural resources that have changed jurisdiction in the past decade. Stakeholders can use WOTUS-ML to support decision-making for individual water resources or aggregate policy design and evaluation, and potentially save large costs by reducing uncertainty and delays. This represents one set of insights that ML algorithms can provide for more general regulatory implementation problems where regulators must repeatedly interpret and apply a law. Regulators have expressed mixed views on geophysical map tools (*2*, *7*). Our analysis provides a basis for caution, in that WOTUS-ML has imperfect accuracy. It also provides a basis for cautious optimism, in the ways WOTUS-ML can provide insight on one of the most complex and controversial US environmental policies.

Besides clarifying which waters the CWA regulates, WOTUS-ML can support stakeholder decision-making. A developer can use WOTUS-ML to learn the estimated probability that the CWA regulates a site. ACE can use WOTUS-ML to provide input to determining jurisdiction. The White House and EPA can use WOTUS-ML to forecast impacts of rules changing which waters are jurisdictional. State environmental agencies can use WOTUS-ML to help regulate more than federal law requires (e.g., to support enforcement of a state-level wetland rule approximating *Rapanos* under a federal *Sackett* rule). Environmental or industry associations can use WOTUS-ML to provide statistics for court briefs.

We offer a template for applying existing algorithms to regulatory implementation problems, where agencies repeatedly interpret court rulings or laws. ACE engineers interpret language from *Rapanos*, CWR, or NWPR using data and visits to determine whether water resources are regulated. Research has used related algorithms for other types of policy problems—for example, where optimal decisions depend on future events such as whether a defendant will jump bail (*26*), or predicting environmental conditions such as whether an aquifer has arsenic (*27*). ML has predicted court decisions (*28*), which differs from modeling regulators' implementation of such decisions. We show how ML can clarify regulatory implementation of environmental law, with potential relevance to many regulations requiring practitioners to interpret and apply textual directives.

REFERENCES AND NOTES

1. S. M. P. Sullivan, M. C. Rains, A. D. Rodewald, W. W. Buzbee, A. D. Rosemond, *Science* **369**, 766–768 (2020).
2. US Environmental Protection Agency and US Army Corps of Engineers, "Economic Analysis for the Final "Revised Definition of `Waters of the United States'" Rule" (2022); https://www.epa.gov/system/files/documents/2022-12/2022_WOTUS%20EA_FinalRule_508.pdf.
3. C. Taylor, H. Druckenmiller, *Am. Econ. Rev.* **112**, 1334–1363 (2023).
4. S. Zellmer, *Notre Dame Law Rev.* **88**, 2323 (2013); https://scholarship.law.nd.edu/ndlr/vol88/iss5/12/.
5. C. Davenport, "Trump Removes Pollution Controls on Streams and Wetlands," *New York Times*, 22 January 2020.
6. S. McFetridge, "Judge temporarily blocks clean water rule in 24 states", AP News, 12 April 2023.
7. US Environmental Protection Agency and US Army Corps of Engineers, "Mapping and the Navigable Waters Protection Rule" (2020); https://19january2021snapshot.epa.gov/sites/static/files/2020-01/documents/nwpr_fact_sheet_-_mapping.pdf.
8. A. Wittenberg, K. Bogardus, "EPA claims 'no data' on impact of weakening water rule. But the numbers exist" (E&E News, 2018); https://www.science.org/content/article/epa-claims-no-data-impact-weakening-water-rule-numbers-exist.
9. K. A. Fesenmyer, S. J. Wenger, D. S. Leigh, H. M. Neville, *Freshw. Sci.* **40**, 252–258 (2021).
10. *Sackett v. EPA*. "Brief *amicus curiae* of the National Association of Wetland Managers, the Association of Floodplain Managers, the American Planning Association, the American Water Works Association, and the New England Interstate Water Pollution Control Commission" (2022); https://www.supremecourt.gov/DocketPDF/21/21-454/228256/20220617104921002_21-454%20Brief.pdf.
11. US Army Corps of Engineers, "Compliance with the Federal Endangered Species Act" (2023); https://www.spk.usace.army.mil/Missions/Regulatory/Permitting/Endangered-Species-Agency-Consultations/.
12. K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition" in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, (2016), pp. 770-778; https://doi.org/10.1109/CVPR.2016.90.
13. Materials and methods are available as supplementary materials.
14. L. M. Cowardin, V. Carter, F. C. Golet, E. T. LaRoe, "Classification of wetlands and deepwater habitats of the United States" (FWS/OBS-79/31, U.S. Fish and Wildlife Service, 1979).

15. US Department of Agriculture, National Agriculture Imagery Program (January 2023); https://datagateway.nrcs.usda.gov/GDGHome_DirectDownLoad.aspx.
16. US Department of Agriculture, Gridded National Survey Geographic Database (October 2022); https://www.nrcs.usda.gov/resources/data-and-reports/gridded-national-soil-survey-geographic-database-gnatsgo.
17. US Environmental Protection Agency, National Hydrography Dataset Plus V2 (October 2022); https://www.epa.gov/waterdata/nhdplus-national-data.
18. S. A. R. Colvin *et al*., *Fisheries* **44**, 73–91 (2019).
19. US Fish & Wildlife Service, National Wetlands Inventory (October 2022); https://www.fws.gov/program/national-wetlands-inventory/download-state-wetlands-data.
20. US Army Corps of Engineers, US Army Corps of Engineers Regulatory Boundary (January 2023); https://geospatial-usace.opendata.arcgis.com/datasets/70805e1a8fd74e42b0a9585088d6d151_0/about.
21. US Army Corps of Engineers, "Corps of Engineers Wetlands Delineation Manual" (Y-87-1, 1987); https://usace.contentdm.oclc.org/digital/collection/p266001coll1/id/4530.
22. A. Altmann, L. Toloşi, O. Sander, T. Lengauer, *Bioinformatics* **26**, 1340–1347 (2010).
23. USCS § 33 1804 (2023).
24. US Environmental Protection Agency, "Land-Use Scenarios: National-Scale Housing-Density Scenarios Consistent with Climate Change Storylines" (EPA/600/R-08/076F, 2009); https://assessments.epa.gov/iclus/document/&deid=203458.
25. T. E. Dahl, "Status and Trends of Wetlands in the Conterminous United States 2004 to 2009" (U.S. Fish & Wildlife Service, 2011); https://www.fws.gov/wetlands/documents/Status-and-Trends-of-Wetlands-in-the-Conterminous-United-States-2004-to-2009.pdf.
26. J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, S. Mullainathan, *Q. J. Econ.* **133**, 237–293 (2018).
27. J. Podgorski, M. Berg, *Science* **368**, 845–850 (2020).
28. D. M. Katz, M. J. Bommarito 2nd, J. Blackman, *PLOS ONE* **12**, e0174698 (2017).
29. S. Greenhill *et al*., (2023). Code for Machine Learning Predicts Which Rivers, Streams, and Wetlands the Clean Water Act Regulates [Dataset], Zenodo; https://doi.org/10.5281/zenodo.10108709.
30. US Geological Survey, 3-Dimensional Elevation Program (November 2022); https://www.usgs.gov/3d-elevation-program.
31. US Environmental Protection Agency, U.S. EPA Ecoregions, Level IV Ecoregion (January 2023); https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states.
32. Prism Climate Group, Parameter-elevation Regressions on Independent Slopes Model, 30-year Normals (January 2023); https://prism.oregonstate.edu/normals/.
33. US Geological Survey, National Land Cover Database (December 2022); https://www.usgs.gov/centers/eros/science/national-land-cover-database#overview.
34. US Census Bureau, TIGER/Line Shapefiles (December 2022); https://www2.census.gov/geo/tiger/TIGER2022/.
35. S. Greenhill *et al*., (2023). Training data from: Machine Learning Predicts Which Rivers, Streams, and Wetlands the Clean Water Act Regulates [Dataset], Dryad; https://doi.org/10.5061/dryad.m63xsj47s.
36. S. Greenhill *et al*., Prediction data from: Machine Learning Predicts Which Rivers, Streams, and Wetlands the Clean Water Act Regulates [Dataset], Dryad (2023); https://doi.org/10.5061/dryad.z34tmpgm7.

ACKNOWLEDGMENTS

Participants at AERE, the California Water Boards, the Energy Institute at Haas, EPA, Global Policy Lab, Stanford, and TWEEDS seminars, E. Benami, M. Burke, J. Corona, C. Costello, T. Doley, J. Hewitt, S. Hsiang, N. Kohli, R. Kwok, M. Massey, A. McGartland, S. Mullainathan, K. Swann, and C. Taylor provided useful comments and discussions. We thank B. Walker for research assistance. **Funding:** This material is based upon work supported by the following: Ciriacy-Wantrup Postdoctoral Fellowship (S.W.); The Giannini Foundation (J.S.S.); Google Cloud Research Credits Program (H.D., D.A.K., J.S.S.); NIH grant 1R01AG079914-01 (D.A.K., J.S.S.). **Author contributions:** Conceptualization: J.S.S.. Data curation: S.G., H.D., S.W., D.A.K., J.S.S. Formal analysis: S.G., H.D., S.W., D.A.K., J.K.M., N.Y., A.T., J.S.S. Funding acquisition: H.D., D.A.K., M.G., A.T., J.S.S. Investigation: S.G., H.D., S.W., D.A.K., M.G., J.K.M., N.Y., A.T., J.S.S. Methodology: S.G., H.D., S.W. Software: S.G., H.D., S.W., D.A.K., J.S.S. Visualization: S.G., H.D., S.W., J.S.S. Writing – original draft: S.G., H.D., S.W., D.A.K., J.S.S. Writing – review and editing: S.G., H.D., S.W., D.A.K., J.S.S. **Competing interests:** D.A.K. serves on the scientific advisory board of the EPA and serves as an expert witness on water pollution litigation. During the research, J.S.S. served as an adviser on trade and environment to the European Commission's Directorate General of Trade. D.A.K. and J.S.S. wrote a public report on EPA's analyses of the CWR and the NWPR as part of the External Environmental Economics Advisory Committee, funded by the Sloan Foundation. All authors of this paper are inventors on a patent pending with serial number 63/513,464, submitted by UC Berkeley, which covers WOTUS-ML. The software has a Creative Commons Attribution Non Commercial No Derivatives 4.0 International license, which freely allows use for research and other noncommercial purposes. Potential commercial users should contact the Office of Technology Licensing at UC Berkeley. **Data and materials availability:** All code is available on Zenodo (*29*). All data are publicly available online (*15–17*, *19*, *20*, *30–34*). All data used to train the models are on Dryad (*35*), as is a subset of the prediction data, as well as model weights and all data required to produce the results presented here (*36*). The model is freely available for research and noncommercial purposes (*29*, *35*, *36*), but its commercial use is limited (patent pending).

SUPPLEMENTARY MATERIALS

Submitted 21 April 2023; accepted 18 December 2023
10.1126/science.adi3794

IMMUNOLOGY

The transcription factor ZEB2 drives the formation of age-associated B cells

Dai Dai[1,2,3]†, Shuangshuang Gu[1]†, Xiaxia Han[1]†, Huihua Ding[1,2], Yang Jiang[1], Xiaoou Zhang[4,5], Chao Yao[6], Soonmin Hong[1], Jinsong Zhang[6], Yiwei Shen[1], Guojun Hou[1,2], Bo Qu[1,2], Haibo Zhou[1,2], Yuting Qin[1,2], Yuke He[1,2], Jianyang Ma[1,2], Zhihua Yin[7], Zhizhong Ye[7], Jie Qian[1], Qian Jiang[8], Lihua Wu[8], Qiang Guo[1], Sheng Chen[1], Chuanxin Huang[9], Leah C. Kottyan[10,11], Matthew T. Weirauch[10,11], Carola G. Vinuesa[2,12]*, Nan Shen[1,2,3,7,10,11]*

Age-associated B cells (ABCs) accumulate during infection, aging, and autoimmunity, contributing to lupus pathogenesis. In this study, we screened for transcription factors driving ABC formation and found that zinc finger E-box binding homeobox 2 (ZEB2) is required for human and mouse ABC differentiation in vitro. ABCs are reduced in *ZEB2* haploinsufficient individuals and in mice lacking *Zeb2* in B cells. In mice with toll-like receptor 7 (TLR7)–driven lupus, ZEB2 is essential for ABC formation and autoimmune pathology. ZEB2 binds to +20-kb myocyte enhancer factor 2b (*Mef2b*)'s intronic enhancer, repressing MEF2B-mediated germinal center B cell differentiation and promoting ABC formation. ZEB2 also targets genes important for ABC specification and function, including *Itgax*. ZEB2-driven ABC differentiation requires JAK-STAT (Janus kinase-signal transducer and activator of transcription), and treatment with JAK1/3 inhibitor reduces ABC accumulation in autoimmune mice and patients. Thus, ZEB2 emerges as a driver of B cell autoimmunity.

Age-associated B cells (ABCs) are a distinct effector B cell subset found at increased numbers in aged female mice, infection models, and systemic autoimmune diseases (*1*). ABCs are identified as $CD11c^+CD11b^+CD21^-CD23^-$T-$bet^+$ in mice (*2*, *3*) and $CD11c^+CD21^-CD27^-CXCR5^-$ $FCRL5^+IgD^-$T-bet^+ in humans (*4*, *5*). In autoimmune settings, these B cells are enriched for autoantibody specificities and are thought to be antigen-experienced. Moreover, there is evidence that ABCs can persist in tissues and rapidly differentiate into antibody-secreting cells (ASCs) upon antigen reencounter or innate stimulation (*1*).

The transcription factors (TFs) T-bet (T box expressed in T cells), IRF5 (interferon regulatory factor 5), and IRF8 are highly expressed in ABCs and have been put forward as functional regulators of ABC differentiation (*6*–*9*). However, except for immunoglobulin G 2a/c (IgG2a/c) isotype switching, T-bet is dispensable for ABC accumulation and maintenance of ABC features (*9*–*11*). IRF5 and IRF8 are broadly expressed in other B cell subsets and have been reported to be involved in cell activation, proliferation, differentiation, and function (*12*–*14*). To determine the TF(s) essential for ABC formation, we screened all TFs expressed by these cells and identified ZEB2, a member of the zinc-finger E homeobox-binding protein family, as the key regulator required for ABC specification and differentiation in mice and humans.

[1]Shanghai Institute of Rheumatology, Shanghai Renji Hospital, Shanghai Jiaotong University School of Medicine (SJTUSM), Shanghai, China. [2]Centre for Personalised Immunology (CACPI), Shanghai Renji Hospital, SJTUSM, Shanghai, China. [3]State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Shanghai Renji Hospital, SJTUSM, Shanghai, China. [4]Shanghai Key Laboratory of Maternal and Fetal Medicine, Clinical and Translational Research Center of Shanghai First Maternity and Infant Hospital, Shanghai, China. [5]Frontier Science Center for Stem Cell Research, School of Life Sciences and Technology, Tongji University, Shanghai, China. [6]Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [7]Shenzhen Futian Hospital for Rheumatic Diseases, Shenzhen, China. [8]Department of Medical Genetics, Capital Institute of Pediatrics, Beijing, China. [9]Center for Immune-Related Diseases at Shanghai Institute of Immunology, Ruijin Hospital, SJTUSM, Shanghai, China. [10]Center of Autoimmune Genomics and Etiology, Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [11]Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA. [12]Francis Crick Institute, London, UK.
*Corresponding author. Email: nanshensibs@gmail.com (N.S.); carola.vinuesa@cric.ac.uk (C.G.V.)
†These authors contributed equally to the work

Screen for TFs that direct ABC differentiation

To gain insights into the nature of ABCs, we sorted peripheral B cells from a patient with new-onset systemic lupus erythematosus (SLE) (table S1) and performed droplet-based single-cell RNA sequencing (scRNA-seq). Seven distinct clusters were revealed through unsupervised clustering with a two-dimensional uniform manifold approximation and projection (UMAP) (Fig. 1A and fig. S1, A and B). These clusters were assigned to known peripheral B cell subsets, including transitional B cells, naïve B cells, activated naïve B cells, ABCs, memory B cells, plasmablasts, and plasma cells, by comparing differentially expressed genes with established landmark genes (*5*, *15*) (fig. S1, A to C). ABCs preferentially expressed genes that encode the key surface markers *CD19*, *CD86*, *FCRLA*, *FCRL3*, *FCRL5*, *FCGR2B*, *MS4A1*, and *ITGAX*, and they lacked *CD27*, *CR2*, *CXCR5*, *FCER2*, and *IGHD* (Fig. 1B and fig. S1C). We found 43 differentially expressed TFs: 27 up-regulated and 16 down-regulated (Fig. 1C). We sorted mouse $CD19^+CD11c^+CD21^-$ B cells from the bm12-induced lupus mouse model and valida[illegible] 40 mouse homologs of the differentially [illegible] pressed TFs identified in the human scRNA-seq data (Fig. 1D and fig. S2, A to C). Among these, 13 up-regulated (*Tbx21*, *Zeb2*, *Plek*, *Litaf*, *Tfeb*, *Nfatc2*, *Zbtb32*, *Srebf2*, *Jazf1*, *Jun*, *Sox5*, *Tfec*, and *Batf*) and 3 down-regulated TFs (*Ets1*, *Mbd4*, and *Fli1*) were identically regulated in human and mouse ABCs and were therefore considered potential transcriptional regulators of ABC differentiation.

To identify which of these 16 TFs were driving ABC differentiation, we retrovirally infected B cells from Cas9 transgenic mice and CD45.1 congenic mice with single-guide RNA (sgRNA) plasmids targeting each TF and coexpressing blue fluorescent protein (BFP) and then cultured these cells with the ABC differentiation cocktail (*7*, *16*) (fig. S3A). The ratio between live BFP^+ $CD45.1^-$ and BFP^+ $CD45.1^+$ ABCs was determined, and genome editing was validated with a sequence specific for *Itgax* (fig. S3, B and C). We identified seven TFs that could significantly ($P < 0.05$) alter ABC formation (Fig. 1, E and F, and fig. S3D). Except for *Zeb2*, ablation of the other six TFs predominantly influenced cell viability (fig. S3, E and F). Two different sgRNAs targeting *Zeb2* in separate $Cas9^+$ B cell cultures led to reduced ABC formation (Fig. 1G), excluding an off-target editing effect. To determine whether any of the other six TFs was required for human ABC lineage formation, we transduced Cas9-gRNA ribonucleoprotein (RNP) complexes through electroporation and cultured edited cells with the ABC differentiation cocktail (*4*–*6*) (fig. S4, A and B). Ablation of *TBX21*, *ZEB2*, and *SREBF2* dampened ABC differentiation in both human and mouse B cells, but *TBX21* and *SREBF2* deficiency also led to altered B cell viability (Fig. 1H and fig. S4, C to E). Gene editing of *ETS1* and *JUN* had opposite effects, and *BATF* and *FLI1* did not change human ABCs ($P < 0.05$) (Fig. 1H and fig. S4C). To exclude the possibility of the off-target effect, two different gRNAs targeting *ZEB2* were used separately, and both were found to impair human ABC induction (Fig. 1, I and J). Thus, collectively, ZEB2 emerges as the most promising putative ABC transcriptional regulator.

ZEB2 haploinsufficiency impairs human ABC formation

ZEB2 is pivotal in early fetal development and cancer progression because it drives the epithelial-to-mesenchymal transition (*17*). Loss-of-function heterozygous gene variants lead to Mowat-Wilson syndrome (MWS), a rare genetic disorder in which *ZEB2* haploinsufficiency causes intellectual disability, distinctive facial features, seizures, and a predisposition to Hirschsprung disease (*18*, *19*). Although mouse studies have revealed ZEB2's role in immune cell differentiation and function (*20*), insights into the immunological consequences of reduced ZEB2 function in MWS patients are limited.

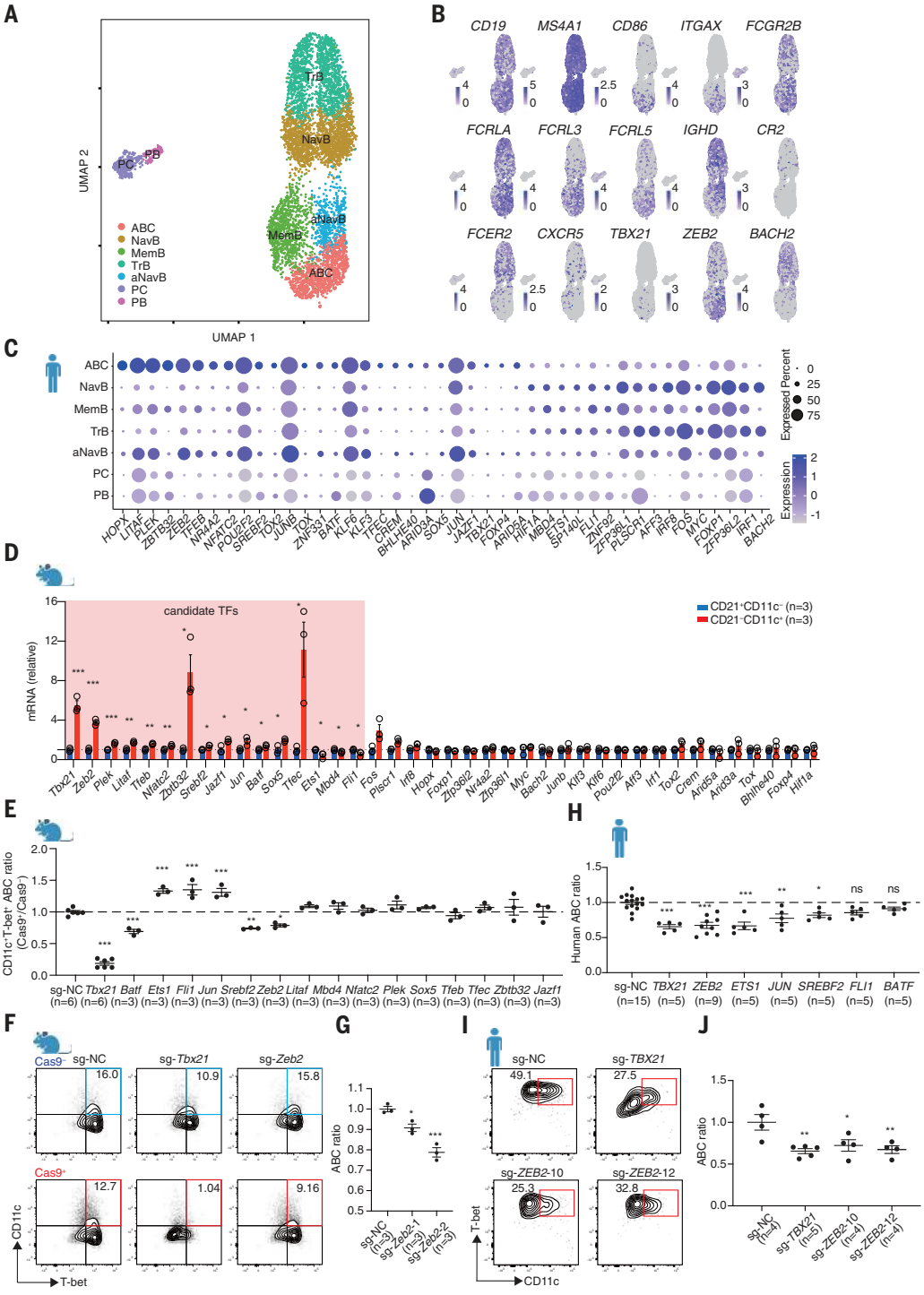


Fig. 1. CRISPR/Cas9–based screen of transcription factors for ABC differentiation. (**A**) scRNA-seq of CD19$^+$ peripheral B cells isolated from a patient with new-onset SLE. Seven clusters were defined as transitional B cells (TrB), naïve B cells (NavB), activated naïve B cells (aNavB), age-associated B cells (ABC), memory B cells (MemB), plasmablasts (PB), and plasma cells (PC). (**B**) UMAP plots displaying expression of select genes distinguishing ABCs. (**C**) Dot plots of 43 differentially expressed transcription factors (TFs) in ABCs. (**D**) Relative expression of mouse homolog genes [encoding equivalent TFs in (C)] in CD21$^+$CD11c$^-$ and CD21$^-$CD11c$^+$ B cells sorted from bm12-induced lupus mice. (**E**) Mouse ABC ratio in groups targeting indicated genes. The ratio defined by comparing Cas9$^+$CD11c$^+$T-bet$^+$ with Cas9$^-$CD11c$^+$T-bet$^+$ in the coculture was normalized with sg-NC (negative control). (**F**) Flow cytometry plots of ABCs (CD11c$^+$T-bet$^+$) derived from B cells transuded with sgRNA targeting *Tbx21* or *Zeb2*. (**G**) ABC ratio in groups with two distinct sgRNAs targeting *Zeb2*. (**H**) The human ABC ratio in groups targeting indicated genes. This ratio was defined by normalizing the frequency of CD27$^-$IgD$^-$CD11c$^+$T-bet$^+$ B cells with the sg-NC group. (**I** and **J**) Representative plots (I) and ABC ratio (J) from B cells electroporated with Cas9-gRNA (RNP) complex targeting *TBX21* and *ZEB2*. *n* represents distinct samples (biological repeats). Data are representative of three to four independent experiments. Data are means ± SEM. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$; ns, not significant; unpaired Student's *t* test for (D) and ordinary one-way analysis of variance (ANOVA) with Dunnett's multiple comparisons test for (E), (H), (G), and (J).

We examined peripheral blood mononuclear cells (PBMCs) from five unrelated MWS patients (tables S2 and S3) and identified five de novo heterozygous germline *ZEB2* variants through whole-exome sequencing (Fig. 2, A and B). *ZEB2* deficiency dramatically decreased ABC frequency (Fig. 2C and fig. S5A). Detailed B cell profiling revealed seven prominent B cell clusters (Fig. 2, D and E). In MWS patients, ABC frequency was notably reduced by 71% and was accompanied by a significant decline in activated naïve B cells and ABC progenitors (*5*) (Fig. 2F and fig. S5B). The switched memory B cells were decreased by 46%, and DN1 B cells—an alternative trajectory for effector B cells—were increased in MWS patients (fig. S5B) (*5*). These changes were confirmed by using manual gating with specific markers (Fig. 2G and fig. S5, C and D). We further studied ZEB2's regulatory effects on ABC formation by isolating B cells from MWS patients and inducing ABC differentiation in vitro. Corroborating our in vivo findings, in vitro ABC formation was also impaired (Fig. 2, H and I, and fig. S5, E and F). Thus, *ZEB2* loss-of-function variants confirm that ZEB2 is required for human ABC formation.

ZEB2 determines ABC pathogenicity in lupus

To further explore ZEB2's role in ABC formation, we generated mice selectively lacking *Zeb2* in B cells by crossing *Zeb2*-floxed mice with *Cd19*-cre mice (fig. S6, A to C). *Zeb2* deficiency in B cells reduced ABC differentiation by more than 50% (Fig. 3A). Even mice hemizygous for *Zeb2* in B cells (B-*Zeb2*^{Het}) displayed reduced ABC formation, providing evidence of *Zeb2* haploinsufficiency in mice. Moreover, overexpression of *Zeb2* in splenic B cells promoted ABC formation, indicating that *Zeb2* is sufficient for ABC differentiation in vitro (Fig. 3B).

ABCs comprise a distinctive effector B cell subset that arises during immune responses to nucleic acid–related antigens (*1*) and that develops separately from the germinal center (GC) pathway (*21*, *22*). To define how ZEB2 affects pathogenic ABCs, we investigated the consequences of *Zeb2* deficiency in B cells by using two lupus mouse models [lupus induced by the toll-like receptor 7 (TLR7) agonist imiquimod (IMQ) and bm12 cell transfer] as well as an acute lymphocytic choriomeningitis virus (LCMV) infection model. B cell–intrinsic *Zeb2* deficiency significantly impaired ABC formation in all three models (Fig. 3, C and D, and figs. S6, D to F, and S7, A to C). Furthermore, GC B cells increased in B cell–intrinsic *Zeb2*-deficient mice in the acute bm12-induced and LCMV infection models (fig. S7, D and E), suggesting a competitive relationship between GC B cells and ABCs. However, ZEB2 neither directly instructed GC B cell differentiation nor promoted antibody responses to ABC-irrelevant protein antigens (fig. S7, F and G). Detailed profiling of ABCs in IMQ-induced lupus confirmed that ABCs were phenotypically distinct from CD38$^-$GL-7$^+$ GC B cells (fig. S8A). A significant proportion of CD19^{hi}CD11c$^+$IgD$^-$ ABCs exhibited CD38$^+$GL-7$^-$ memory markers, while also displaying a distinct hyperactivated state (CD95$^+$CD80$^+$) in comparison with other memory B cells (fig. S8A). *Zeb2* deficiency selectively affected the distribution and hyperactivation of CD11c$^+$ memory-like B cells, without affecting the frequency of the CD11c$^-$ memory B cell subset (fig. S8B). A subpopulation of ABCs acquired a phenotype (CD38$^+$GL-7$^+$) consistent with precursors of GC (pre-GC) B cells (fig. S8C), suggesting that like conventional memory B cells, ABCs could contribute to secondary GCs, seeding a chronic GC response. ABCs reside at the preplasma cell stage and can quickly differentiate into plasma cells (*1*). Such chronic GC responses and terminally differentiated plasma cells were reduced in IMQ-induced B-*Zeb2*^{KO} mice, likely because of reduced replenishment from ABCs (fig. S8, D to F). Thus, rather than broadly promoting all effector B cell responses, ZEB2 selectively drives ABCs and their progeny. Moreover, although these cells develop extrafollicularly, their progeny may participate in GC responses in the context of chronic inflammation.

TLR7-driven lupus is GC-independent and mostly ABC-dominant (*21*). We therefore examined whether ZEB2 regulated ABC-mediated autoimmunity in lupus induced by IMQ, a TLR7 agonist. *Zeb2* deficiency in B cells significantly ameliorated splenomegaly (Fig. 3E) and decreased serum antinuclear (ANA) and double-stranded DNA (dsDNA) autoantibodies (Fig. 3, F and G). ABCs are particularly pathogenic owing to secretion of antibodies of the IgG2a/c isotype (*1*). Compared with non-ABCs, ABCs secreted the highest levels of IgG2c isotype antibodies upon restimulation. By contrast, residual ABC-like cells isolated from B-*Zeb2*^{KO} mice were unable to produce comparable amounts of IgG2c antibodies (Fig. 3H). Similarly, B-*Zeb2*^{KO} mice treated with IMQ produced much fewer IgG2c autoantibodies (Fig. 3I). In lupus nephritis (LN), ABCs correlate with tissue damage (*21*, *23*) and are known to produce proinflammatory cytokines and chemokines, such as CCL5, CXCL10, interferon-γ (IFN-γ), and interleukin-6 (IL-6) (*8*). In IMQ-treated B-*Zeb2*^{KO} mice, kidney-infiltrating ABCs were significantly decreased (Fig. 3J and fig. S8G) as was tissue damage (Fig. 3K). The residual CD11c$^+$CD21$^-$ B cells from B-*Zeb2*^{KO} mice also produced reduced quantities of CCL5 and CXCL10 ex vivo (Fig. 3L). Thus, ZEB2 is essential for ABC-mediated autoimmunity and the proinflammatory properties of ABCs.

ZEB2 controls the lineage specification and cellular identity of ABCs

To investigate the consequences of ZEB2 regulation of gene transcription, RNA sequencing was performed on sorted *Zeb2*-deficient B cells after in vitro ABC induction. ABC signature genes including *Itgax*, *Itgam*, *Itgb2*, *Nkg7*, *Tbx21*, *Zbtb32*, and *Fcer2a* (*4*, *5*, *8*, *24*–*26*) exhibited inverse expression patterns after *Zeb2* deficiency (Fig. 4, A and B). Gene set enrichment analysis (GSEA) revealed that *Zeb2*-deficient B cells lacked expression of the "ABC up-regulated" gene set, whereas it was enriched in the "ABC down-regulated" gene set from the public dataset GSE99480 (*8*) (Fig. 4, C and D).

To elucidate ZEB2's direct targets in ABCs, we performed high-throughput sequencing of the regulome by using assay for transposase-accessible chromatin sequencing (ATAC-seq), CUT & Tag, and CUT & RUN, leading to the identification of 4338 genes annotated by 6733 accessible sites with ZEB2 binding. Among the genes differentially expressed in *Zeb2*-deficient cells, we found 33 candidate direct targets: 22 repressed and 11 activated by ZEB2 (fig. S9A). Myocyte enhancer factor 2b (*Mef2b*), an essential TF for GC development (*27*), was repressed by ZEB2. This direct regulation was mapped to a conserved region ~20 kb downstream of *Mef2b*'s exon 1 TSS, enriched with enhancer-associated features in both human and mouse (Fig. 4E and fig. S9, B to D). We validated ZEB2 suppression of *Mef2b* expression in *Zeb2*-deficient and *Zeb2*-overxpressing B cells and confirmed opposing expression patterns of *Zeb2* and *Mef2b* in ABCs and GC B cells from public datasets (*22*) (fig. S9, E to H). MEF2B can directly regulate *S1pr2* (*27*), and *Zeb2*-deficient B cells up-regulated *S1pr2* expression (fig. S9E). Thus, ZEB2 appears to foster ABC differentiation by directly repressing *Mef2b* to constrain GC B cells, which is in alignment with our observations of the bm12 and LCMV models (fig. S7, D and E).

Additionally, ZEB2-specific peaks from CUT & Tag were matched to motifs of GATA3, FOSL2, and ZEB2 (fig. S10, A to C), which is consistent with existing ZEB2 chromatin immunoprecipitation sequencing (ChIP-seq) data (fig. S10D). We identified a ZEB2-specific peak residing in the promoter of *Itgax*, containing a ZEB2-binding sequence (Fig. 4F). *Zeb2* deficiency altered the chromatin accessibility of the ABC-specific opening in the *Itgax* promoter, further confirming that ZEB2 controls transcription of ABC signature genes. CD11c (*Itgax*), an important α-subunit member of β2 integrins, can pair with the β subunit CD18 to form heterodimeric cell-surface receptors important for immune-cell adhesion and recruitment to tissues (*28*), which is a superior property of ABCs among B cells (*4*, *21*, *23*). In kidney biopsies from patients with LN (table S4), CD11c$^+$ B cells were found in affected tissues, constituting approximately 50% of total B cells, with IgD$^-$CD27$^-$CD11c$^+$ ABCs comprising about 20% (fig. S11, A and B). We validated the enhanced migratory capacity of ABCs in vitro, which was modulated by CD11c blockade and *Zeb2* deficiency (fig. S11, C to G). Thus, ZEB2 plays a crucial role

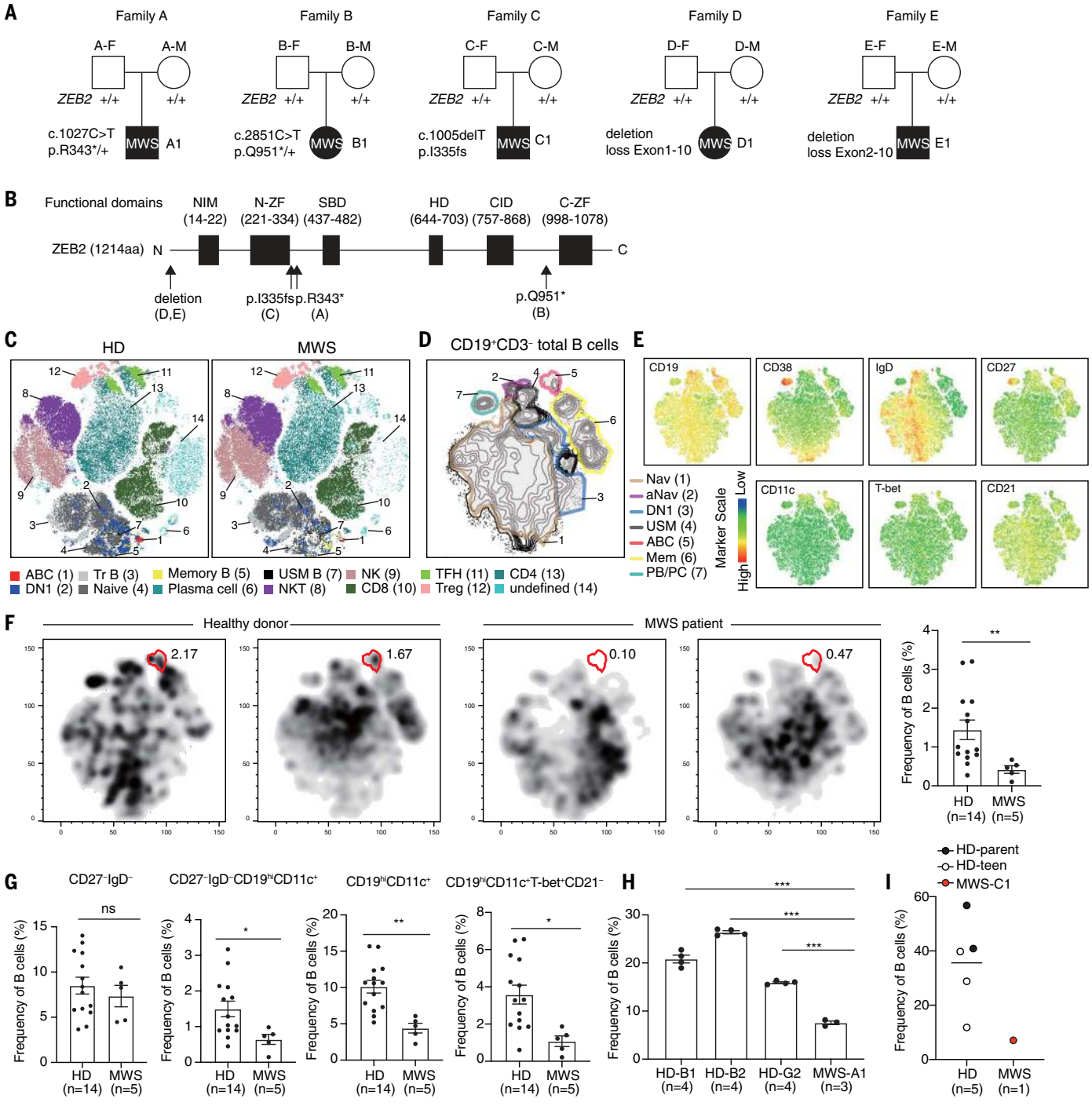


Fig. 2. *ZEB2* is required for human ABC formation. (**A**) Family pedigrees showing de novo heterozygous mutations of *ZEB2* in Mowat-Wilson syndrome (MWS) patients. (**B**) Schematic of the general linear structure of the functional domain composition of ZEB2 protein. The black arrows show the location of the ZEB2 mutation described in (A). (**C**) T-distributed stochastic neighbor embedding (t-SNE) plots of lymphocyte clusters in PBMCs of a healthy donor (HD) and a MWS patient by flow cytometry. (**D**) t-SNE plots of peripheral B cell clusters for MWS patients and healthy donors. Seven B cell clusters were identified based on lineage marker expression as naïve B cells (Nav), activated naïve B cells (aNav), CXCR5$^+$ double-negative B cells (DN1), unswitched memory B cells (USM), age-associated B cells (ABC), memory B cells (Mem), plasmablasts (PB), and plasma cells (PC). (**E**) t-SNE plots of peripheral B cells displaying CD11c, CD19, CD21, CD27, CD38, IgD, and T-bet expression. (**F**) Representative t-SNE plots and frequency of ABCs (red gate) in peripheral B cells from MWS patient and HDs. (**G**) Frequency of CD27$^-$IgD$^-$, CD27$^-$IgD$^-$CD19^{hi}CD11c$^+$, CD19^{hi}CD11c$^+$, and CD19^{hi}CD11c$^+$T-bet$^+$CD21$^-$ in PBMCs from MWS patient and HDs analyzed by manual gating. (**H** and **I**) Representative plots and frequency of in vitro–induced ABCs derived from B cells of MWS-A1 (H), MWS-C1 (I), and HDs. *n* represents distinct samples [biological repeats, except (H)]. Data are means ± SEM. **P* < 0.05, ***P* < 0.01, ****P* < 0.001; ns, not significant; unpaired Student's *t* test for (G) and (H) with Welch's correction for (F).

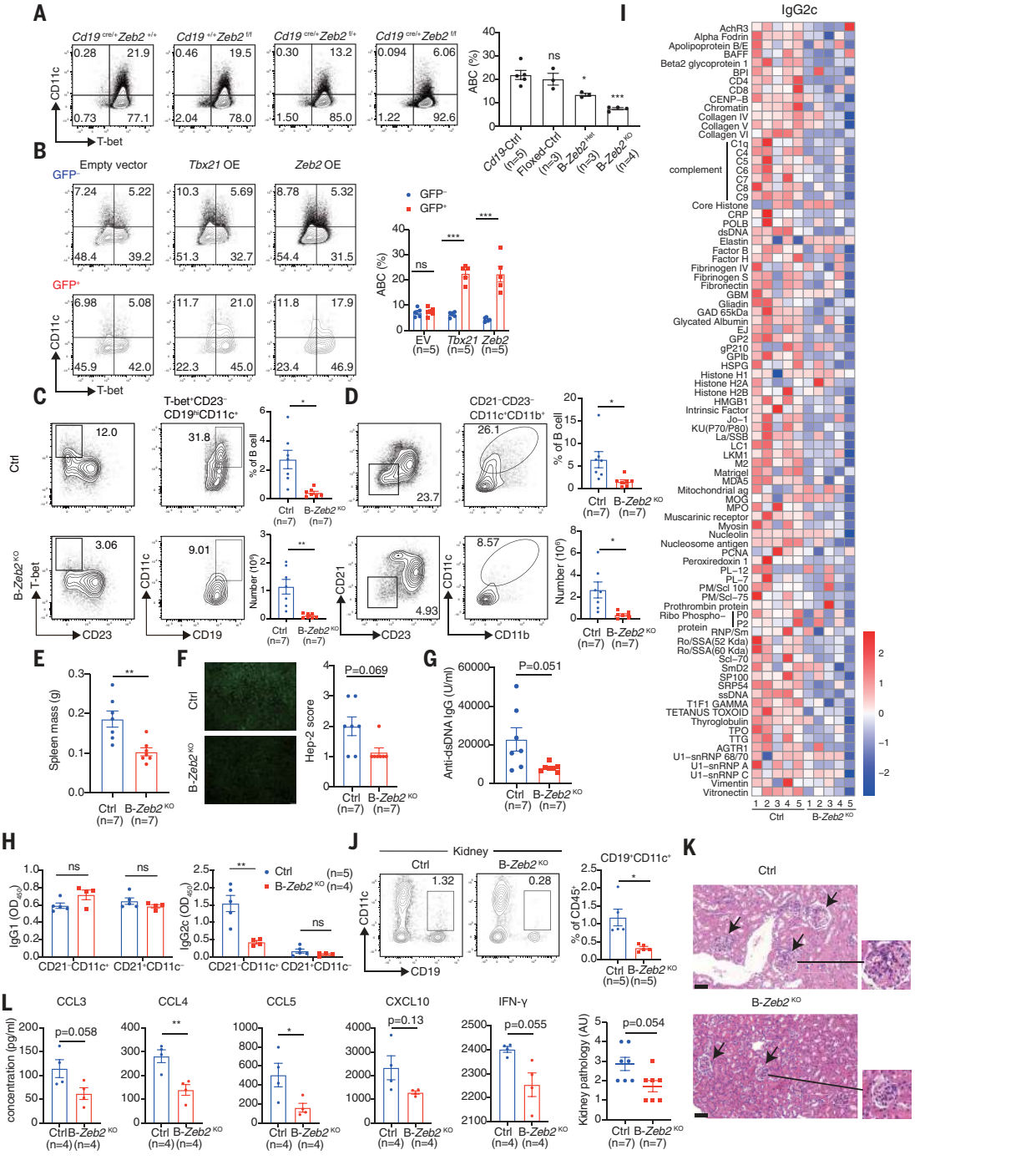


Fig. 3. *Zeb2* deficiency impairs ABC formation and alleviates lupus pathogenesis. (**A**) Representative plots and frequency of in vitro–induced ABCs ($CD19^+CD11c^+T\text{-}bet^+$) derived from splenic B cells of $Zeb2^{+/+}Cd19^{Cre/+}$ (*Cd19*-Ctrl), $Zeb2^{f/f}Cd19^{+/+}$ (Floxed-Ctrl), $Zeb2^{f/+}Cd19^{cre/+}$(B-$Zeb2^{Het}$), and $Zeb2^{f/f}Cd19^{cre/+}$(B-$Zeb2^{KO}$) mice. (**B**) Representative plots and frequency of ABCs ($CD19^+CD11c^+T\text{-}bet^+$) in GFP^+ (infected) and GFP^- (uninfected) B cells transduced with empty plasmid, *Tbx21*, or *Zeb2* cDNA sequence. GFP, green fluorescent protein; EV, empty vector; OE, overexpression. (**C** and **D**) Representative plots, frequency, and absolute number of splenic ABCs identified by $T\text{-}bet^+CD23^-CD19^{hi}CD11c^+$ (C) and $CD21^-CD23^-CD11c^+CD11b^+$ (D) in IMQ-induced B-$Zeb2^{KO}$ and $Cd19^{Cre/+}$ (Ctrl) mice. (**E** to **G**) Spleen mass (E), ANA (F), and anti-dsDNA (G) in serum from mice described in (C) and (D). Scale bars in (F), 100 μm. (**H**) The IgG1 and IgG2c antibody titers in the culture supernatants from $CD21^-CD11c^+$ and $CD21^+CD11c^-$ B cells sorted from mice described in (C) and (D). (**I**) Autoantigen microarray showing the relative IgG2c-isotype autoantibody levels in the serum of mice described in (C) and (D). (**J**) Representative plots and frequency of renal ABCs ($CD19^+CD11c^+$) from mice described in (C) and (D). (**K**) (Right) Hematoxylin and eosin (H&E) staining and (left) pathology assessment of kidneys from mice described in (C) and (D). Scale bars, 50 μm. (**L**) The concentration of cytokine and chemokine in the culture supernatants described in (H). *n* represents distinct samples (biological repeats). Data are representative of two to three independent experiments. Data are means ± SEM. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$; ns, not significant; unpaired Student's *t* test [(B), (E), (H), and (L) for CCL3, CCL4, and CCL5] with Welch's correction [(C), (D), (G), (J), and (L) for CXCL10 and IFN-γ]; Mann-Whitney *U* test [(F) and (K)]; and ordinary one-way ANOVA with Dunnett's multiple comparisons test (A).

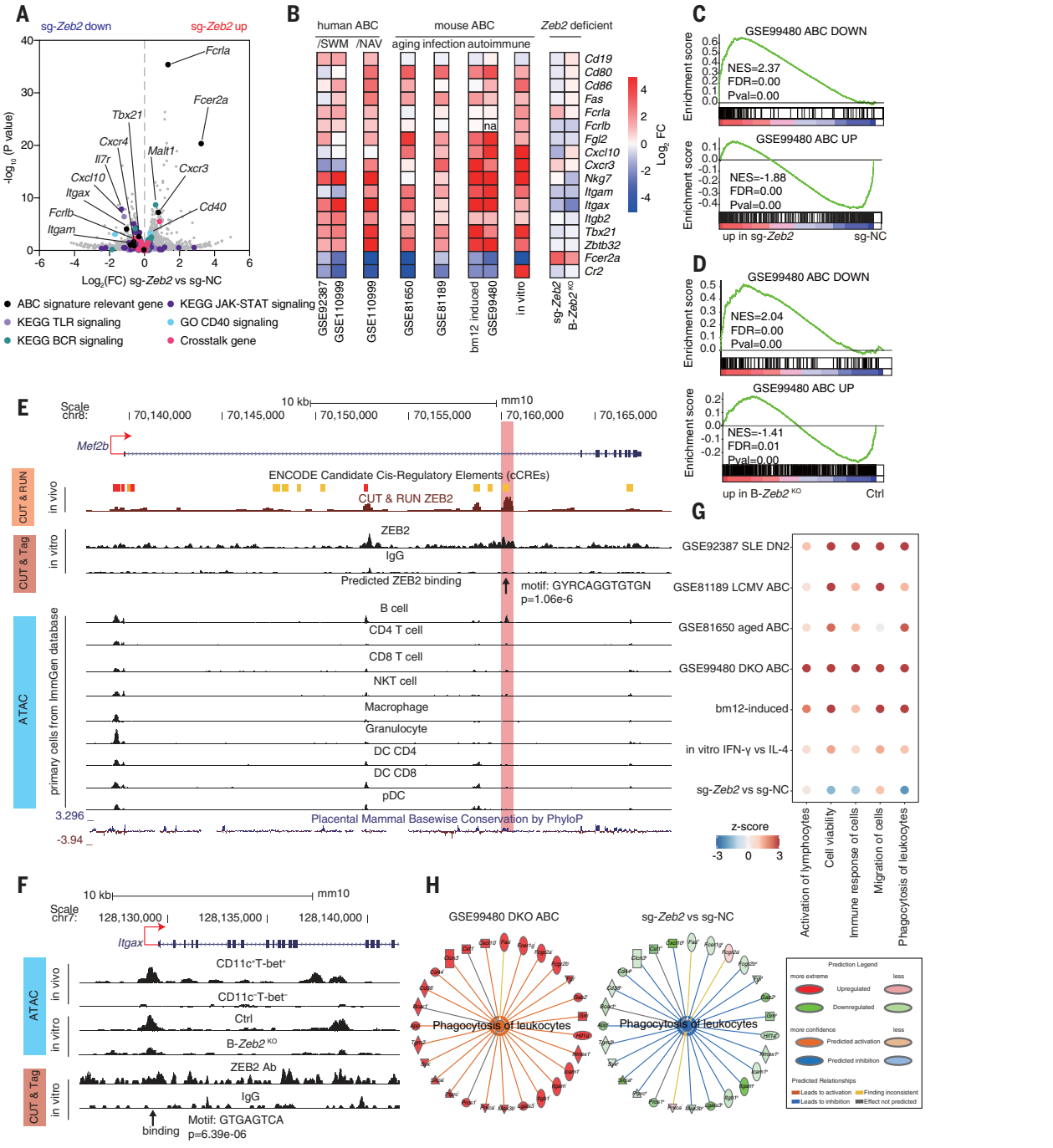


Fig. 4. *Zeb2* regulates specification and cellular identity of ABCs. (**A**) Volcano graph showing transcriptional profiles in *Zeb2*-edited B cells. ABC-signature genes; genes associated with BCR, TLR, JAK-STAT, and CD40 signaling; and cross-talk genes were labeled with colored dots. GO, Gene Ontology. (**B**) Heatmap showing expression of representative ABC-signature genes in RNA-seq of public datasets GSE92387, GSE110999, GSE81650, and GSE81189, and in RNA-seq of ABCs sorted from bm12-induced lupus mice (bm12-induced), public dataset GSE99480, ABC-polarized B cells (in vitro), ABC-polarized B cells derived from *Zeb2*-edited B cells (sg-*Zeb2*), and *Zeb2*-knockout B cells (B-*Zeb2*^{KO}). (**C** and **D**) GSEA showing the enrichment of the ABC-down gene set and ABC-up gene set from GSE99480 in ABC-polarized B cells derived from *Zeb2*-deficient B cells. *Zeb2* deficiency was mediated through sg-*Zeb2* editing (C) or deletion (D). NES, normalized enrichment score; FDR, false discovery rate. (**E**) CUT & RUN, CUT & Tag, and ATAC-seq tracks display ZEB2 binding around the *Mef2b* locus, visualized by the University of California, Santa Cruz (UCSC) genome browser. The chromatin accessibility in mouse primary immune cell subsets is from the public ImmGen database. DC, dendritic cell; pDC, plasmacytoid dendritic cell; PhyloP, phylogenetic *P* value. (**F**) ATAC and CUT & Tag tracks display ZEB2 binding around the *Itgax* locus in ex vivo–sorted ABCs and ABC-polarized B cells. (**G**) Dot plot showing the activation z-score of predicted biological function in RNA-seq datasets mainly described as in (B). DKO, double knockout. (**H**) Network diagram representing phagocytosis pathway in ABC dataset (GSE99480) and sg-*Zeb2* versus sg-NC dataset by IPA. The color of each node and the sign attached to each gene symbol indicate change in the gene expression: up-regulated (red, asterisk) and down-regulated (green, hash). The connecting lines indicate the predicted relationship between nodes and biological function: Orange represents activation, blue represents inhibition, and gray represents effect not predicted.

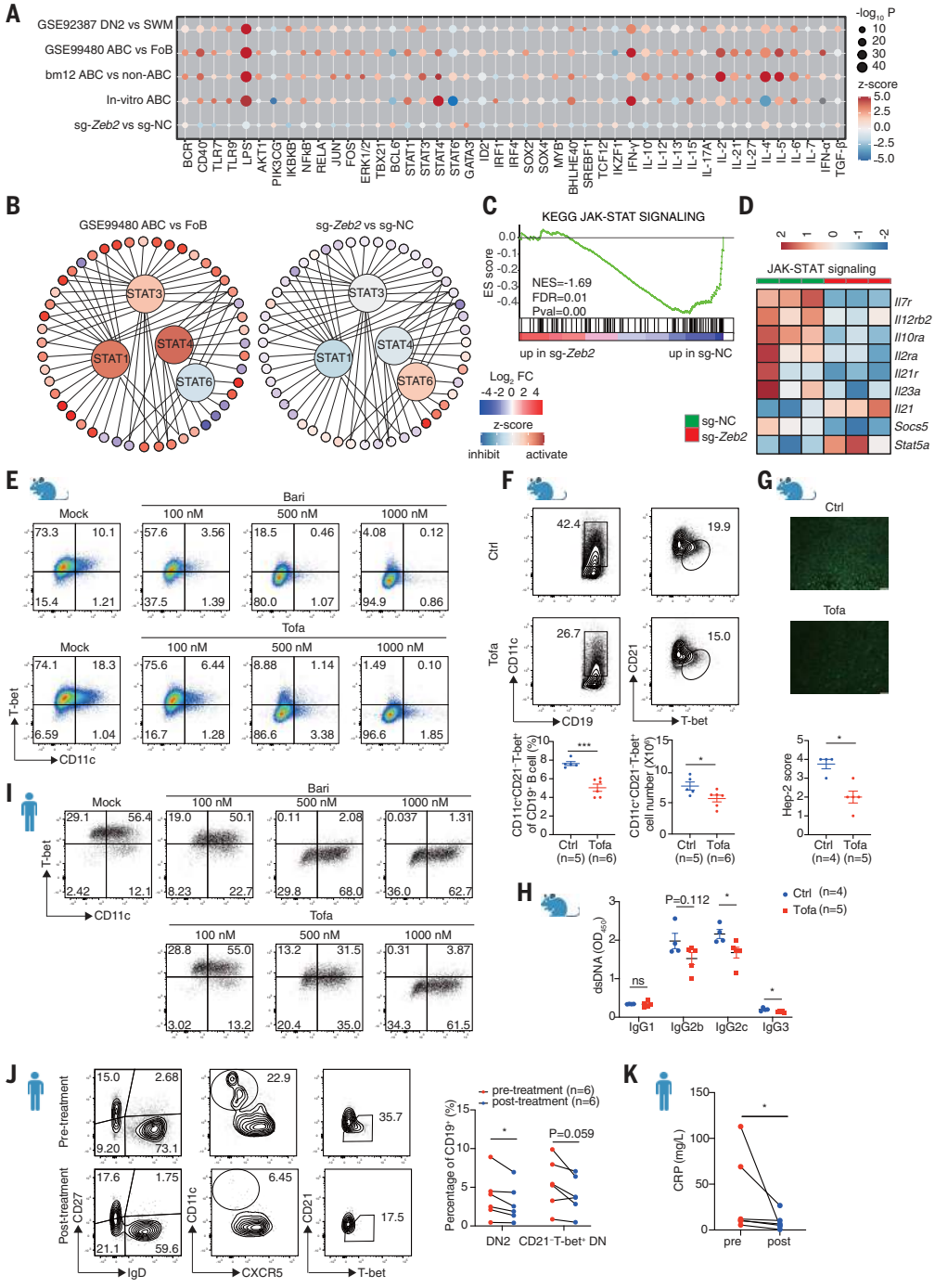


Fig. 5. ZEB2-JAK-STAT axis controls ABC differentiation. (**A**) Activation *z*-score heatmap for IPA-predicted upstream regulators in indicated datasets mainly described in Fig. 4B. *P* values are log-normalized and presented by the size of the plot. The color of the plot indicates the activated (red) or inhibited (blue) regulation of the predicted regulators. (**B**) Upstream regulator analysis showing STAT1, STAT3, and STAT4 activated and STAT6 inhibited in ABCs (GSE99480), whereas the opposite effects were observed in the sg-*Zeb2* versus sg-NC dataset. The color of the surrounding circles indicates the change in the expression: red (up-regulated) and blue (down-regulated). The color of the central circles indicates predicted regulators: orange (activated) and blue (inhibited). (**C**) GSEA showing impaired JAK-STAT pathway in *Zeb2*-edited B cells. (**D**) Heatmap showing expression of selected JAK-STAT signaling genes in sg-NC^{+} and sg-*Zeb2*$^{+}$ B cells. (**E**) Flow cytometry plots of in vitro–induced mouse ABCs with different concentrations of baricitinib (Bari) or tofacitinib (Tofa). (**F** to **H**) Frequency and absolute number of ABCs (F), ANA titer (G), and anti-dsDNA Ig titers (H) in serum from bm12-induced lupus mice treated with tofacitinib. Scale bars in (G), 100 μm. (**I**) Flow cytometry plots of in vitro–induced human ABCs with different concentrations of baricitinib or tofacitinib. (**J**) Flow cytometry plots of human ABCs in PBMCs from patients with rheumatoid arthritis before and after JAK-STAT inhibitor tofacitinib treatment for 4 weeks. (**K**) The change of complement-reactive protein (CRP) level in RA patients described in (J). *n* represents distinct samples (biological repeats). Data are representative of two to three independent experiments [(E) through (I)]. Data are means ± SEM. **P* < 0.05, ***P* < 0.01, ****P* < 0.001; ns, not significant; unpaired Student's *t* test [(F) and (H)]; Mann-Whitney *U* test (G); and paired Student's *t* test [(J) and (K)].

in orchestrating ABC specification by directly suppressing other effector B cell subsets and inducing the ABC signature.

ZEB2 drives distinct functional properties of ABCs

To better define the function of ABCs, we applied ingenuity pathway analysis (IPA) to a public dataset (GSE99480) (*8*). Among the top 35 significantly increased predicted functions (table S5), ABCs shared five features: "enhanced viability," "migration," "activation," "immune response," and "phagocytosis/engulfment" (fig. S12A). These were validated across several transcriptomes (Fig. 4G). Selected transcripts linked to these biological functions formed a network (fig. S12A). Pathway analysis further supported our finding that ABC-enriched pathways were linked to these five functional features (fig. S12B). ABCs have also been characterized by a hyperactivation state, long-term survival, and a distinct migration/distribution pattern in published studies (*4*, *5*, *29*). ABCs also exhibited a characteristic phagocytic capacity, identified by enriched phagosome formation and Fc-receptor pathways (fig. S12B). The ability of ABCs to both perform typical B cell functions and co-opt myeloid markers such as CD11c—as well as cytotoxic molecules such as NKG7, granzyme A, and perforin—has been previously described (*3*, *4*). In line with ZEB2's critical role in ABC function, *Zeb2* editing in B cells dampened their viability, immune response, and phagocytosis/engulfment (Fig. 4, G and H).

To experimentally test the phagocytic capacity of ABCs, we incubated splenic B cells with apoptotic thymocytes labeled with pHrodo and monitored apoptotic cell internalization (fig. S12C). $CD19^{hi}CD11c^{+}$ B cells exhibited markedly enhanced uptake evidenced by an increased $pHrodo^{+}$ fraction and signal intensity (fig. S12D). In vitro–generated ABCs were also able to engulf apoptotic cells, and this capacity was dampened by *Zeb2* deficiency (fig. S12, E and F). Thus ABCs exhibit distinct biological functions that are regulated by ZEB2.

The ZEB2-JAK-STAT axis governs ABC differentiation

To elucidate the signaling pathways by which ZEB2 influences ABC formation, we performed upstream regulator analysis (URA) in IPA on both public and our own datasets. As anticipated, BCR, CD40, TLR, and key downstream pathways such as NF-κB and AKT, were predicted to be activated in ABCs (Fig. 5A). Regulatory effects of cytokines IFN-γ, IL-10, and IL-21, along with their JAK-STAT (Janus kinase-signal transducer and activator of transcription) signals, were also detected (Fig. 5, A and B). Specifically, STAT1, STAT3, and STAT4 were activated, whereas STAT6 was inhibited (Fig. 5, A and B), which is consistent with previous findings (*1*). *Zeb2* deficiency altered STAT signals (Fig. 5, A and B), reflecting opposite expression pattern of STAT target genes in *Zeb2*-edited B cells (Fig. 5B). GSEA and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis produced similar findings, providing support for an important role of JAK-STAT signaling in the function of ZEB2 on ABCs (Fig. 5, C and D, and fig. S13A). We also confirmed that gene expression altered by *Zeb2* deficiency largely overlapped with the transcriptional program affected by inhibition of JAK-STAT signaling (fig. S13, B and C).

JAK-STAT inhibitors, such as baricitinib and tofacitinib, have proven effective in dampening intracellular cytokine signaling (*30*, *31*). We therefore tested their effects on mouse ABC formation and found that they impaired in vitro ABC differentiation in a dose-dependent manner (Fig. 5E and fig. S13, D to G). Tofacitinib administration reduced ABC accumulation and splenomegaly, lowered autoantibody titers, and decreased ABC-relevant cytokines in a manner likely to be B cell–intrinsic (Fig. 5, F to H, and fig. S14, A to E). Human ABC differentiation was also inhibited by these drugs (Fig. 5I and fig. S14F). Furthermore, tofacitinib treatment decreased circulating ABCs in rheumatoid arthritis (RA) patients (Fig. 5J and fig. S14G) and ameliorated systemic inflammation (Fig. 5K). Thus, targeting the JAK-STAT pathway can block ABC differentiation in both mice and humans, making it a promising strategy for the treatment of ABC-mediated autoimmunity.

Discussion

We have identified ZEB2 as an essential TF that drives human and mouse ABC differentiation, antinuclear antibody formation, proinflammatory cytokine and chemokine production, and ABC migration to inflamed tissues. ZEB2 drives the ABC gene signature, including *Itgax*, and suppresses *Mef2b*, which causes activated B cells to deviate from GCs and differentiate extrafollicularly. Differentiation of ABCs in a GC-independent manner has raised questions about the role of GCs in autoimmunity (*21*, *22*). GC reactions comprise several tolerance checkpoints that are lacking in ABC development. Although ZEB2 appears to be essential for ABC formation, the upstream physiological signals and cells that turn on *Zeb2* expression in vivo remain unclear. ZEB2 is likely to act in concert with other transcription or epigenetic factors, including IRF5, T-bet, and metabolic regulators, to shape a regulatory complex in ABCs, mirroring ZEB2's regulatory programs in natural killer (NK) cells (*32*) and CD8 T cells (*33*).

Our study highlights the innate ability of ABCs to phagocytose apoptotic cells, a function that may underpin TLR7 activation and self-antigen presentation to T cells, as well as explaining their hyperactivated status. The requirement of the JAK-STAT pathway to exert *Zeb2*-mediated ABC development and pathogenicity offers promising therapeutic prospects through JAK inhibitors. These insights extend to conditions in which ABCs are expanded and may exert pathogenic roles such as aging.

REFERENCES AND NOTES

1. M. P. Cancro, *Annu. Rev. Immunol.* **38**, 315–340 (2020).
2. Y. Hao, P. O'Neill, M. S. Naradikian, J. L. Scholz, M. P. Cancro, *Blood* **118**, 1294–1304 (2011).
3. A. V. Rubtsov *et al.*, *Blood* **118**, 1305–1315 (2011).
4. S. Wang *et al.*, *Nat. Commun.* **9**, 1758 (2018).
5. S. A. Jenks *et al.*, *Immunity* **49**, 725–739.e6 (2018).
6. E. Zumaquero *et al.*, *eLife* **8**, e41641 (2019).
7. S. L. Stone *et al.*, *Immunity* **50**, 1172–1187.e7 (2019).
8. M. Manni *et al.*, *Nat. Immunol.* **19**, 407–419 (2018).
9. E. Ricker *et al.*, *Nat. Commun.* **12**, 4813 (2021).
10. S. W. Du, T. Arkatkar, H. M. Jacobs, D. J. Rawlings, S. W. Jackson, *Eur. J. Immunol.* **49**, 170–178 (2019).
11. R. C. Levack, K. L. Newell, M. Popescu, B. Cabrera-Martinez, G. M. Winslow, *J. Immunol.* **205**, 1050–1058 (2020).
12. C. Lien *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4664–4668 (2010).
13. H. Xu *et al.*, *Nat. Immunol.* **16**, 1274–1281 (2015).
14. S. De *et al.*, *Front. Immunol.* **8**, 1938 (2018).
15. D. R. Glass *et al.*, *Immunity* **53**, 217–232.e5 (2020).
16. M. S. Naradikian *et al.*, *J. Immunol.* **197**, 1023–1028 (2016).
17. C. Vandewalle *et al.*, *Nucleic Acids Res.* **33**, 6566–6578 (2005).
18. D. R. Mowat *et al.*, *J. Med. Genet.* **35**, 617–623 (1998).
19. N. Wakamatsu *et al.*, *Nat. Genet.* **27**, 369–370 (2001).
20. C. L. Scott, K. D. Omilusik, *Trends Immunol.* **40**, 431–446 (2019).
21. G. J. Brown *et al.*, *Nature* **605**, 349–356 (2022).
22. W. Song *et al.*, *Immunity* **55**, 290–307.e5 (2022).
23. C. Wu *et al.*, *Ann. Rheum. Dis.* **78**, 1090–1100 (2019).
24. L. M. Russell Knode *et al.*, *J. Immunol.* **198**, 1921–1927 (2017).
25. B. E. Barnett *et al.*, *J. Immunol.* **197**, 1017–1022 (2016).
26. X. Han *et al.*, *Arthritis Rheumatol.* **75**, 1203–1215 (2023).
27. P. Brescia *et al.*, *Cancer Cell* **34**, 453–465.e9 (2018).
28. Y. Zhang, H. Wang, *Immunology* **135**, 268–275 (2012).
29. J. L. Johnson *et al.*, *Immunity* **52**, 842–855.e6 (2020).
30. J. J. O'Shea, A. Kontzias, K. Yamaoka, Y. Tanaka, A. Laurence, *Ann. Rheum. Dis.* **72**, ii111–ii115 (2013).
31. P. G. Traves *et al.*, *Ann. Rheum. Dis.* **80**, 865–875 (2021).
32. M. J. van Helden *et al.*, *J. Exp. Med.* **212**, 2015–2025 (2015).
33. C. X. Dominguez *et al.*, *J. Exp. Med.* **212**, 2041–2056 (2015).

ACKNOWLEDGMENTS

We thank Z. Liu, Y. Ma, Q. Hu, Y. Hu, Y. Chen, and S. Zhou for providing experimental help; J. Li and X. Xu for provding the LCMV Amstrong virus; X. Song for suggestions; J. Qin and J. Huang for sample collection; and Y. Yu for reagents support. We acknowledge BioRender.com for providing tools to create schematic illustration. **Funding:** This research was supported by the National Natural Science Foundation of China (31630021, 31930037, 82071843, and 81901637); the core grant from the Francis Crick Institute (C2228) to C.G.V., the National Human Genetic Resources Sharing Service Platform (2005DKA21300); the Shanghai Municipal Key Medical Center Construction Project (2017ZZ01024-002); the Shenzhen Science and Technology Project (JCYJ20180504170414637 and JCYJ20180302145033769); the Shenzhen Futian Public Welfare Scientific Research Project (FTWS2018005); and the Sanming Project of Medicine in Shenzhen (SZSM201602087). **Author contributions:** Conceptualization: N.S., D.D., and C.G.V. Methodology: D.D. and S.G. Investigation: D.D., S.G., X.H., H.D., Y.J., X.Z., C.Y., S.H., J.Z., G.H., B.Q., H.Z., Y.Q., Y.H., J.M., Z.Yin, Z.Ye, J.Q., Q.J., L.W., Q.G., S.C., C.H., L.C.K., and M.T.W. Visualization: D.D., S.G., and Y.J. Funding acquisition: N.S., C.G.V., D.D., Z.Ye, and S.C. Supervision: N.S. and C.G.V. Writing – original draft: D.D. and G.S.S. Writing – review and editing: N.S. and C.G.V. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The

scRNA-seq, RNA-seq, ATAC-seq, CUT & Tag, and CUT & RUN sequencing data are deposited in GEO under accession numbers GSE242615, GSE242607, and GSE242611. Plasmids are available upon establishment of an MTA with Shanghai Jiaotong University.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf8531
Materials and Methods
Figs. S1 to S14
Tables S1 to S9
References (*34*–*51*)
MDAR Reproducibility Checklist

Submitted 17 November 2022; resubmitted 19 April 2023
Accepted 15 December 2023
10.1126/science.adf8531

SYNTHETIC BIOLOGY

Establishing a synthetic orthogonal replication system enables accelerated evolution in *E. coli*

Rongzhen Tian*, Fabian B. H. Rehm, Dariusz Czernecki, Yangqi Gu, Jérôme F. Zürcher, Kim C. Liu, Jason W. Chin*

The evolution of new function in living organisms is slow and fundamentally limited by their critical mutation rate. Here, we established a stable orthogonal replication system in *Escherichia coli.* The orthogonal replicon can carry diverse cargos of at least 16.5 kilobases and is not copied by host polymerases but is selectively copied by an orthogonal DNA polymerase (O-DNAP), which does not copy the genome. We designed mutant O-DNAPs that selectively increase the mutation rate of the orthogonal replicon by two to four orders of magnitude. We demonstrate the utility of our system for accelerated continuous evolution by evolving a 150-fold increase in resistance to tigecycline in 12 days. And, starting from a GFP variant, we evolved a 1000-fold increase in cellular fluorescence in 5 days.

The evolution of new function in living organisms is the result of continuous genomic mutation and selection within a population. This process is slow, and the rate of evolution is fundamentally limited by the critical mutation rate (*1*). Directed evolution commonly sidesteps the limitation on in vivo mutation rate by generating genetic diversity in vitro (*2*), but this does not enable the continuous evolution of genes within an organism. The mutation rate of cells can be transiently increased, but high levels of untargeted mutation lead to a catastrophic mutational load on the genome and are unsustainable. Genes inserted in viral genomes can be mutated by iteratively infecting new mutagenic cells (*3*–*6*). This approach sidesteps the challenge of increasing the rate of sustained mutation on genes in cells and can be extended to select for some phenotypes (*7*). However, this strategy is limited to evolving genes that are small enough to be packaged into viruses and to selecting phenotypes that can be coupled to infectivity; furthermore, selection occurs in cells under conditions of replicative stress, which may further limit the cellular phenotypes that can be explored.

Strategies that direct mutations to specific, targeted DNA sequences within cells without substantially increasing the genomic mutation rate offer the possibility of driving accelerated, sustainable, continuous, cellular evolution of target sequences (*8*–*17*). Pioneering work has taken advantage of an existing natural linear plasmid that functions in the yeast cytosol and is replicated by a dedicated, protein-primed DNA polymerase that does not copy the yeast genome as a natural orthogonal replication system (*12*, *13*). By recombining target genes into this existing linear plasmid system in yeast and generating mutagenic orthogonal DNA polymerases, a continuous evolution system was developed in this host. This system has been used to evolve metabolic pathways and antibodies and provided key insights into evolutionary trajectories (*12*, *13*, *18*). However, the system cannot be used for engineering bacterial genetic elements and requires additional steps to engineer the established replicons in vivo. Moreover, the doubling time of yeast makes this system theoretically slower than bacterial systems. Recent work has shown that target genes can also be recombined into a natural linear plasmid in *Bacillus thuringiensis*, and this system can also be used to generate a mutagenic orthogonal replication system (*14*). However, there are very limited genetic tools in this organism and the host is not widely used or well characterized.

Escherichia coli is the workhorse of molecular biology and is widely used in both fundamental discovery science and industrial production (*19*). It is the best characterized organism, and many of its biochemical pathways have been characterized in detail. It has a rapid doubling time (~20 min) and is a preferred host for gene cloning and protein expression, and a vast repertoire of genetic tools have been developed for this organism over many years. An outstanding challenge over the past decade has been to discover a stable orthogonal replication system that operates in *E. coli* and thereby enables accelerated continuous evolution in this host. However, despite substantial effort, no stable orthogonal replication system has been discovered in *E. coli*.

Results

Establishing a synthetic linear DNA replication system in vivo

PRD1 is a lytic phage that infects *E. coli*, undergoes uncontrolled replication, and lyses cells in 60 min (Fig. 1A and fig. S1) (*20*). Its linear, double-stranded genome encodes at least 25 gene products from five annotated operons under the control of eight annotated promoters and terminators (*21*). The ends of the linear genome are composed of inverted terminal repeats (ITRs) that form the binding site for the terminal protein (TP) and function as origins of replication (Fig. 1B) (*22*). The early operons contain the genes responsible for replication of the PRD1 genome. The left early operon encodes the TP and the DNA polymerase (DNAP), and the right early operon encodes phage single-stranded DNA binding protein (SSB) and double-stranded DNA-binding protein (DSB). The central operons contain the genes encoding the remaining structural and lytic protein components of PRD1.

To generate a synthetic system for the controlled replication of a linear replicon (Fig. 1A), we separated the four genes that we hypothesized might be essential for in vivo replication of the PRD1 genome from the structural and lytic genes and combined them into a single synthetic replication operon controlled by an isopropyl-β-D-thiogalactopyranoside (IPTG)–inducible promoter, PtacIPTG (Fig. 1B). We hypothesized that this synthetic replication operon might be sufficient to direct the replication of any linear double-stranded DNA flanked by PRD1 ITR sequences in *E. coli* without leading to the uncontrolled replication and cell lysis mediated by the parent phage. We integrated the synthetic replication operon into the genome of *E. coli* to create a strain primed for replicating a linear replicon composed of linear double-stranded DNA flanked by PRD1 ITR sequences.

We created a Kan^R-GFP linear replicon composed of a kanamycin resistance gene and a *GFP* gene under the control of constitutive promoters flanked by 110-bp PRD1 ITR sequences on each end; the sequence was amplified by polymerase chain reaction (PCR) (Fig. 1C). We electroporated this replicon into *E. coli* bearing the synthetic replication operon in their genome and plated the cells on agar plates containing IPTG to express the operon and kanamycin to maintain the replicon. We obtained 9 ± 4 colonies per 100 μl of competent cells that grew on kanamycin and exhibited green fluorescent protein (GFP) fluorescence, which is consistent with the linear replicon being present in cells (Fig. 1D). We did not observe growth on kanamycin when the linear replicon was electroporated into cells that did not contain the synthetic replication operon, demonstrating that the synthetic replication operon is necessary for the maintenance of the linear replicon (Fig. 1D). Purification of the linear replicon and analysis by agarose gel electrophoresis confirmed the presence of the linear replicon in cells (Fig. 1E). Taken together, our experiments demonstrate that we have created a linear replicon that requires the synthetic replication operon for its maintenance and replication.

Medical Research Council Laboratory of Molecular Biology, Cambridge, UK.
*Corresponding author. Email: rtian@mrc-lmb.cam.ac.uk (R.T.); chin@mrc-lmb.cam.ac.uk (J.W.C.)

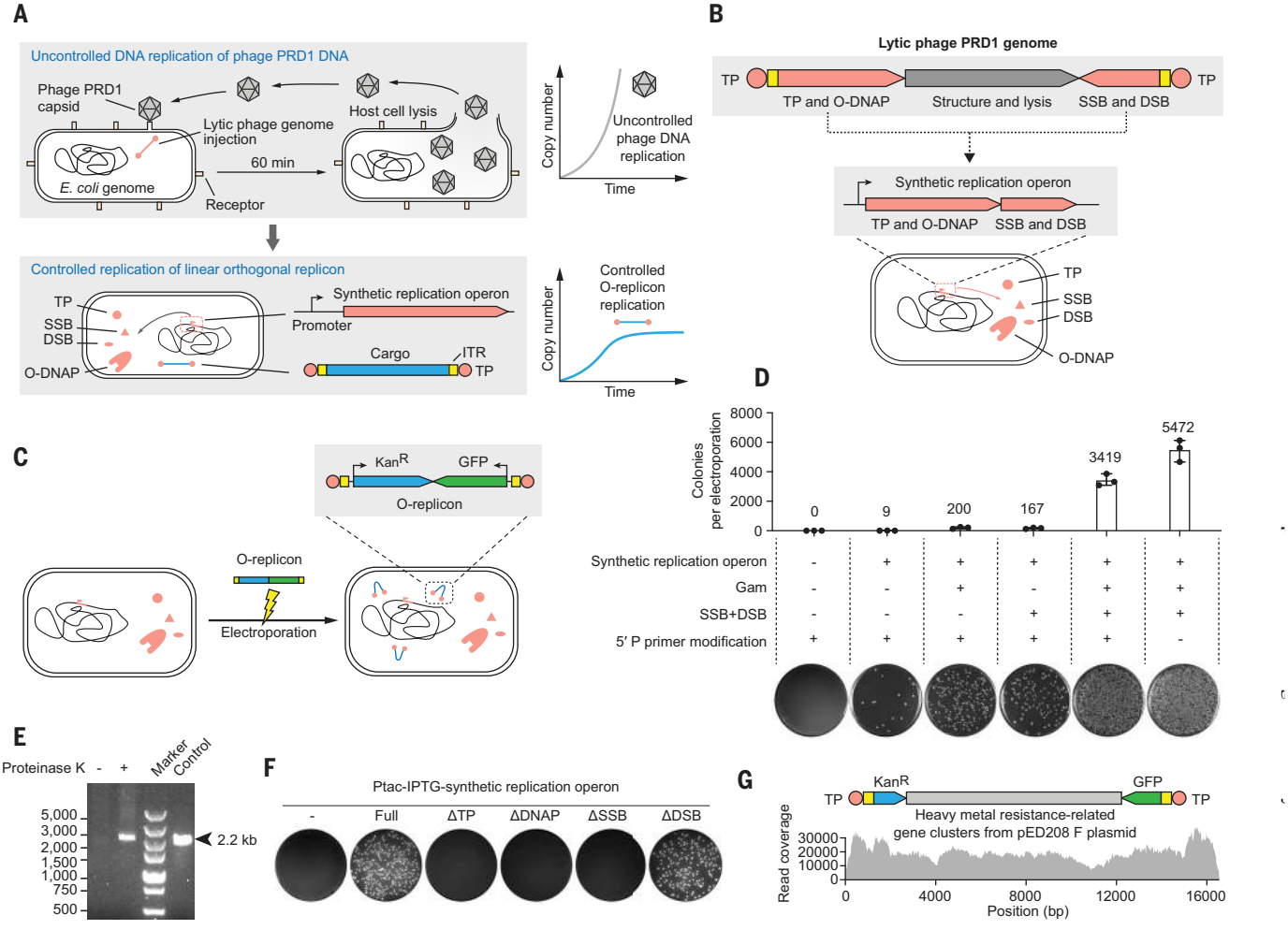


Fig. 1. Establishing a synthetic orthogonal replicon in *E. coli*. (**A**) PRD1 undergoes uncontrolled replication upon infecting *E. coli* and rapidly lyses host cells. Constructing a synthetic operon enables controlled replication of an orthogonal replicon. ITRs are shown in yellow. (**B**) We combined genes for replication of the orthogonal replicon to generate a synthetic replication operon. (**C**) Orthogonal replicons can be established by electroporating into *E. coli* cells harboring a genomic synthetic replication operon. The linear Kan^R-GFP replicon consists of flanking 110-bp ITR, a kanamycin resistance gene, and a GFP gene. (**D**) Efficiency of establishing orthogonal replicons by electroporating 3 μg of Kan^R-GFP PCR product into 100 μl (~10^9 cells). Expression of *gam*, *ssb*, and *dsb* genes from a helper plasmid increased efficiency (n = 3, error bars indicate ± SD). (**E**) Extraction of the Kan^R-GFP orthogonal replicon from cells. Proteinase K addition was needed to remove the terminal proteins. The control is a PCR product. (**F**) Essentiality of genes in the synthetic replication operon for establishing the orthogonal replicon. (**G**) A 16.5-kb orthogonal replicon. Shown is the Illumina sequencing read coverage.

We next sought to increase the efficiency with which the linear replicon could be established in cells. We expressed the Gam protein from the lambda phage, which inhibits host nucleases (RecBCD and SbcCD) and thereby protects linear double-stranded DNA from degradation (*23*), and found that it increased the number of colonies 22-fold. Overexpression of PRD1 *ssb* and *dsb* increased colony formation comparably to *gam* (Fig. 1D). Overexpressing *gam*, PRD1 *ssb* and *dsb* together increased the number of colonies 380-fold with respect to the original system. Switching to using non-phosphorylated primers with overexpressed *gam*, *ssb*, and *dsb* generated the most efficient transformation system and increased the number of colonies 608-fold with respect to the original system (Fig. 1D). The helper plasmids used to express *gam*, *ssb*, and *dsb*, were easily cured from cells once the linear replicon was established in cells (fig. S2).

We electroporated the Kan^R-GFP linear replicon into *E. coli* cells transformed with a single-copy plasmid bearing a synthetic replication operon in which one of each of the four genes in the operon (encoding TP, DNAP, SSB, and DSB) was disrupted (Fig. 1F). These experiments demonstrated that the DNAP, TP, and SSB, but not DSB, are necessary for establishing the linear replicon. Because the linear replicon is replicated by the PRD1-derived DNAP, but not the host DNAPs, we refer to it as an orthogonal replicon.

Using the most efficient transformation system, we established in vivo replication for a 16.5-kb orthogonal replicon (Fig. 1G). This demonstrates that we can use the system for large cargos.

Orthogonal replicon is stably inherited

To investigate the stability of the orthogonal replicon through many cell divisions, we followed the percentage of cells that maintain the Kan^R-GFP orthogonal replicon, as judged by the percentage of cells positive for GFP fluorescence shown by fluorescence-activated cell sorting (FACS), over 300 generations (Fig. 2A and fig. S3). In the presence of kanamycin, the Kan^R-GFP orthogonal replicon was stably maintained for 300 generations (Fig. 2A). In the absence of kanamycin, GFP fluorescence began to decay after ~50 generations (Fig. 2A). We obtained similar results using alternative

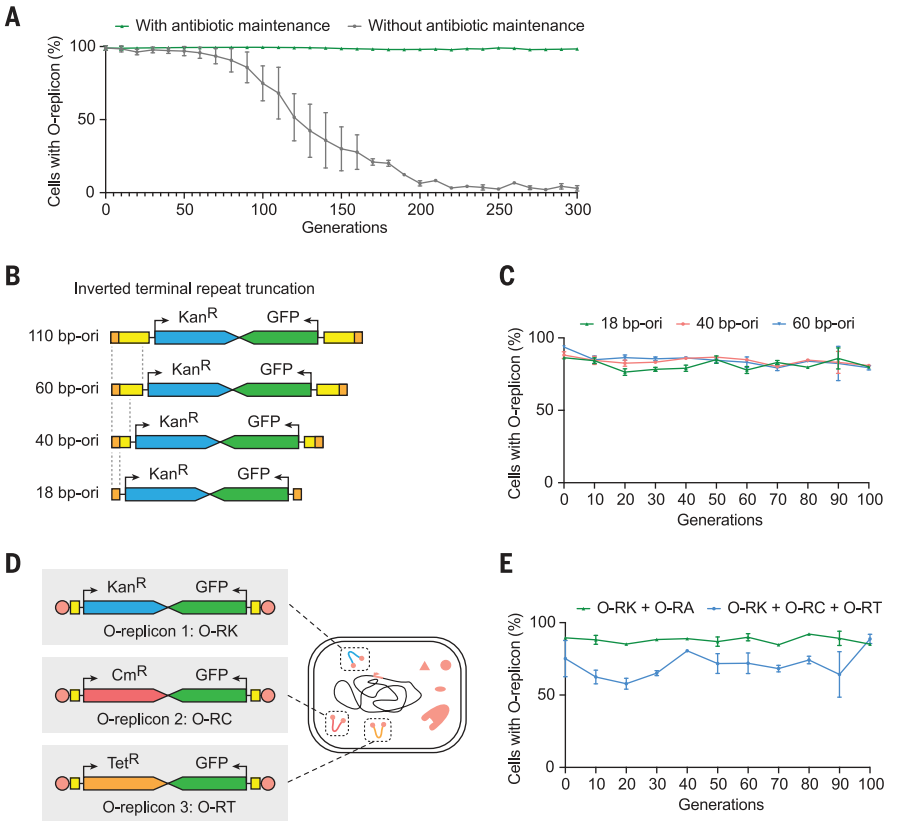


Fig. 2. Orthogonal replicons are stably maintained for hundreds of generations. (**A**) Stability of the Kan^R-GFP orthogonal replicon over 300 generations with or without kanamycin, as assessed by maintenance of GFP fluorescence using flow cytometry. The synthetic replication operon was under the control of a P*dnaKJ* promoter. (**B**) The ITR origins were iteratively truncated to establish a minimal origin for an orthogonal replicon. A single 18-bp repeat is shown in orange and the rest of the ITR in yellow. (**C**) Stability over 100 generations of the truncated orthogonal replicons shown in (B) in the presence of kanamycin as assessed by maintenance of GFP fluorescence using flow cytometry. (**D**) Multiple distinct orthogonal replicons were sequentially transformed and co-maintained under selection. (**E**) Stability, over 100 generations, of two or three co-maintained orthogonal replicons, as shown in (D), in the presence of the corresponding antibiotics as assessed by the maintenance of GFP fluorescence using flow cytometry. For the doubly transformed cells, O-RA corresponds to an orthogonal replicon carrying Amp^R. For all experiments, n = 3 and data are shown as mean ± SD.

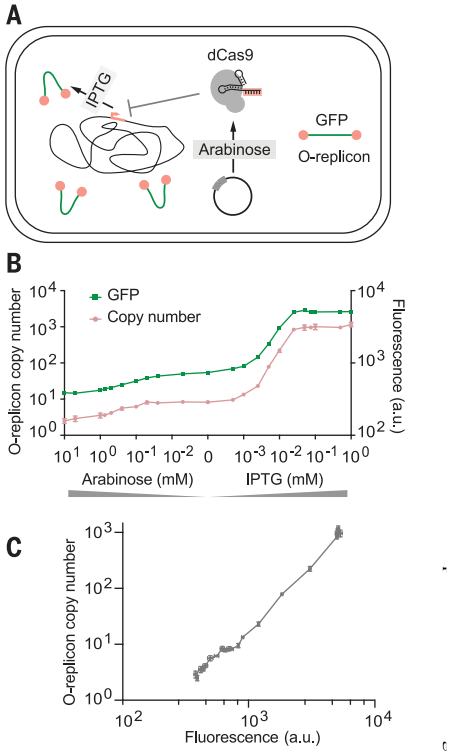


Fig. 3. Control of orthogonal replicon copy number over a 465-fold range. (**A**) Control of the orthogonal replicon copy number is achieved by inducing expression of the synthetic replication operon through IPTG addition or by down-regulating its expression using an arabinose-inducible dCas9 targeted to the IPTG-responsive Ptac promoter. (**B**) Orthogonal replicon copy number, as determined by quantitative PCR, and GFP fluorescence (normalized to OD_{600}) were measured at different arabinose or IPTG concentrations. (**C**) Correlation between Kan^R-GFP orthogonal replicon copy number and GFP fluorescence. Data from (B) were replotted for (C). For all experiments, n = 3 and data are shown as mean ± SD.

promoters to control the genomically integrated synthetic replication operon (fig. S3). These experiments demonstrated that the orthogonal replicon can be stably maintained in cells for many generations, as required for directed evolution using orthogonal replication systems.

Defining minimal origins of replication for the orthogonal replicon

To determine the minimal origin length required to establish an orthogonal replicon in cells, we prepared linear DNA with iteratively truncated ITRs (Fig. 2B). We found that replicons with ITRs truncated to 60, 40, or 18 bp could readily be established and maintained under selection for at least 100 generations (Fig. 2C and fig. S4). Linear DNA bearing 10 bp of the ITRs did not enable the replicon to be established, which is in agreement with an in vitro study of minimal PRD1 replication origins (*24*). Aligning the left origin of the PRD1 phage to the left origins of other *Tectiviridae* phages that prey on *E. coli* revealed that this minimal 18-bp sequence was conserved (fig. S1B).

Maintaining multiple distinct orthogonal replicons simultaneously

To test whether multiple distinct orthogonal replicons could be maintained in the same cell simultaneously, we sequentially established orthogonal replicons carrying different selection markers in cells (Fig. 2D). We found that at least three orthogonal replicons could be maintained simultaneously, under selection, for at least 100 generations (Fig. 2E). This may enable the directed evolution of multigene pathways without the requirement for these genes to be on a contiguous stretch of DNA.

Controlling orthogonal replicon copy number

Next, we modulated the copy number of an orthogonal replicon (expressing GFP from a constitutive promoter) in cells containing the IPTG-inducible synthetic replication operon (Fig. 3A and fig. S5). Cells also contained an arabinose-inducible dCas9 targeted to repress the IPTG-responsive Ptac promoter on the synthetic replication operon (fig. S6). By addition of arabinose or IPTG to cells, we modulated the copy number of the orthogonal replicon 466-fold, from 2.5 to 1166 copies per cell (Fig. 3B). We observed an increase in fluorescence resulting from GFP expression with increasing

orthogonal replicon copy number (Fig. 3C and fig. S7). In all cases, we further validated the precise orthogonal replicon copy numbers using quantitative PCR. These experiments demonstrated that we could regulate the copy number of the orthogonal replicon, and therefore gene expression from the orthogonal replicon, over a wide dynamic range.

Mutagenic DNA polymerases for the orthogonal replicon

To measure the mutation rate of the orthogonal replicon, we performed fluctuation analysis (*25*, *26*). We introduced a Kan^R-Cm^R(Q38TAG) orthogonal replicon that contains an amber stop codon (TAG) at position 38 of the chloramphenicol resistance gene (Cm^R) into cells containing a genomically encoded synthetic replication operon with a wild-type (WT) DNAP. We switched from using the genomically encoded WT DNAP to primarily using the plasmid-encoded DNAP of interest for replicating the orthogonal replicon by dCas9-meditated suppression of the genomically encoded synthetic replication operon and induction of plasmid encoded synthetic replication operons with mutagenic polymerases (fig. S8). After 10 generations, we measured the fraction of Cm-resistant cells resulting from point mutations that convert the TAG stop codon to sense codons. Because a single copy of the intact Cm^R gene is sufficient to confer chloramphenicol resistance, we also measured the copy number of the orthogonal replicon (fig. S9). We used this information to calculate the apparent genomic mutation rate [μ, in substitutions per base pair per generation (s.p.b.)] for the DNAP at the TAG codon in the orthogonal replicon (fig. S9). The apparent mutation rate for the WT DNAP was 8.96×10^{-10} s.p.b. We designed nine DNAPs (fig. S9) with the goal of increasing the mutation rate of the orthogonal replicon. The mutant DNAPs increased the apparent mutation rate to between 2.3×10^{-8} and 7.6×10^{-6} s.p.b. (fig. S9). The error rates of some of these DNAPs exceed the threshold of 4×10^{-7} s.p.b. that has previously been experimentally shown to lead to loss of viability with any further increase in mutagenesis (*27*). We focused on characterizing two mutant DNAPs, N71D and Y127A, because these mutant DNAPs supported linear orthogonal replicon copy numbers comparable to the WT DNAP (fig. S9). The mutation rate of the DNAPs N71D and Y127A were 9.13×10^{-7} and 5.61×10^{-7} s.p.b., respectively (Fig. 4).

O-DNA polymerases do not copy the genome

To measure the genomic mutation rate in cells containing each DNAP (WT, N71D, and Y127A), we introduced a Cm^R(Q38TAG) gene into the genome of strains containing the synthetic replication operon and switched from using the genomically encoded WT DNAP to primarily using the plasmid encoded DNAP of interest to copy the orthogonal replicon. After 10 generations, we measured the fraction of Cm-resistant cells resulting from point mutations that convert the TAG stop codon in the genome to sense codons and calculated the genomic mutation rate (μ) at the TAG codon. The genomic mutation rates with each mutant DNAP were indistinguishable from the genomic mutation rate in unmodified WT cells (Fig. 4). Moreover, the genomic mutation rates that we measured (6.4×10^{-10} s.p.b.) were comparable to those previously reported for *E. coli* (*28*, *29*). We conclude that the DNAP mutants can increase the mutation rate for replication of the orthogonal replicon without affecting the mutation rate of the genome, which is replicated by host DNAPs. The mutation rate for replication of the orthogonal replicon by the N71D and Y127A mutant DNAPs is approximately three orders of magnitude higher than the mutation rate of the genome. Overall, we conclude that the DNAP for the orthogonal replicon is an orthogonal DNAP (O-DNAP) and the orthogonal replicon and the synthetic replication operon (which contains the O-DNAP) constitute an *E. coli* orthogonal replication system (EcORep).

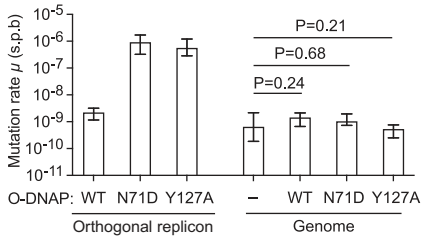


Fig. 4. Mutagenic orthogonal DNA polymerases selectively mutate the orthogonal replicon but not the genome. Determination of genomic or orthogonal replicon mutation rate (μ, in s.p.b.) for the O-DNAP and its engineered variants. The mutation rate was measured after 10 generations with fluctuation tests. For assessment of the orthogonal replicon mutation rate, we used an orthogonal replicon–encoded Cm^R gene with a TAG stop codon at position 38, and the O-DNAP variants were expressed from genomically integrated synthetic replication operons. For assessment of the genome mutation rate, we used a genomically encoded Cm^R gene with a TAG stop codon at position 38, and the O-DNAP variants were expressed from p15A plasmids using rhamnose induction. For all experiments, $n = 12$ and data are shown as mean ± upper or lower 95% bounds.

Accelerated continuous evolution of tigecycline resistance

Next, we investigated whether we could use the orthogonal replication system to continuously evolve new functions. We first investigated converting the tetracycline resistance gene *tetA* into a gene that confers resistance to tigecycline. We grew cells containing a Kan^R-TetA orthogonal replicon, which is primarily replicated by mutagenic (plasmid encoded) O-DNAPs in increasing concentrations of tigecycline (fig. S10). We completed 14 passages in 12 days. We performed 12 replicates with O-DNAP (N71D) and 12 replicates with O-DNAP (Y127A), with similar results.

After selection, we switched to replicating the orthogonal replicon with the WT O-DNAP so that it was not subject to further mutation. We obtained pools of cells that grew on 150 μg ml^{-1} tigecycline (Fig. 5A and fig. S11). For comparison, cells containing the WT *tetA* gene on the orthogonal replicon grew on agar plates containing tigecycline at 0.5 μg ml^{-1} but failed to grow on 2.5 μg ml^{-1} tigecycline. We identified numerous mutations across the promoter and 5′-untranslated region (5′-UTR), as well as synonymous and nonsynonymous mutations in the open reading frame (figs. S12 and S13 and data S1). Our experiment directly identifies mutations in *tetA* that have previously been implicated in tigecycline resistance, as well as a series of new mutations (Fig. 5B, data S1, and fig. S11). In contrast to previous work, we increased tolerance to both tigecycline and tetracycline simultaneously (fig. S11) (*10*).

We cloned selected genes into a standard circular plasmid (with a copy number ~5-fold lower than that of the orthogonal replicon). The evolved *tetA* genes conferred tigecycline resistance to 37 μg ml^{-1}, whereas the parent *tetA* gene conferred resistance to 0.25 μg ml^{-1}, and a previously reported *tetA* gene for tigecycline resistance conferred resistance to 0.5 μg ml^{-1} (Fig. 5C and fig. S14) (*10*). We conclude that in 12 days, we evolved tigecycline resistance genes that conferred resistance to 150 times more tigecycline than the starting gene and 74 times more tigecycline than in previous work.

Accelerated continuous evolution of GFP fluorescence

Next, we aimed to continuously evolve a *GFP* gene for increased green fluorescence. We grew cells containing a Kan^R-GFP orthogonal replicon in which "GFP" is a weakly fluorescing T66H variant of sfGFP (fig. S15) on a weak promoter; for simplicity, we refer to our starting variant as "WT GFP." The replicon was primarily replicated by (plasmid-encoded) mutagenic O-DNAPs (fig. S16).

To diversify the *GFP* gene and promoter, cells were diluted 1000-fold from a saturated culture and grown for 12 hours before 1000-fold dilution into fresh medium. This process was repeated four times over 48 hours before cells were sorted for GFP fluorescence. The resulting cells were then grown for a further 48 hours, with sorting for GFP fluorescence

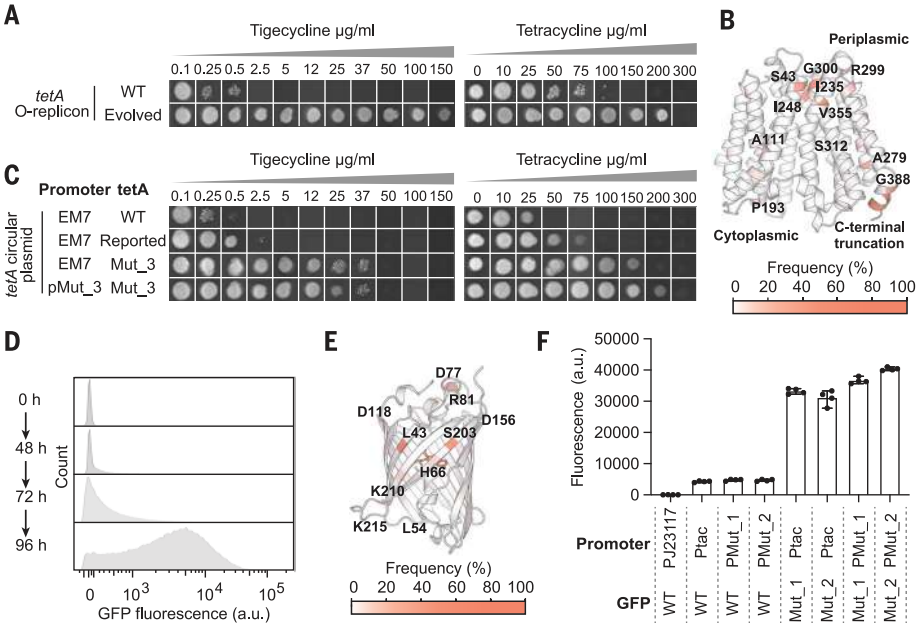


Fig. 5. Accelerated continuous evolution of tigecycline resistance and GFP fluorescence in cells. (**A**) Analysis of an evolved pool (after 14 passages) of cells carrying the Kan^R-TetA orthogonal replicon. Shown is the pool for replicate 10 performed with the N71D O-DNAP; fig. S13 shows other replicates. (**B**) AlphaFold2 model of TetA. Gradient indicates the mutational frequency of each residue. (**C**) Validation of evolved *tetA* on a ColE1 plasmid. Shown is Mut_3 from the replicate 10 pool; fig. S15 shows other mutants. Either the EM7 promoter or the evolved promoter (pMut_3) were used to drive expression. WT *tetA* or a previously reported mutant were assessed for comparison. (**D**) To select for brighter variants of a Kan^R-GFP orthogonal replicon, we iteratively isolated the brightest 0.1% of cells through FACS. Replicate 12 from the Y127A O-DNAP is shown; fig. S18 shows other replicates. (**E**) Structure of GFP (2B3P). Gradient indicates the mutational frequency of each residue. (**F**) Validation of evolved *GFP* and/or evolved promoter variants on a ColE1 plasmid. PMut_1 and Mut_1 were obtained with the N71D O-DNAP; PMut_2 and Mut_2 were obtained with the Y127A O-DNAP. For all experiments, n = 4 and data are shown as mean ± SD.

at 24 and 48 hours. All 12 replicates of this experiment for each of the two O-DNAPs were performed in <5 days (fig. S16). The population of cells progressively increased in fluorescence over the course of the experiment (Fig. 5D and fig. S17).

After selection, we switched to replicating the orthogonal replicon with the WT O-DNAP and reduced the copy number of the orthogonal replicon from 195.1 ± 3.4 and 300.5 ± 24.7 (for the N71D and Y127A O-DNAP mutants, respectively) to 9.1 ± 0.7 and 7.4 ± 0.2 (fig. S16). We then used FACS to identify clones with strong fluorescence.

Sequencing of selected clones identified numerous mutations across the promoter and 5′-UTR, as well as synonymous and nonsynonymous mutations in the open reading frame (Fig. 5E, figs. S18 and S19, and data S2). Our experiment directly identified mutations in the promoter that convert the –10 sequence to a consensus sequence and identified a number of enriched mutations in the coding sequence (fig. S19).

We picked colonies that exhibited strong fluorescence and cloned the corresponding gene into a standard circular plasmid. The selected constructs (pMut_1/GFP Mut_1 and pMut_2/GFP Mut_2) produced 36,586 ± 874 and 40,335 ± 442 arbitrary units (au) of fluorescence, respectively, whereas the starting WT *GFP* gene produced 33 ± 24 au of fluorescence (Fig. 5F and fig. S20). Thus, selection using the orthogonal replication system increased the cellular green fluorescence by >1000-fold in 5 days. Additional experiments demonstrated that mutations in both the promoter and the open reading frame of GFP make contributions to the observed increase in cellular fluorescence (Fig. 5F and figs. S21 and S22).

Discussion

We have established an orthogonal replicon in a living organism, *E. coli*, the most widely used and best characterized host, by endowing cells with a rationally designed synthetic replication operon. Our work demonstrates that orthogonal replication systems can be created de novo to enable the generation of mutagenic continuous cellular evolution systems in organisms beyond the extremely limited set in which natural replicons have been modified in vivo (*12*–*14*). The orthogonal linear double-stranded DNA replicon simply requires 18-bp DNA sequences at each end and can carry diverse cargos, including cargos too large for viral systems. The dynamic range of our control over replicon copy number exceeds that of control systems for circular plasmid copy number (*30*, *31*). Control over copy number allows control over evolutionary dynamics. Low copy number and stringent selection should favor the direct discovery of the desired genotypes that are proximal to the sequence of the starting gene. By contrast, high copy number may favor the exploration of more distal sequence space, thereby enabling the crossing of fitness valleys. These different evolutionary dynamics may be preferred in different circumstances.

EcORep provides a simple, stable, and scalable platform for accelerated continuous evolution in *E. coli*. We anticipate that it will substantially accelerate the development of diverse research tools, biopharmaceutical leads, and strains for the production of industrial chemicals.

REFERENCES AND NOTES

1. M. Lynch, *Trends Genet.* **26**, 345–352 (2010).
2. M. S. Packer, D. R. Liu, *Nat. Rev. Genet.* **16**, 379–394 (2015).
3. N. M. Low, P. H. Holliger, G. Winter, *J. Mol. Biol.* **260**, 359–368 (1996).
4. C. M. Berman *et al.*, *J. Am. Chem. Soc.* **140**, 18093–18103 (2018).
5. J. G. English *et al.*, *Cell* **178**, 748–761.e17 (2019).
6. C. E. Denes *et al.*, *ACS Synth. Biol.* **11**, 3544–3549 (2022).
7. K. M. Esvelt, J. C. Carlson, D. R. Liu, *Nature* **472**, 499–503 (2011).
8. C. L. Moore, L. J. Papa 3rd, M. D. Shoulders, *J. Am. Chem. Soc.* **140**, 11560–11564 (2018).
9. A. Cravens, O. K. Jamil, D. Kong, J. T. Sockolosky, C. D. Smolke, *Nat. Commun.* **12**, 1579 (2021).
10. X. Yi, J. Khey, R. J. Kazlauskas, M. Travisano, *Sci. Adv.* **7**, eabg8712 (2021).
11. S. O. Halperin *et al.*, *Nature* **560**, 248–252 (2018).
12. A. Ravikumar, G. A. Arzumanyan, M. K. A. Obadi, A. A. Javanpour, C. C. Liu, *Cell* **175**, 1946–1957.e13 (2018).
13. A. Ravikumar, A. Arrieta, C. C. Liu, *Nat. Chem. Biol.* **10**, 175–177 (2014).
14. R. Tian *et al.*, *Nat. Chem. Biol.* **19**, 1504–1512 (2023).
15. M. Camps, J. Naukkarinen, B. P. Johnson, L. A. Loeb, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9727–9732 (2003).
16. C. Fabret *et al.*, *Nucleic Acids Res.* **28**, E95 (2000).
17. G. T. Hess *et al.*, *Nat. Methods* **13**, 1036–1042 (2016).
18. A. Wellner *et al.*, *Nat. Chem. Biol.* **17**, 1057–1064 (2021).
19. Z. D. Blount, *eLife* **4**, e05826 (2015).
20. M. M. Poranen *et al.*, *J. Virol.* **80**, 8081–8088 (2006).
21. A. M. Grahn, J. K. Bamford, M. C. O'Neill, D. H. Bamford, *J. Bacteriol.* **176**, 3062–3068 (1994).
22. H. Savilahti, D. H. Bamford, *J. Virol.* **67**, 4696–4703 (1993).
23. K. C. Murphy, *J. Bacteriol.* **173**, 5808–5821 (1991).
24. S. K. Yoo, J. Ito, *J. Mol. Biol.* **218**, 779–789 (1991).
25. P. L. Foster, *Methods Enzymol.* **409**, 195–213 (2006).
26. B. M. Hall, C. X. Ma, P. Liang, K. K. Singh, *Bioinformatics* **25**, 1564–1565 (2009).
27. A. H. Badran, D. R. Liu, *Nat. Commun.* **6**, 8425 (2015).
28. J. Jee *et al.*, *Nature* **534**, 693–696 (2016).
29. H. Lee, E. Popodi, H. Tang, P. L. Foster, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2774–E2783 (2012).

30. S. H. Joshi, C. Yong, A. Gyorgy, *Nat. Commun.* **13**, 6691 (2022).
31. M. V. Rouches, Y. Xu, L. B. G. Cortes, G. Lambert, *Nat. Commun.* **13**, 3908 (2022).
32. R. Tian *et al*., Code for: Establishing a synthetic orthogonal replication system enables accelerated evolution in *E. coli*, version 1, Zenodo (2023); https://doi.org/10.5281/zenodo.10213374.

ACKNOWLEDGMENTS

We thank the FACS facility at the Medical Research Council Laboratory of Molecular Biology (MRC-LMB), especially P. A. Penttilä, Y. Li, F. Zhang, and D. Nolan, for support. **Funding:** This work was supported by the MRC (grants MC_U105181009 and MC_UP_A024_1008 to J.W.C.). D.C. and Y.G. were supported by European Molecular Biology Organization (EMBO) postdoctoral fellowships (ALTF 715-2022 to D.C. and ALTF 93-2023 to Y.G.). F.B.H.R. was supported by a UK Research and Innovation (UKRI) Marie Skłodowska-Curie Actions (MSCA) guarantee fellowship (EP/Y014154/1). **Author contributions:** R.T. conceptualized the project. R.T. and J.F.Z. designed approaches to establishing the O-replication system. R.T. established and investigated the stability of the O-replication system, designed and tested the O-replicon copy number control system, characterized all error-prone DNAP mutants, and performed the accelerated continuous evolution of *tetA* and GFP. F.B.H.R. designed the helper plasmid, characterized the minimal origin length, tested multiple distinct O-replicons co-maintenance, and analyzed structures of evolved proteins obtained from the continuous evolution experiments. D.C. provided the predicted structure of the O-DNAP, rationally designed nine O-DNAP mutants, analyzed structures of evolved proteins obtained from the continuous evolution experiments, and performed all flow cytometry experiments. Y.G. analyzed structures of evolved proteins obtained from the continuous evolution experiments and analyzed the Sanger sequencing–based determination mutations. K.C.L. prepared next-generation sequencing (NGS) samples. J.W.C., F.B.H.R., and R.T. wrote the manuscript with input from all authors. J.W.C. supervised the project. **Competing interests:** The MRC has filed a provisional patent application related to this work on which R.T., J.F.Z., F.B.H.R., D.C., Y.G., and J.W.C. are listed as inventors. J.W.C. is the founder of and J.F.Z. is a consultant for Constructive Bio. K.C.L. declares no competing interests. **Data and materials availability:** All data are available in the main text or supplementary materials. The computer code to analyze NGS data are available from Zenodo (*32*). The authors agree to provide any materials and strains used in this study upon request.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk1281
Materials and Methods
Figs. S1 to S22
References (*33*–*40*)
Data S1 to S3
MDAR Reproducibility Checklist

Submitted 3 August 2023; accepted 28 November 2023
10.1126/science.adk1281

AIR POLLUTION

Total organic carbon measurements reveal major gaps in petrochemical emissions reporting

Megan He[1]†, Jenna C. Ditto[1]‡, Lexie Gardner[1], Jo Machesky[1], Tori N. Hass-Mitchell[1], Christina Chen[1], Peeyush Khare[1]§, Bugra Sahin[1], John D. Fortner[1], Desiree L. Plata[1]¶, Brian D. Drollette[1]#, Katherine L. Hayden[2], Jeremy J. B. Wentzell[2], Richard L. Mittermeier[2], Amy Leithead[2], Patrick Lee[2], Andrea Darlington[2], Sumi N. Wren[2], Junhua Zhang[2], Mengistu Wolde[3], Samar G. Moussa[2], Shao-Meng Li[4], John Liggio[2]*, Drew R. Gentner[1]*

Anthropogenic organic carbon emissions reporting has been largely limited to subsets of chemically speciated volatile organic compounds. However, new aircraft-based measurements revealed total gas-phase organic carbon emissions that exceed oil sands industry–reported values by 1900% to over 6300%, the bulk of which was due to unaccounted-for intermediate-volatility and semivolatile organic compounds. Measured facility-wide emissions represented approximately 1% of extracted petroleum, resulting in total organic carbon emissions equivalent to that from all other sources across Canada combined. These real-world observations demonstrate total organic carbon measurements as a means of detecting unknown or underreported carbon emissions regardless of chemical features. Because reporting gaps may include hazardous, reactive, or secondary air pollutants, fully constraining the impact of anthropogenic emissions necessitates routine, comprehensive total organic carbon monitoring as an inherent check on mass closure.

Gaseous organic compounds are associated with considerable air quality and environmental impacts through exposure to primary emissions (*1*, *2*) and/or after their photochemical reactions and multigenerational oxidative transformations. The latter leads to secondary air pollution, including tropospheric ozone (*3*) and secondary organic aerosol (SOA)—a principal component of particulate matter ($PM_{2.5}$) (*4*) linked to major health and climate effects (*5*, *6*). Governments often mandate monitoring and reporting to develop emissions inventories to track pollutant sources and target regulatory actions. However, emissions monitoring and reporting have historically relied on discrete subsets of compounds limited to smaller hydrocarbons, with the underlying assumption that they cover the majority of carbon and/or reactivity. In reality, the chemical complexity of anthropogenic carbonaceous emissions spans a highly diverse range of molecular sizes and functionalities, including volatile organic compounds (VOCs), intermediate-volatility organic compounds (IVOCs), and semivolatile organic compounds (SVOCs) (*7*, *8*) as well as lower-volatility species in primary organic aerosol. For most research, monitoring, and reporting programs, measuring all of these individual species is not technically, logistically, or financially feasible for either a region or industrial facilities. Consequently, only a subset of carbonaceous compounds (usually VOCs) is routinely measured and/or reported as emissions.

This is particularly relevant for the oil and gas sector, for which emitted hydrocarbons can span the entire VOC-to-SVOC volatility range depending on the deposits, from light hydrocarbons in natural gas reservoirs (*9*) up to SVOCs in the case of unconventional petroleum resources (*10*). Over recent decades, global petroleum production has shifted to more unconventional sources, including heavy oil and bitumen deposits, which together are expected to account for up to 40% of global oil production by 2040 (*11*). One such deposit is Canadian oil sands, which contains an estimated 1.7 trillion barrels of oil and currently produces ~3 million barrels of crude bitumen daily (*12*, *13*), comprising the majority of Canadian oil production (*14*). This global transition to unconventional resources, and the associated diversity of emissions, presents challenges for traditional speciated VOC-focused approaches.

These challenges are evident for Canadian oil sands extraction and processing regions, where a variety of carbonaceous pollutants and their subsequent secondary products have been observed downwind of facilities (*3*, *10*, *15*–*17*). Yet limited reported emissions of individual carbonaceous species cannot be reconciled with incomplete existing emissions measurem (*15*) or explain the diverse magnitude of ondary products observed (*10*, *16*). Hence, these vast oil sands operations provide a key opportunity to examine one major petrochemical sector's reporting discrepancies caused by wide organic compound ranges that are often overlooked by traditional means but affect atmospheric chemistry, leading to SOA and ozone and their associated human and ecosystem health effects.

Using new measurements of total gas-phase organic carbon (TC) emissions from oil sands facilities, we conducted the first carbon closure experiments for any industrial source. These measurements present a powerful approach to capture the full range of organic pollutants, which we used to derive top-down facility-wide TC emissions from surface and in situ oil sands mining operations for comparison with bottom-up industry-reported values. Supported by the most chemically detailed characterization of their emissions to date as well as complementary laboratory experiments, this study demonstrates the magnitude and impact of unmonitored organic gases on emissions reporting, including IVOCs and SVOCs (I/SVOCs) from non–combustion-related sources, highlighting the need to advance routine emissions reporting and monitoring beyond traditional VOCs.

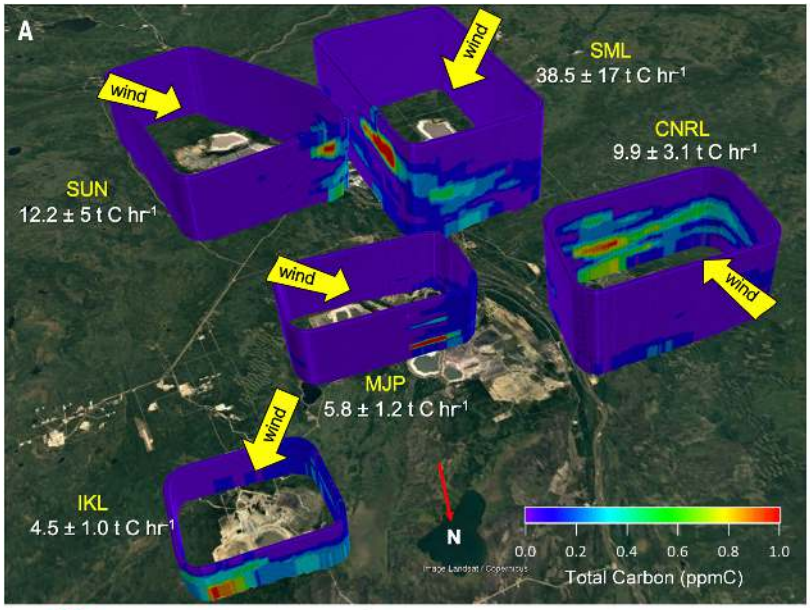
Results

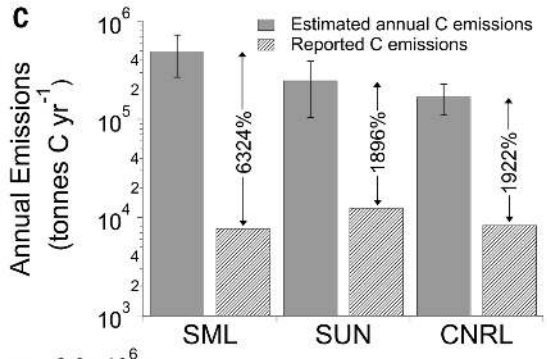
Total observed organic carbon emissions greatly exceed reported emissions

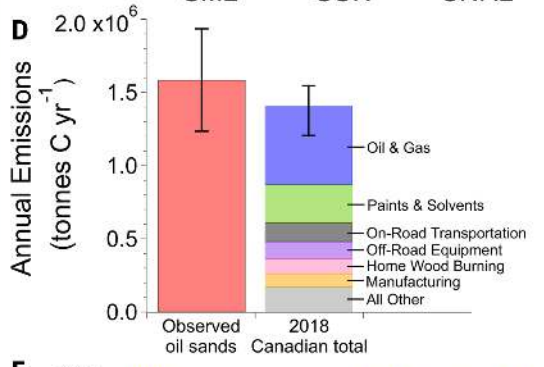
TC concentrations (excluding methane) were measured in April to July 2018 across box-shaped (n = 16) and downwind flights (n = 14) (table S1) in the Athabasca oil sands region (Alberta, Canada) by using an aircraft deployment of paired carbon dioxide (CO_2) analyzers, one with a catalyst-outfitted inlet to convert all organic gases to CO_2 (*18*). Elevated TC concentrations [>0.2 parts per million by carbon (ppmC)] were observed across facility locations and types (six surface mining and six in situ) (examples are given in Fig. 1A), from which emission rates were derived for each facility by using the top-down emission rate retrieval algorithm (TERRA) (Fig. 1, fig. S3, and table S2) (*19*–*21*). Surface mining sites use shallow oil sands reserves, whereas in situ operations extract bitumen from deeper deposits by using various methods, including steam-assisted gravity drainage (*22*).

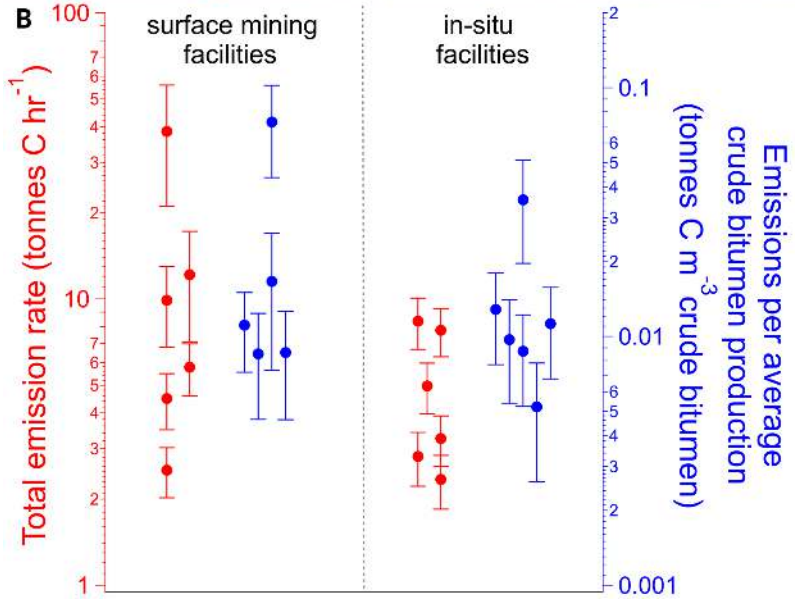
Observed hourly emission rates varied between facilities (2 to 40 tonnes C $hour^{-1}$) but were generally comparable between surface mining and in situ facilities (Fig. 1B and fig. S3) despite substantial differences in on-site operations and typically lower crude bitumen production rates for individual in situ operations. When normalizing the annualized emission rates by reported facility-level annual crude bitumen production (*23*), the average TC emission intensity (excluding methane) across

[1]Department of Chemical and Environmental Engineering, Yale University, New Haven, CT, USA. [2]Air Quality Research Division, Environment and Climate Change Canada, Toronto, ON, Canada. [3]National Research Council Canada, Ottawa, ON, Canada. [4]College of Environmental Sciences and Engineering, Peking University, Beijing, China.
*Corresponding author. Email: drew.gentner@yale.edu (D.R.G.); john.liggio@ec.gc.ca (J.L.)
†Present address: School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA.
‡Present address: Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, St. Louis, MO, USA.
§Present address: Laboratory of Atmospheric Chemistry, Paul Scherrer Institute, AG-5232 Villigen, Switzerland.
¶Present address: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.
#Present address: Ramboll, Westford, MA, USA.









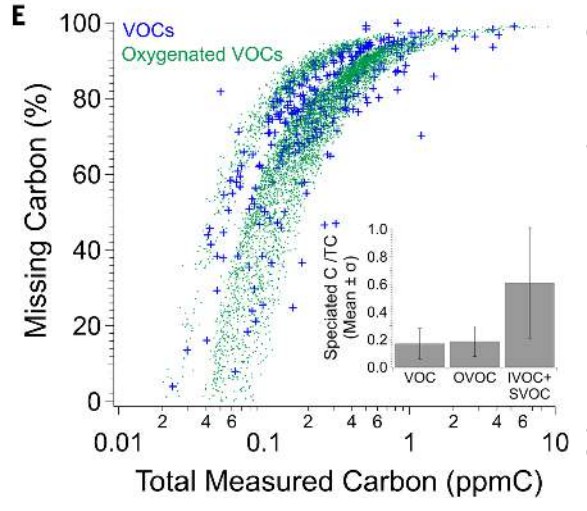


Fig. 1. Observed total gaseous organic carbon emissions, their hydrocarbon intensity, and comparisons with reported emissions. (**A**) Examples of box flights around five major surface mining facilities on different days show elevated downwind total gaseous organic carbon with total emissions derived with TERRA (supplementary materials, materials and methods). Numeric values in white indicate average TC emission rates. (**B**) Hourly carbon emission rates and average annual carbon intensities for surface mining and in situ facilities. Each marker indicates the mean for each site, and error bars indicate the standard deviation (number of flights per site is provided in fig. S3 and table S2). (**C**) Estimated annual gaseous organic carbon emissions compared with the reported emissions converted to carbon mass units for the three highest-emitting (both measured and reported) surface mining facilities (SML, SUN, and CNRL) (table S2), with percent differences. Annual emissions were estimated by using TC/NO_x ratios, and error bars indicate the standard deviation of the derived TC/NO_x ratios (with emissions derived as the TC/NO_x ratio scaled by reported annual NO_x emissions) (fig. S4 and supplementary materials). (**D**) Observed total gaseous organic carbon emissions for the studied facilities compared with the total Canadian annual inventory for 2018 converted to carbon units. (**E**) Percentage of "missing" organic carbon relative to either VOC or OVOC measurements based on canister samples, PTR-ToF-MS, and iodide-CIMS. (Inset) Average contributions of VOCs, OVOCs, and I/SVOCs to total observed organic carbon measurements in concentrated plumes (>0.35 ppmC), which represents the top 75th percentile of TC data. This is not in comparison with emissions inventories. For the purpose of comparing with the discrete speciated VOCs and OVOCs that are predominantly C_{10} and smaller, the IVOC+SVOC value in the inset is inclusive of C_{11} compounds.

sampled surface and in situ facilities was 0.024 ± 0.010 and 0.014 ± 0.006 tonnes C m^{-3} bitumen, respectively (Fig. 1B). These hydrocarbon intensities translate to total facility-wide emissions that are equivalent to 0.3 to 12.1% of production by mass (table S2), which is comparable with the magnitude of loss rates of highly volatile methane from US oil and gas operations (0.3 to 8.9%) (*24*). The magnitude of these emissions emphasizes the importance of total hydrocarbon measurements in capturing infrequently measured nonmethane organic compounds.

Total organic carbon annual emissions for facilities were estimated by using TC-to-NO_x (a combustion tracer) ratios multiplied by

reported annual NO_x emissions (supplementary materials) (*25*, *26*), which has been performed for other pollutants (*27*–*29*). Emission ratios were obtained by means of empirical concentration correlations during box flights and TERRA-derived direct emission ratios (supplementary materials and fig. S5). The density of sources within facilities and nearby atmospheric mixing leads to both combustion and noncombustion gas-phase organic carbon sources being mixed and thus moderately correlated to NO_x in downwind measurements (fig. S5), which is similar in approach to well-correlated anthropogenic tracers downwind of major urban areas (*28*). Across the three highest-emitting facilities—Syncrude Mildred Lake (SML), Suncor (SUN), and Canadian Natural Resources (CNRL)—these average ratio values were 24 ± 11, 13 ± 7, and 17 ± 6 kg C (kg NO_x)$^{-1}$, respectively (Fig. 1C), yielding annual emissions estimates of ≈200,000 to 500,000 tonnes C year^{-1}. Although scaling with TC/NO_x is more robust than simple annual extrapolation (24 hours × 365 days), extrapolation remains within a factor of 2.2 on average (maximum 3), and both methods result in annual estimates far greater than reported emissions (table S2).

These large emission rates were 20 to 64 times greater than those in the Alberta Emissions Inventory Report (AEIR) and Canada's National Pollutant Release Inventory (NPRI), the latter of which is required to include the entire VOC-to-SVOC range for oil sands operations (Fig. 1C and tables S2 and S3). For context, the sum of measured gas-phase organic carbon emissions from all measured surface mining and in situ facilities in 2018 was 1.59 × 10^6 tonnes C year^{-1}, which is approximately equivalent to the VOC emissions reported for the sum of all anthropogenic sources in Canada's Air Pollutant Emissions Inventory (1.40 × 10^6 tonnes C year^{-1}) (carbon mass conversion is provided in the supplementary materials) (Fig. 1D) (*30*). Surveyed facilities included 88 and 50% of 2018 crude bitumen production from surface mining and in situ sites, respectively (table S2) (*23*, *31*). Thus, the oil sands sector alone represents a dominant fraction of country-wide gas-phase organic carbon emissions, even when only including the facilities studied here.

Measured VOCs only account for a fraction of total measured organic carbon (fig. S1), reflecting the need to report the full range of organic volatilities across oil sands and other anthropogenic sectors. Even when including measurements of oxygenated VOCs (OVOCs) with two on-board high-resolution mass spectrometers [proton transfer reaction–time of flight mass spectrometry (PTR-ToF-MS) and iodide–chemical ionization mass spectrometry (iodide-CIMS)] (table S4), a substantial fraction of carbon remains "missing" relative to the total carbon observations (Fig. 1E). In this case, "missing" indicates that the sum of speciated carbon is less than the total measured carbon. At lower total carbon concentrations (background air), most of the observed total carbon was speciated. In concentrated oil sands plumes with TC concentrations >0.35 ppmC, VOC and OVOC measurements were only responsible for 17 ± 11% and 19 ± 11% of carbon, respectively (Fig. 1E, inset). Conversely, the I/SVOCs observed in integrated low time-resolution adsorbent tube samples represented a greater fraction (61 ± 40%) (Fig. 1E and fig. S2), highlighting the abundant contributions of I/SVOCs to total oil sands–related emissions and their insufficient bottom-up quantification in reported emissions (table S3).

Abundant complex mixtures of oil sands–derived I/SVOCs near facilities

Time- and spatially integrated samples of I/SVOCs were collected during box flight segments (for example, Fig. 1A) and downwind transects and analyzed by means of gas chromatography on both unit-resolution and high-resolution mass spectrometers [gas chromatography–electron ionization–mass spectrometry (GC-EI-MS) and gas chromatography–time of flight (GC-ToF)], which revealed abundant complex mixtures of I/SVOCs near both surface mining and in situ facilities (Figs. 2 and 3). IVOCs (C_{12} to C_{18}) and SVOCs (C_{19} to C_{25}) were uncharacteristically abundant relative to VOCs (Fig. 1E) and were observed around various facilities, as shown in selected flight samples in Fig. 2A (additional examples are available in figs. S6 and S7). The relative abundances and composition varied between and around facilities, with maxima ranging from C_{17} to C_{22} (Fig. 2A, figs. S8 and S9, and tables S5 and S6), which may suggest varying on-site sources and emissions

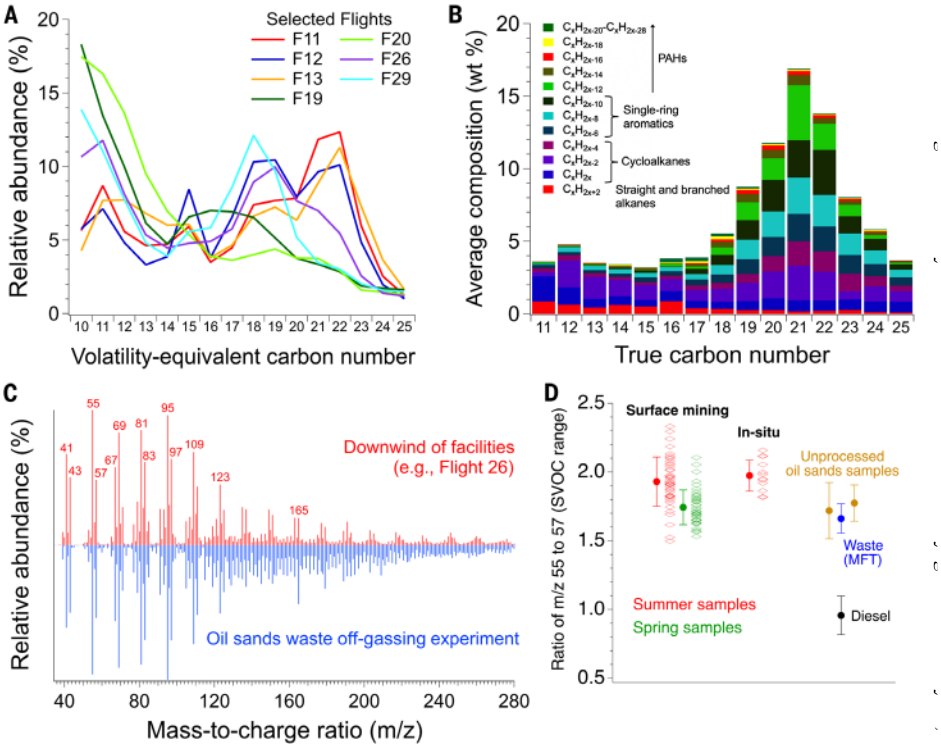


Fig. 2. Chemically speciated observations of abundant gas-phase I/SVOC mixtures near surface and in situ mining facilities are indicative of oil sands origin. (**A**) Relative volatility distributions of observed complex I/SVOC mixtures vary between facilities, shown as total ion chromatograms (across C_{10} to C_{25} by means of GC-EI-MS) in which each line is a sample from selected flights (other flights are available in fig. S6). (**B**) Average chemical composition of I/SVOC emissions across flight samples by means of high-resolution GC-ToF is consistent with the characteristics of oil sands bitumen indicating depleted acyclic (linear or branched) alkanes and relatively more mono-, bi-, and tri-cyclic alkanes. (**C**) SVOC mass spectra (by means of GC-EI-MS) from aircraft samples (flight 26) share similar characteristic mass spectral fragments with oil sands, including for multicyclic alkanes (for example, *m/z* 69, 81, 83, 95, 109, and 123), shown here with an average spectrum from oil sands MFT waste off-gassing experiments (Fig. 5). Flight 26 was chosen as an example with marked enhancement downwind of a concentrated area of facilities. (**D**) Average (±SD) of *m/z* 55/57 ratios measured with GC-EI-MS across all flight adsorbent tube samples (individual points) further demonstrates consistent reduced abundances of acyclic alkanes, shown with ratios from other oil sands materials (extractions of two types of unprocessed oil sands and MFT waste) and diesel fuel (*7*) for comparison.

pathways. There are stark differences in the observed concentrations when compared with that of urban areas. Average concentrations of primary gas-phase IVOCs were 6.3 ± 1.9 μg m^{-3} in greater Los Angeles, with primary gas-phase SVOC estimates of 0.6 μg m^{-3} (*1*). We observed average I/SVOC concentrations of 104 ± 93 μg m^{-3} (range, 10.2 to 409 μg m^{-3}) across flight samples, accompanied by corresponding total carbon enhancements (fig. S2).

Detailed chemical speciation of offline samples provided I/SVOC composition at the molecular formula level, with variations in I/SVOCs across samples (figs. S8 and S9). The average distribution based on all flight samples (Fig. 2B) exhibited a ~C_{20} to C_{22} maximum with aliphatic (alkane), single ring-aromatic, and polycyclic aromatic hydrocarbon (PAH) formulas comprising 43, 39, and 18% of the mass across the IVOC to SVOC range, respectively (Fig. 2B). These observed complex I/SVOC mixtures were consistent with the composition of oil sands materials in prior literature (*32*, *33*) and our own analysis of oil sands material samples (Fig. 2, C and D). This includes the large aromatic content substantially exceeding aliphatics in raw oil sands (*33*) and depleted levels of acyclic alkanes, with a large fraction of mono- through tetra-cyclic alkanes in the ~C_{15} to C_{23} range observed in Athabasca bitumen (*32*). Airborne measurements show relatively minor contributions from acyclic (linear or branched) alkanes, and prominent mass-to-charge (*m*/*z*) fragments associated with mono- and multicyclic alkanes (Fig. 2, B to D) (*34*). These cyclic-to-acyclic alkane ratios are elevated across flights and oil sands materials and are atypical of observations of common I/SVOC sources (such as diesel fuel combustion) (Fig. 2D) (*7*), further supporting that the observed I/SVOCs are oil sands–derived.

I/SVOC enhancements were often observed around and directly downwind of both surface mining and in situ facilities (Fig. 3). For example, flights 25 and 26 initially focused on a forest fire (*18*, *35*) upwind of oil sands operations, with the fifth screen, which was downwind of all major surface mining facilities, showing a marked enhancement in SVOC abundances and mass spectra indicative of oil sands–derived emissions (Figs. 2C and 3, A and B, and fig. S7B). The strong vertical gradient in screen 5′s transects (Fig. 3B) with higher SVOC abundances at lower altitudes implies ground-level emissions (<500 m). Enhancements were also observed directly downwind of in situ facilities (for example, flight 29) (Fig. 3, C and D), providing additional evidence of I/SVOC emissions from in situ operations.

An important role for noncombustion carbon emissions

Measurements of TC were compared with established combustion tracers (NO_y; sum of all

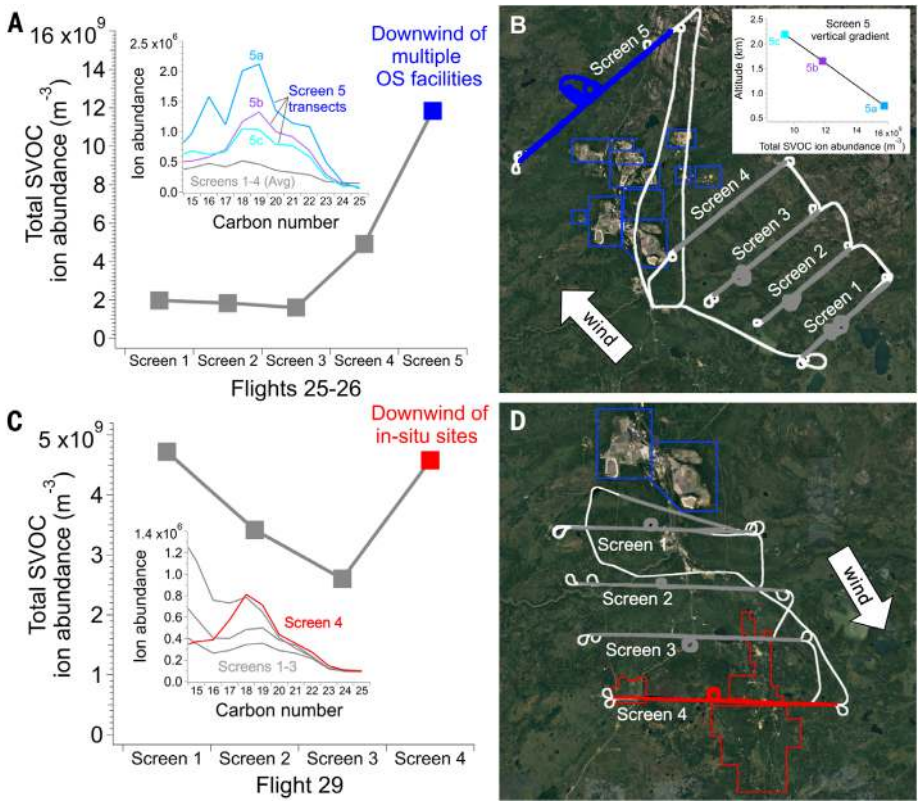


Fig. 3. Offline measurements of semivolatile organic compounds observed downwind of surface mining and in situ facilities. (**A**) Total SVOC ion abundances by using GC-EI-MS compared across upwind (screens 1 to 4) to downwind (screen 5) samples during consecutive flights 25 and 26. (Inset) The distribution of *n*-alkane volatility-equivalent C_{15}-to-C_{25} ion abundances (screen 5 samples are labeled 5a, 5b, and 5c, indicating different altitudes of the screen's transects above sea level). (**B**) Map of the five screens across flights 25 and 26, with screen 5 downwind of all surface mining facilities (outlined). (Inset) Vertical gradient with larger enhancements at lower altitudes for the three screen 5 transects. (**C**) Total SVOC ion abundances for flight 29 screens, in which screens 1 to 3 (gray) are downwind of surface mining facilities and screen 4 (red) is immediately downwind of multiple in situ facilities. (Inset) The distribution of C_{15}-to-C_{25} ion abundances. (**D**) Map of screens in flight 29 with outlined surface mining (blue) and in situ (red) facilities. Flights 25, 26, and 29 were conducted during daytime: flights 25 and 26, 8:45 to 17:20, and flight 29, 9:45 to 14:30, local time.

oxides of nitrogen). These TC/NO_y ratios were used to examine the relative contributions of organic carbon emitted from combustion-related (for example, vehicles and equipment) versus non–combustion-related sources (for example, evaporative and fugitive emissions). In addition to expected combustion-related emissions, there was clear evidence for substantial non–combustion-related emissions.

For example, flight 29's TC and NO_y measurements (Fig. 4A) show correlated enhancements across plume transects in screens 1 to 3 downwind of surface mining facilities (Fig. 3D), which is indicative of co-located emissions, although not necessarily co-emitted from the same on-site source(s). TC/NO_y ratios remained similar across the first three screens (~0.08 ppmC ppb^{-1}), with downwind transport and dilution of the plume from surface mining facilities. However, in screen 4, after intercepting emissions from the in situ facilities, the ratio increased to 0.19 ± 0.41 ppmC ppb^{-1}, with abundant TC enhancements from in situ facilities that were not correlated with NO_y, indicating that they were no longer co-located with combustion-related sources. The corresponding I/SVOC enhancements (Fig. 3, C and D) and spatially resolved analysis of screen 4 show clear TC enhancements at the lowest flight altitude (Fig. 4B and fig. S10), despite continued dilution of the upwind NO_y plume.

Elevated TC/NO_y ratios were also observed across other flights (flights 3, 4, 7 to 14, 17, 18, 20 to 22, 24, 29, and 30). The ratios are indicative of major contributions from non–combustion-related emissions pathways because they are substantially greater (≥10× on average) than the expected ratios from combustion-related sources (such as gasoline and diesel engines) in North American emissions inventories (Fig. 4C) (*30*, *36*). This necessitates further efforts to constrain both combustion and noncombustion sources at oil sands operations, including a broader consideration of organic carbon emissions (such as I/SVOCs).

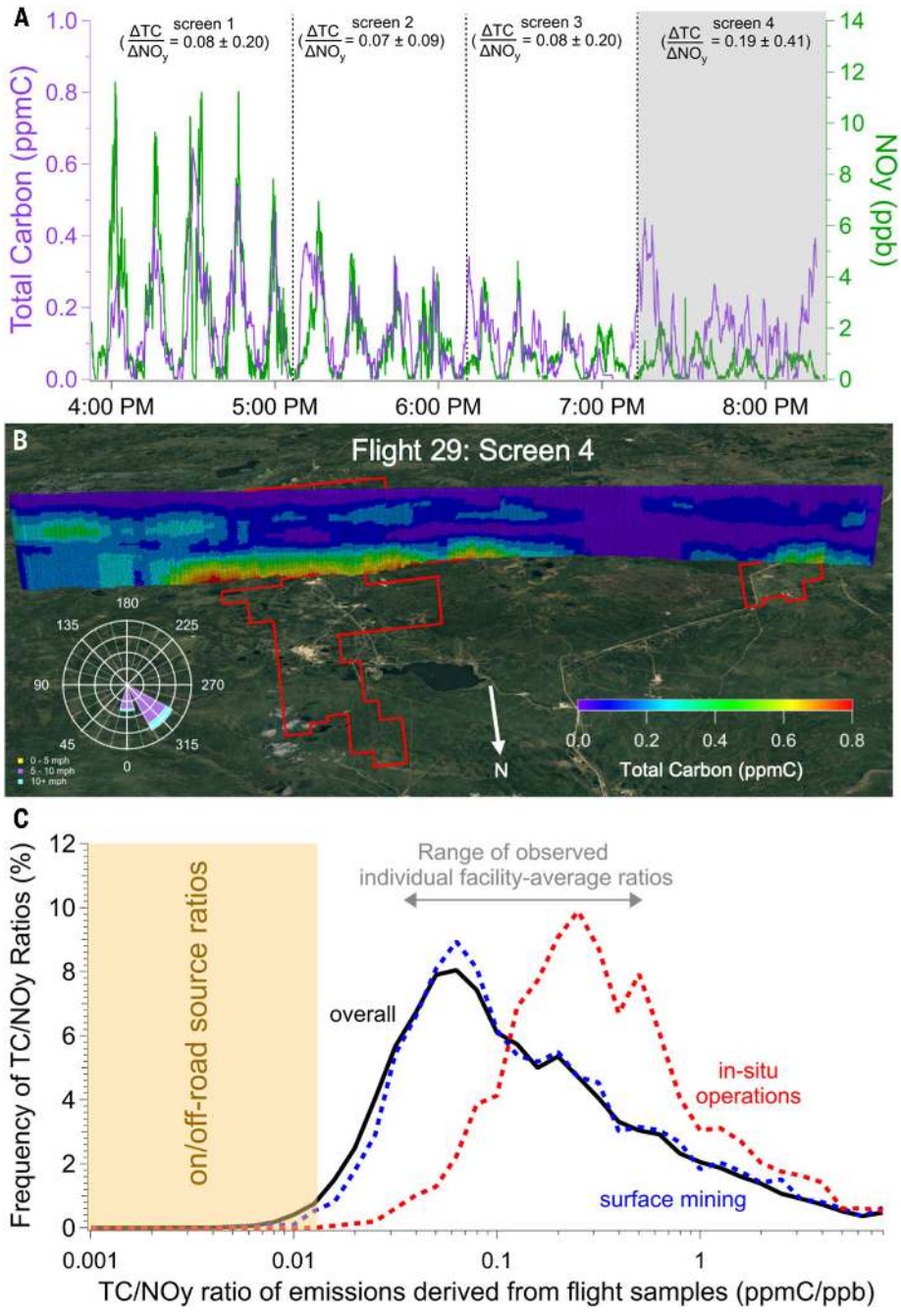


Fig. 4. Total gaseous organic carbon enhancements and their ratios to nitrogen oxide combustion tracers (NO_y) highlight the importance of non–combustion-related emissions. (**A**) Comparison of TC and NO_y background-subtracted concentrations across the four screens of flight 29, shown with average TC/NO_y ratios (above the 50th percentile). (**B**) Spatially resolved observations of TC corresponding to flight 29 screen 4 in (A) show enhancements in TC downwind of in situ facilities (red outline delineates Long Lake, Surmont, and JACOS Hangingstone) near ground level (Fig. 3C, map). Additional details on flight 29 can be found in the supplementary materials and figs. S10 and S11. (**C**) TC/NO_y ratios (10 s averages) across both facility types (solid black line), as well as flights around surface mining only (flights 11, 13, 20, 21, 22, 24, and 29, screens 1 to 3; dashed blue line) and in situ facilities only (flights 12, 18, and 29, screen 4; dashed red line). Background-subtracted concentrations above the 50th percentile within each data subset were used to focus the analysis on more concentrated plumes (>0.07 ppmC overall, >0.11 ppmC for surface mining only, and >0.06 ppmC for in situ only), with all data shown in figs. S12 and S13. The range of known ratios from on- and off-road sources (*30*, *36*) is shown for comparison. The surface mining and in situ distributions are not additive to the "overall" distribution, which encompasses additional flights.

Considering potential noncombustion emission pathways

Oil sands extraction and processing encompass multifaceted operations that vary with facility type, extraction methods, processing capabilities, and various on-site activities that may contribute to non–combustion-related emissions. Such I/SVOC emissions can be expected during mining operations from raw oil sands off-gassing and fugitive emissions during extraction and processing (*37*). Yet emissions may extend past processing stages, warranting holistic lifecycle-wide consideration of potential sources, including waste management. For example, tailings ponds are managed open pits that contain wastewater and by-products of the bitumen separation process, and off-gassing of I/SVOCs from tailings ponds has been hypothesized as a major source (*37*). However, available field measurement methods have been limited predominantly to VOCs (*3*, *15*), and the presence of water inhibits emissions owing to rate-limiting multiphase partitioning processes (*37*).

Decades of oil sands surface mining have resulted in large volumes of accumulated fluid tailings waste (water and solids), necessitating tailings reclamation measures to reduce the volume of tailings waste stored in ponds (*38*). We evaluated I/SVOC emissions from a tailings drying technique, which is used by the oil sands industry to process aged or fresh fine tailings. In 2018, 252 Mm^3 of treated fluid tailings were reported industry-wide (*38*). We specifically examined off-gassing emissions from mature fine tailings (MFT), an older mixture of fine particles (sand, silt, and clay) with residual bitumen that remains suspended in tailings ponds and is particularly difficult to separate from wastewater. Although several methods exist, we emulated atmospheric fines drying (also called thin lift drying or tailings reduction operations) through a series of bench-top experiments. Other methods such as accelerated drying techniques may vary. Yet all dried tailings (fine or coarse) are typically kept in either temporary storage areas, transferred to dedicated disposal areas, or used in construction projects (such as roads or dykes)—all of which are open to the atmosphere for some duration (*39*, *40*).

Although MFT off-gassing was initially relatively low, emissions increased markedly once the MFT were dry and remained elevated for weeks at environmentally relevant temperatures (Fig. 5A and table S7). Without the inhibiting water barrier, the diffusion of I/SVOCs through the dried MFT continued over the experiments' duration (up to 9 weeks), with increased emissions at higher surface temperatures and irradiation to simulate solar exposure (Fig. 5B and table S8). Whereas initial MFT off-gassing over the first 11 days included some VOCs (fig. S15), the emissions shifted toward the I/SVOC range after drying and aging and under increased

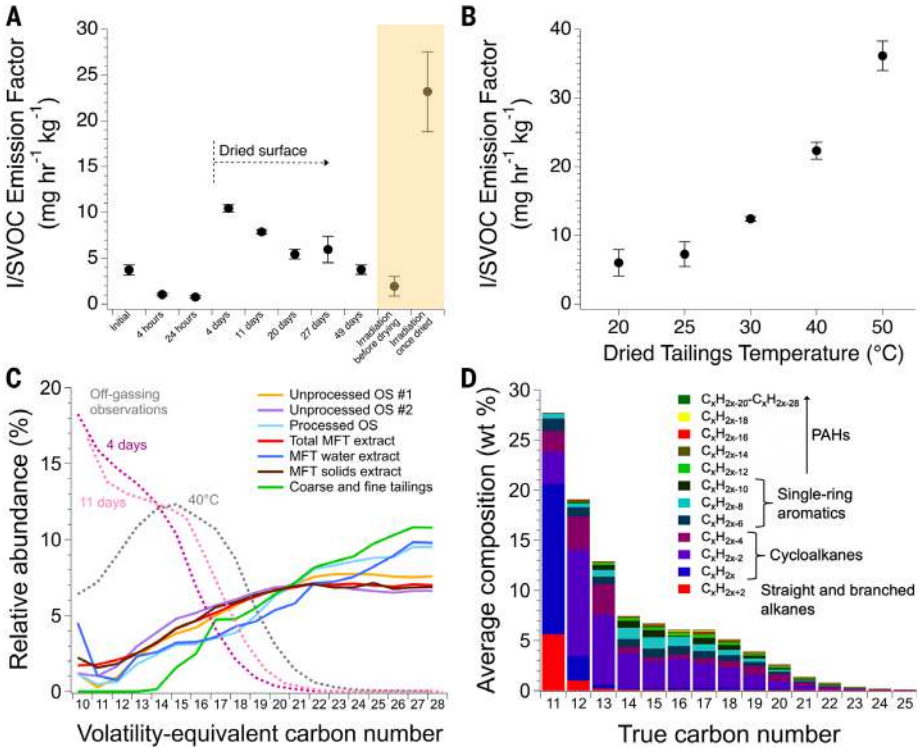


Fig. 5. Dried oil sands waste releases substantial quantities of I/SVOCs from hydrocarbon reservoirs sorbed to suspended tailings solids. (**A**) I/SVOC emission factors from MFT at various stages of drying-aging for an industry-supplied undried sample, and with irradiation of undried and dried tailings. (**B**) Temperature dependence of emissions from dried tailings [units in (A) and (B) are in *n*-alkane equivalent mass per mass of dry tailings, and error bars reflect emission factor SDs]. (**C**) Demonstration of underlying I/SVOC reservoirs in unprocessed oil sands, processed oil sands, and waste products as a function of *n*-alkane volatility-equivalent carbon number (by means of GC-EI-MS ion abundance). Examples of off-gassing emissions from MFT (including several days after application) are indicated with dashed lines for comparison. (**D**) Average I/SVOC composition observed in fresh and partially dried MFT off-gassing (samples both with and without irradiation at ~25°C are included here) by means of GC-TOF.

temperatures or irradiation (Fig. 5, A and B). The chemical speciation of MFT off-gassing emissions exhibits an enhancement in cyclic alkanes similar to that of ambient measurements, with characteristic fragments of multicyclic alkanes and limited acyclic alkanes (Figs. 2, C and D, and 5D; fig. S14; and table S9). The reservoir of I/SVOCs was not just present in MFT but was also observed in a range of materials spanning unprocessed oil sands, processed oil sands, and waste products (Fig. 5C). So although multiple similar on-site sources may exist, these observations show that tailings drying could be an important source of I/SVOCs, making substantial contributions to the total organic carbon observed in the flights.

Discussion

The magnitude of TC emissions observed from oil sands facilities far exceeds industry reports, with observed emissions [1.59 ± 0.35 million tonnes (Mt) C year^{-1}] being equivalent to the total Canadian anthropogenic emissions of organic carbon (Fig. 1, C and D). Total oil sands organic carbon emissions also far surpass reactive organic gas emissions from total anthropogenic sources (stationary, mobile, and chemical products) in the largest US megacities (such as Los Angeles) (~0.1 Mt C year^{-1} in the South Coast Air Basin) (*41*). These findings demonstrate that complete coverage of a wider volatility range of emissions is necessary to effectively inform science and policy because speciated VOC reporting alone is insufficient to capture the entire range of carbon emissions (Fig. 1, C and E, and fig. S2). Although oil sands operations are required to report "analytically unresolved hydrocarbon" (AUHC; including I/SVOCs), only a small number of operators reported such emissions in 2018, with negligible contributions to total reported organic carbon (0 to 2.6 × 10^{-3} Mt C year^{-1}) in 2018 (*42*). In subsequent years, reported AUHC has increased in magnitude (up to 1.5 × 10^{-2} Mt C year^{-1} in 2021) yet remain a minor contributor (~1%) to total measured emissions. Given that I/SVOCs are estimated to represent ~60% of carbon in concentrated plumes (Fig. 1E), the full air quality and environmental impacts of oil sands operations cannot be evaluated without more realistic inclusion of IVOCs and SVOCs in emissions reporting.

Although a diverse range of on-site sources—including from extraction, processing, and tailings ponds (*37*)—likely contributes to the observed total organic carbon emissions from surface mining facilities, our laboratory experiments identified potential unintended consequences of tailings reduction strategies to reduce the volumes of tailings water. This warrants further measurements and consideration of off-gassing emissions resulting from oil sands waste management strategies, especially given that the dewatering of tailings by means of a variety of other forced-drying techniques will similarly produce dried solids without an inhibiting water layer. Although not currently considered a VOC-SVOC source, this is a timely issue because the surface area of nonfluid tailings in the oil sands has grown considerably with increased oil sands production over the past several decades, with 119 km^2 (in 2020) representing 40% of total waste surface area (*43*). Hence, the potential for dried waste products to emit large amounts of reactive I/SVOCs to the atmosphere suggests that reducing liquid waste by such methods opens up potentially large and unexpected pathways of atmospheric pollution.

Prior work has focused on surface mining operations, but total gaseous organic carbon emissions from in situ facilities also greatly exceed reported emissions. With total carbon emissions per bitumen production (hydrocarbon intensity) comparable with that of surface mining (Fig. 1B), further examination of their emissions is warranted because the proportion of bitumen production from in situ extraction will increase beyond ~50% over the coming decade (*13*).

Effective emissions mitigation to achieve co-benefits across air quality–, health-, climate-, and energy-related goals requires accurate representation in inventories. This cannot be accomplished without the combination of both bottom-up and top-down approaches to examine closure and reveal emissions that require further scrutiny. In the case of both oil sands operations and many other anthropogenic sources, routinely monitoring all gas-phase organic carbon emissions with complete speciation is often infeasible for researchers, operators, and regulators. For I/SVOCs, the challenges associated with measuring their inherently complex mixtures demonstrate how total organic carbon observations would enable inclusive, routine carbon coverage across an anthropogenically ubiquitous class of compounds that drive secondary organic $PM_{2.5}$ formation (*7*, *8*, *44*, *45*). The total organic carbon approach here can also be a valuable tool used to capture a broader range of chemical species across the VOC-SVOC range, thus identifying the presence of previously unknown hydrocarbons or functionalized organic compounds, unknown or underconstrained

sources, spatiotemporally variable emissions, hotspots, or noncompliance. This facilitates the quantification of sources in emissions inventories and the modeling of their contributions to both primary hazardous pollutant concentrations and to secondary pollutant formation. Future applications can similarly improve emissions reporting across many other anthropogenic sources and locations because accounting for life cycle wide emissions of chemically diverse compound classes by means of total carbon monitoring presents a vastly simpler approach with inherent mass closure checks for industry, scientists, and policy-makers alike.

REFERENCES AND NOTES

1. Y. Zhao *et al*., *Environ. Sci. Technol.* **48**, 13743–13750 (2014).
2. E. N. Kelly *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22346–22351 (2009).
3. I. J. Simpson *et al*., *Atmos. Chem. Phys.* **10**, 11931–11954 (2010).
4. J. L. Jimenez *et al*., *Science* **326**, 1525–1529 (2009).
5. S. Feng, D. Gao, F. Liao, F. Zhou, X. Wang, *Ecotoxicol. Environ. Saf.* **128**, 67–74 (2016).
6. S. Fuzzi *et al*., *Atmos. Chem. Phys.* **15**, 8217–8299 (2015).
7. D. R. Gentner *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18318–18323 (2012).
8. Q. Lu, Y. Zhao, A. L. Robinson, *Atmos. Chem. Phys.* **18**, 17637–17654 (2018).
9. G. Pétron *et al*., *J. Geophys. Res. Atmos.* **119**, 6836–6852 (2014).
10. J. Liggio *et al*., *Nature* **534**, 91–94 (2016).
11. L. Rosa, K. F. Davis, M. C. Rulli, P. D'Odorico, *Earths Futur.* **5**, 158–170 (2017).
12. Government of Alberta, "Environmental management of Alberta's oil sands" (2009); https://open.alberta.ca/publications/9780778576778.
13. Alberta Energy Regulator, "Crude Bitumen - In Situ Production" (2023); https://www.aer.ca/providing-information/data-and-reports/statistical-reports/st98/crude-bitumen/production/in-situ.
14. Alberta Energy Regulator, "ST3: Alberta Energy Resource Industries Monthly Statistics" (2017); https://www.aer.ca/providing-information/data-and-reports/statistical-reports/st3.
15. S. M. Li *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, E3756–E3765 (2017).
16. J. Liggio *et al*., *Atmos. Chem. Phys.* **17**, 8411–8427 (2017).
17. J. R. Brook *et al*., *J. Air Waste Manag. Assoc.* **69**, 661–709 (2019).
18. K. L. Hayden *et al*., *Atmos. Chem. Phys.* **22**, 12493–12523 (2022).
19. M. Gordon *et al*., *Atmos. Meas. Tech.* **8**, 3745–3765 (2015).
20. J. Liggio *et al*., *Nat. Commun.* **10**, 1863 (2019).
21. B. M. Erland *et al*., *Atmos. Meas. Tech.* **15**, 5841–5859 (2022).
22. A. D. Charpentier, J. A. Bergerson, H. L. MacLean, *Environ. Res. Lett.* **4**, 014005 (2009).
23. Alberta Energy Regulator, "Statistical Series 39 Alberta Mineable Oil Sands Plant Statistics Monthly Supplement" (2018); https://static.aer.ca/prd/documents/sts/ST39-2018.pdf.
24. J. Peischl *et al*., *J. Geophys. Res. Atmos.* **123**, 7725–7740 (2018).
25. Environment and Climate Change Canada, "National Pollutant Release Inventory" (2019); https://pollution-waste.canada.ca/national-release-inventory/archives/index.cfm?lang=en.
26. Government of Alberta, "Alberta Emissions Inventory Report (AEIR) Air Emission Rates" (2020); https://open.alberta.ca/opendata/aeirairemissionrates#summary.
27. I. B. Konovalov *et al*., *Atmos. Chem. Phys.* **16**, 13509–13540 (2016).
28. F. Liu *et al*., *Atmos. Chem. Phys.* **20**, 99–116 (2020).
29. S. N. Wren *et al*., *PNAS Nexus* **2**, pgad140 (2023).
30. Environment and Climate Change Canada, "Canada's Air Pollutant Emissions Inventory" (2018); https://data-donnees.ec.gc.ca/data/substances/monitor/canada-s-air-pollutant-emissions-inventory.
31. Alberta Energy Regulator, "ST53: Alberta In Situ Oil Sands Production Summary" (2018); https://www.aer.ca/providing-information/data-and-reports/statistical-reports/st53.
32. O. P. Strausz *et al*., *Energy Fuels* **24**, 5053–5072 (2010).
33. L. He, X. Li, G. Wu, F. Lin, H. Sui, *Energy Fuels* **27**, 4677–4683 (2013).
34. M. H. Erickson, M. Gueneron, B. T. Jobson, *Atmos. Meas. Tech.* **7**, 225–239 (2014).
35. J. C. Ditto *et al*., *Atmos. Chem. Phys.* **21**, 255–267 (2021).
36. California Air Resources Board, "2015 Estimated Annual Average Emissions Statewide" (2016); https://www.arb.ca.gov/app/emsinv/2017/emseic1p_query.php.
37. B. D. Drollette, D. R. Gentner, D. L. Plata, *Environ. Sci. Technol.* **54**, 9872–9881 (2020).
38. Alberta Energy Regulator, "State of Fluid Tailings Management for Mineable Oil Sands, 2020" (2021); https://static.aer.ca/prd/documents/reports/2020-State-Fluid-Tailings-Management-Mineable-OilSands.pdf.
39. Canadian Natural Upgrading, "Canadian Natural Muskeg River Mine Fluid Tailings Management Report" (Canadian Natural Upgrading, 2020); https://www.aer.ca/providing-information/by-topic/tailings.
40. Alberta Energy Regulator, "State of Fluid Tailings Management for Mineable Oil Sands, 2018" (2019); https://static.aer.ca/prd/documents/oilsands/2018-State-Fluid-Tailings-Management-Mineable-OilSands.pdf.
41. California Air Resources Board, "CEPAM2019v1.03 Emission Projection Data 2017: Estimated Annual Average Emissions Los Angeles County" (2017); https://ww2.arb.ca.gov/applications/emissions-county.
42. Environment and Climate Change Canada, "2018 National Pollutant Release Inventory (NPRI) Facility-Reported Data" (2018); https://www.canada.ca/en/services/environment/pollution-waste-management/national-pollutant-release-inventory.html.
43. G. Chow-Fraser, A. Rougeot, *50 Years of Sprawling Tailings: Mapping Decades of Destruction by Oil Sands Tailings* (CPAWS Northern Alberta, Environmental Defence Canada, 2022).
44. B. A. Nault *et al*., *Atmos. Chem. Phys.* **21**, 11201–11224 (2021).
45. A. Hodzic *et al*., *Atmos. Chem. Phys.* **10**, 5491–5514 (2010).

ACKNOWLEDGMENTS

The authors acknowledge the substantial technical and scientific contributions toward the success of this study from the AQRD technical and data teams, the NRC team, and the leadership of S. Cober. We thank A. Goldstein and R. Harley (UC Berkeley) for the use of the diesel fuel composition data used in Fig. 2D. **Funding:** J.C.D., M.H., L.G., J.M., T.N.H.-M., C.C., P.K., and D.R.G. acknowledge support from the National Science Foundation (AGS-1764126 and CBET-2011362) and GERSTEL for their collaboration with the thermal desorption unit used as part of this study, and M.H. also acknowledges the Goldwater Scholarship Foundation. S.-M.L. acknowledges the support of the Ministry of Science and Technology of China (grant 2019YFC0214700). The project was partially funded by Environment and Climate Change Canada's Climate Change and Air Pollutants Program. This work was also partially funded under the OSM program, and the results are independent of any position of the program. **Author contributions:** Conceptualization: D.R.G., J.L., and S.-M.L. Methodology: J.L., D.R.G., and S.-M.L. Investigation: All authors. Visualization: M.H., D.R.G., J.L., J.C.D., and L.G. Supervision: D.R.G., J.L., S-M.L., and J.C.D. Writing – original draft: M.H. Writing – review and editing: J.L., D.R.G., M.H., and all authors. **Competing interests:** The authors declare no competing interests in this study. **Data availability:** Flight campaign data are publicly available at https://donnees.ec.gc.ca/data/air/monitor/ambient-air-quality-oil-sands-region/pollutant-transformation-aircraft-based-multi-parameters-oil-sands-region/?lang=en. Other data and analysis relevant to the interpretation of the results can be found in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj6233
Materials and Methods
Supplementary Text
Figs. S1 to S15
Tables S1 to S10
References (*46*–*70*)

Submitted 13 July 2023; accepted 15 December 2023
10.1126/science.adj6233

ECOLOGY

Disruption of an ant-plant mutualism shapes interactions between lions and their primary prey

Douglas N. Kamaru[1,2], Todd M. Palmer[3], Corinna Riginos[1,4], Adam T. Ford[5], Jayne Belnap[6], Robert M. Chira[7], John M. Githaiga[7], Benard C. Gituku[2], Brandon R. Hays[8], Cyrus M. Kavwele[9,10], Alfred K. Kibungei[2], Clayton T. Lamb[5], Nelly J. Maiyo[2], Patrick D. Milligan[2,11], Samuel Mutisya[2], Caroline C. Ng'weno[12], Michael Ogutu[2], Alejandro G. Pietrek[13], Brendon T. Wildt[1], Jacob R. Goheen[1]*

Mutualisms often define ecosystems, but they are susceptible to human activities. Combining experiments, animal tracking, and mortality investigations, we show that the invasive big-headed ant (*Pheidole megacephala*) makes lions (*Panthera leo*) less effective at killing their primary prey, plains zebra (*Equus quagga*). Big-headed ants disrupted the mutualism between native ants (*Crematogaster* spp.) and the dominant whistling-thorn tree (*Vachellia drepanolobium*), rendering trees vulnerable to elephant (*Loxodonta africana*) browsing and resulting in landscapes with higher visibility. Although zebra kills were significantly less likely to occur in higher-visibility, invaded areas, lion numbers did not decline since the onset of the invasion, likely because of prey-switching to African buffalo (*Syncerus caffer*). We show that by controlling biophysical structure across landscapes, a tiny invader reconfigured predator-prey dynamics among iconic species.

Mutualisms are among the most widespread and economically important species interactions, creating and maintaining terrestrial, aquatic, and marine ecosystems (*1*, *2*). Because virtually every species on Earth participates in one or more mutualisms, their disruption can erode biodiversity through a combination of the direct loss of species, altered flows of mass and energy through ecological communities, and the inhibition of evolutionary trajectories (*3*). Although the loss of mutualisms is a global phenomenon (*3*), empirical studies linking mutualism disruption to broader community dynamics, particularly those across expansive areas, remain scarce.

The potential for mutualism disruption to reverberate across entire landscapes is especially strong when mutualism underpins the persistence of foundation species, i.e., spatially dominant and highly connected species within ecological networks that can amplify diversity and modulate critical ecosystem processes (*4*–*6*). Mutualisms involving these foundation species (or "foundational mutualisms") create and maintain habitats through biophysical structure [e.g., corals and their dinoflagellate associates (*7*), seagrasses and sulfide-oxidizing lucinid bivalves (*8*), and whistling-thorn trees (*Vachellia drepanolobium*) and their protective ant associates (*9*)]. As such, foundational mutualisms may modify species interactions through nontrophic pathways; for example, by generating refugia for competitors or prey species and cover or vantage points for predators. In the aftermath of disrupted foundational mutualisms, shifts in trophic dynamics may occur where biophysical structure shapes the frequency and outcomes of encounters among predators and their prey. Within such systems, spatially structured interactions, encompassing landscapes of fear (in which spatial variation in predation risk affects prey distributions) (*10*, *11*), predator-prey shell games (in which predators attempt to anticipate locations of prey, and prey respond by attempting to be spatially unpredictable) (*12*), and competition (*13*), should hinge on foundation species, and thus on foundational mutualisms.

Effects of ant invasion on defenses of a foundation tree

Across tens of thousands to hundreds of thousands of square kilometers in East Africa (*14*, *15*), the foundational whistling-thorn tree forms near-monocultures, comprising >70% (and often 98 to 99%) of woody stems where it occurs (*9*, *16*) (Figs. 1 and 2A). The whistling-thorn tree is a myrmecophyte, providing food (extrafloral nectar) and shelter (swollen-thorn domatia) in exchange for defense by a guild of native acacia ants (*Crematogaster* spp.) (*17*). Protection by acacia ants is particularly effective at deterring lethal herbivory by elephants (*Loxodonta africana*), thereby stabilizing savanna tree cover across entire landscapes (*9*). Over the past two decades, invasion of the big-headed ant (*Pheidole megacephala*), thou[illegible] to originate from an island in the Indian Oc[illegible], has disrupted this foundational mutualism in Laikipia, Kenya (*18*). Where big-headed ants encounter whistling-thorn trees, they numerically overwhelm and completely exterminate *Crematogaster* spp. ants, killing adult ants and consuming eggs, larvae, and pupae (*18*). However, big-headed ants do not protect whistling-thorn trees from herbivory, thus increasing the vulnerability of invaded trees to browsing by elephants. Consequently, in invaded areas, elephants browse and break trees at five to seven times the rate of that in uninvaded areas (*18*) (Figs. 1 and 2B).

We hypothesized that disruption of this foundational ant-tree mutualism would affect interactions between lions (*Panthera leo*) and their most common prey, plains zebra (hereafter referred to as "zebra"; *Equus quagga*). Zebra are unselective grazers (*19*) that require large volumes of grass to meet their nutritional needs, and they comprise around 50% of wild ungulates killed by lions on Ol Pejeta Conservancy in Laikipia (figs. S1 and S2). We tested two predictions regarding lion-zebra dynamics and mutualism disruption by means of big-headed ant-invasion (Fig. 1): (i) Big-headed ant invasion increases browsing by elephants, thereby generating greater visibility or "openness" relative to uninvaded areas, and (ii) greater visibility, mediated by big-headed ant invasion, shapes interactions between lions and zebra through some combination of increased selection for visibility by zebra (if zebra choose habitats on the basis of perceived safety) (*20*), avoidance of increased visibility by lions (if lions choose habitats on the basis of prey accessibility) (*21*), or a reduction in the hiding cover necessary for lions to hunt successfully (*22*). Additionally, we sought to quantify whether and how any changes in the catchability of zebra triggered by big-headed ant invasion manifested as changes in lion population size through time.

Effects of mutualism disruption on savanna openness

To test our first prediction, we measured differences in visibility across a 364-km^2 landscape that varied in both tree cover and in the occurrence of big-headed ants. We measured visibility associated within three blocks of four replicated 2500-m^2 plots (fig. S1A). For each replicate block, a pair of plots was established on each side of a big-headed ant invasion front; one plot in each pair experimentally excluded "megaherbivores" [elephants, giraffes (*Giraffa camelopardalis*), and rhinoceros (*Diceros bicornis* and *Cerototherum simum*)] with electrified fencing. Big-headed ant invasion fronts advance ~50 m per year (*23*). Therefore, we established "uninvaded" plots 0.5 to 2.5 km in front of invasion fronts (fig. S1, A and B) to ensure that uninvaded plots would not be

[1]Department of Zoology & Physiology and Program in Ecology & Evolution, University of Wyoming, Laramie, WY, USA. [2]Ol Pejeta Conservancy, Private Bag 10400, Nanyuki, Kenya. [3]Department of Biology, University of Florida, Gainesville, FL, USA. [4]The Nature Conservancy, Lander, WY, USA. [5]Department of Biology, University of British Columbia, Kelowna, BC, Canada. [6]Southwest Biological Science Center, US Geological Survey, Moab, UT, USA. [7]School of Biological Sciences, University of Nairobi, Nairobi, Kenya. [8]Nicholas School of the Environment, Duke University, Durham, NC, USA. [9]School of Mathematics & Statistics and School of Biodiversity, University of Glasgow, Glasgow, Scotland. [10]School of Natural Resources, Karatina University, Nyeri, Kenya. [11]Department of Biology, University of Nevada, Reno, NV, USA. [12]Lewa Wildlife Conservancy, Private Bag 60300, Isiolo, Kenya. [13]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Salta, Argentina.
*Corresponding author. Email: jgoheen@uwyo.edu

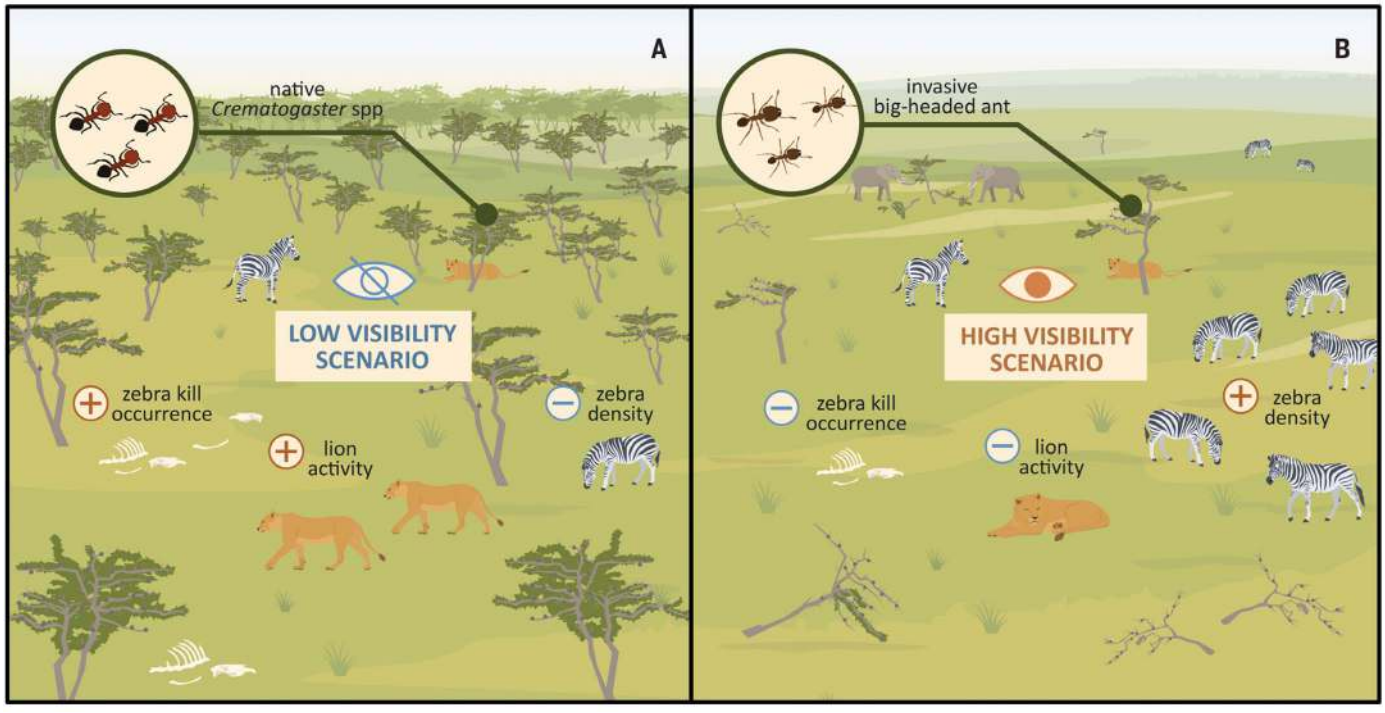


Fig. 1. Illustrated predictions by which disruption of the foundational ant-plant mutualism shapes spatial patterns of lion predation. (**A**) In uninvaded whistling-thorn tree savanna, native acacia ants defend whistling-thorn trees against browsing by elephants, such that tree density is high and visibility is low. In turn, lower visibility is predicted to be associated with zebra kills through some combination of reduced zebra density (if lower densities increase risk of predation via delayed detection of lions), increased lion activity (if lions are more active in denser stands of whistling-thorn trees), and increased hunting success of lions (if hunting success is predicated on the hiding cover afforded by whistling-thorn trees). (**B**) In invaded whistling-thorn tree savanna, big-headed ants kill acacia ants, rendering trees vulnerable to browsing by elephants and resulting in higher visibility. Higher visibility is predicted to be associated with reduced occurrence of zebra kills through some combination of increased zebra density (if lower densities increase risk of predation through delayed detection of lions), reduced lion activity (if lions are more active in denser stands of whistling-thorn trees), and reduced hunting success of lions (if hunting success is predicated on the hiding cover afforded by whistling-thorn trees).

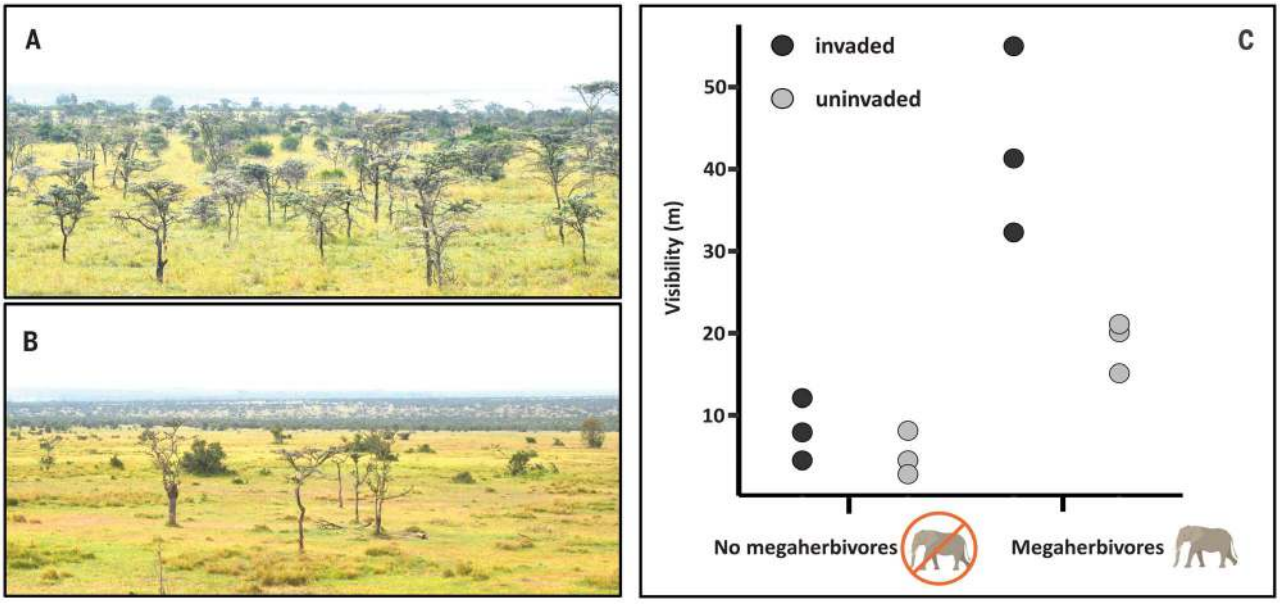


Fig. 2. Big-headed ant invasion increases visibility. (**A**) An uninvaded whistling-thorn tree savanna. (**B**) An invaded landscape in which elephants have browsed, broken, and killed whistling-thorn trees. (**C**) After a 3-year period in open (unfenced) plots accessible to megaherbivores, visibility was 2.67 times higher in plots invaded by big-headed ants relative to uninvaded plots accessible to megaherbivores (two-way ANOVA invasion-megaherbivore interaction: $F_{1,8} = 8.14$, $P = 0.02$).

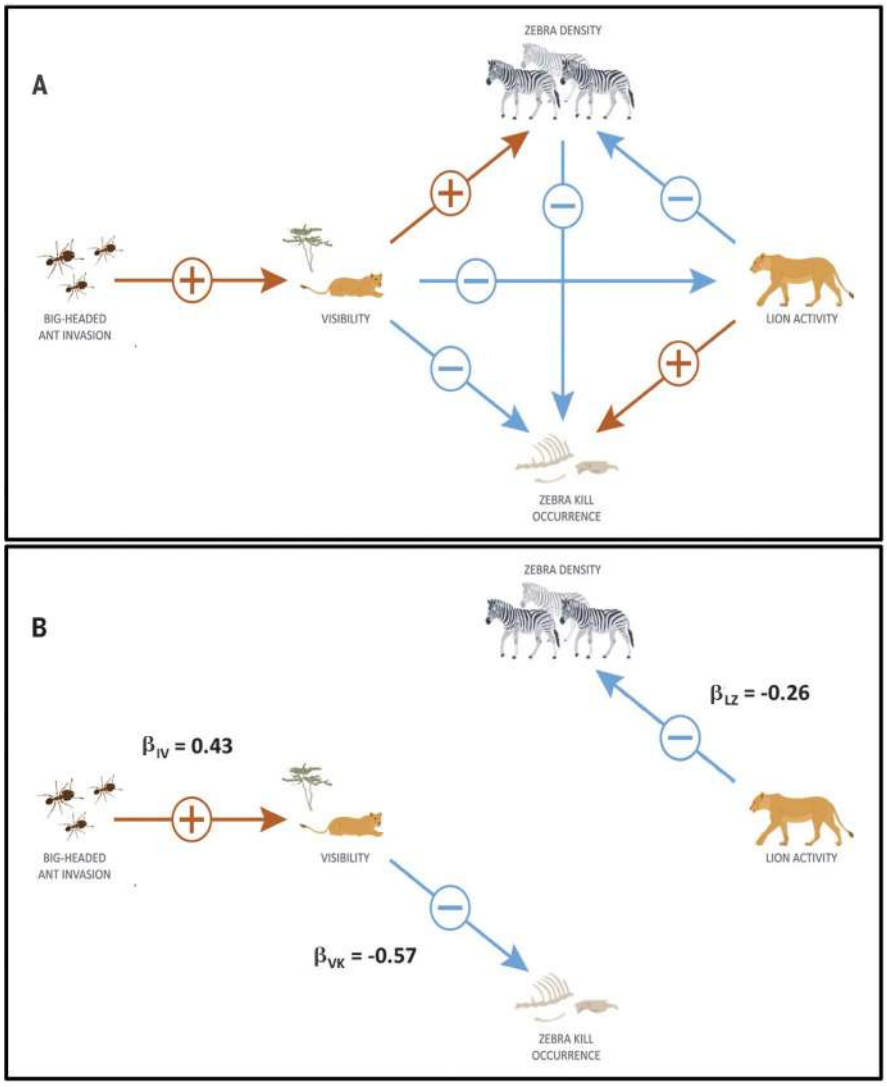


Fig. 3. Zebra kill occurrence is reduced in areas invaded by big-headed ants because of heightened visibility stemming from disruption of the foundational mutualism. (**A**) The full model depicting hypothesized relationships among big-headed ant invasion, visibility, zebra density, lion activity, and zebra kill occurrence. Orange and blue arrows represent hypothesized positive and negative effects among variables, respectively. Within this full model, 21 paths were nested and evaluated using *d*-separation and subsequent model selection (figs. S8 and S9 and tables S8 and S9). (**B**) The best-supported nested path model (nested path model 17) is illustrated. This model was statistically indistinguishable (i.e., within 2 AIC_C units) from nested path models 19 and 9, each of which contained an additional linkage from zebra density to zebra kill occurrence (tables S8 and S9). In both cases, and contrary to our inferred hypothesis, the path coefficient for zebra density to zebra kill occurrence was positive, although its 95% confidence limits for both nested path models 19 and 9 encompassed zero. Similarly, nested path model 9 contained an additional linkage from lion activity to zebra kill occurrence, but the 95% confidence limits on this path coefficient encompassed zero. Confidence limits for path coefficients for each linkage in nested path model 17 did not encompass zero (big-headed ant invasion and visibility: unstandardized β_{IV} = 13.45 ± 2.75 SE; visibility and zebra kill occurrence: unstandardized β_{VK} = −0.09 ± 0.02 SE; lion activity and zebra density: unstandardized β_{LZ} = −1.07 ± 0.39 SE). Standardized β coefficients are reported next to each arrow.

invaded during our 3-year study. Over the course of our experiment, changes in tree cover could arise in two ways. First, and within megaherbivore exclusion (fenced) plots, any differences relative to baseline conditions (i.e., those from uninvaded, open plots in 2017) would reflect differences in tree growth and survival in the absence of megaherbivores for both invaded and uninvaded areas (*24*). Second, within open (unfenced) plots, megaherbivore browsing (which we predicted would be higher in invaded areas) would reduce tree growth and survival (*24*). We expected both processes in tandem to result in the highest visibility within invaded, open plots and the lowest visibility within megaherbivore exclusion plots (regardless of invasion status), with intermediate visibility within uninvaded, open plots. We attributed differences in visibility caused by megaherbivore exclusion and big-headed ant invasion to elephant browsing (as opposed to other megaherbivores) for the following reasons: (i) Acacia ants are especially effective at deterring browsing by elephants, such that browsing rates increase by nearly an order of magnitude on branches from which acacia ants have been experimentally removed (*9*); (ii) elephants are singular in their ability to promote visibility by breaking and knocking over adult trees; and (ii) elephants comprise around 60 to 70% of the biomass density of megaherbivores at Ol Pejeta Conservancy and in Laikipia, more generally (*25*). In open plots, big-headed ant invasion was associated with 2.67 times higher visibility after 3 years (uninvaded, open mean = 18.06 m ± 3.00 SE; invaded, open mean = 48.17 m ± 6.80 SE) (Fig. 2C). By contrast, visibility did not differ as a function of big-headed ant invasion for megaherbivore exclusion plots (Fig. 2C). Relative to uninvaded, open plots (reflecting baseline conditions of a "natural" savanna), changes in visibility were driven by a combination of greater growth and survival of trees within megaherbivore exclusion plots, and reduced growth and survival of trees in invaded, open plots (fig. S3 and tables S1 and S2). Thus, big-headed ant invasion rendered whistling-thorn trees largely defenseless against elephants, leading to higher browsing rates and more open landscapes characterized by higher visibility (Fig. 2).

Savanna openness and predation risk to zebra

Testing our second prediction entailed quantifying zebra density, lion activity, big-headed ant occurrence, and visibility at zebra kills. To quantify zebra density, we built time-varying, spatially explicit density surfaces from resource selection functions to measure habitat selection and population density of zebra (*26*). We incorporated three habitat features into these resource selection functions: glades (nutrient-rich lawns arising from old livestock corrals), water sources, and human settlements (in the event that zebra selected for such settlements as protection against lions; i.e., the "human shield" hypothesis) (*27*), in addition to a distance to glade–distance to water source interaction. To quantify lion activity, we captured and fit GPS collars to six lionesses from distinct prides representing approximately 50 adult individuals and 30 cubs (or 95% of the lions at our study site) (*26*). In Laikipia, lions form cohesive groups, such that movements of a single lioness are representative of pride-level movements (*28*, *29*). Estimates of zebra density and

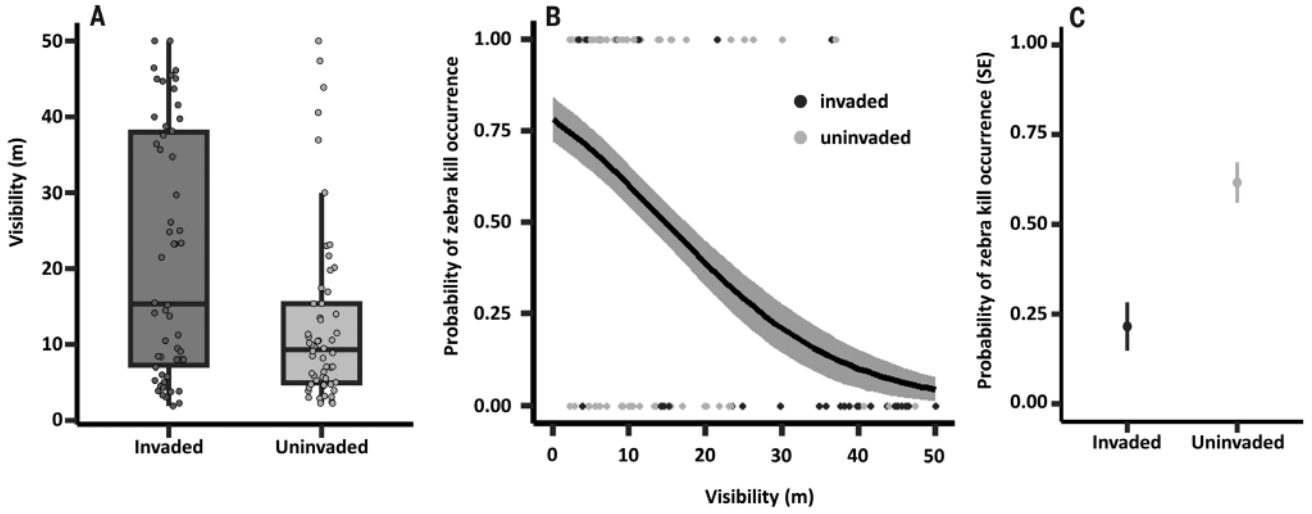


Fig. 4. Pairwise relationships underlying zebra kill occurrence from the best-supported nested path model. (**A**) Visibility and big-headed ant invasion, (**B**) probability of zebra kill occurrence and visibility, and (**C**) predicted probability of zebra kill occurrence by invasion status. The shading in (B) indicates 95% confidence limits. Visibility was higher in areas invaded by big-headed ants, and probability of zebra kill occurrence was lower where visibility was higher. At the median level of visibility for invaded (29.69 m) versus uninvaded (9.31 m) whistling-thorn tree savanna, the probability of zebra kill occurrence was 2.87 times higher in uninvaded than in invaded savanna (0.62 ± 0.06 SE versus 0.22 ± 0.07 SE).

lion activity were then incorporated with visibility as predictors of kill occurrence. Within nested path models, visibility was modeled as an outcome of big-headed ant invasion and as a predictor of kill occurrence, zebra density, and lion activity (Fig. 3A).

To quantify big-headed ant occurrence and visibility at zebra kills, we first assessed spatial variation in the distribution of lion-killed zebra using a clustering algorithm from GPS-collared lions (*26*). Because lions at our study site and elsewhere in East Africa ambush their prey (as opposed to chasing them over long distances) (*21*, *22*), the locations of kills are reasonable proxies for locations of successful hunts (*29*). Within whistling-thorn tree savanna, we identified 55 zebra kill sites by investigating GPS "clusters," which are defined as ≥2 successive GPS relocations occurring within 100 m of each other (given that pride members feed together at kills, we were able to identify kill sites from the movement patterns of telemetered individuals, even if they did not make the kill themselves) (*29*). Within 5 to 10 days, we visually confirmed the prey species at each GPS cluster (*26*). At each zebra kill site, we collected data on visibility and the presence of big-headed ants (*26*), as well as estimated zebra density (from resource selection function-derived density estimates; fig. S4) and lion activity (from telemetry-derived utilization distributions; fig. S6). Across the broader landscape, patterns of visibility mirrored our experimental results, such that the visibility of invaded locations was 13.45 m ± 2.75 SE greater than that of uninvaded locations (Fig. 4A).

We expected zebra to aggregate in high-visibility areas, including areas invaded by big-headed ants and other openings in the tree layer. Further, lion activity is known to be correlated with tree cover (or visibility) in our study system and elsewhere (*21*, *22*, *29*). Thus, several potential predictors (big-headed ant invasion, visibility, zebra density, and lion activity) of predation risk to zebra were correlated, making it challenging to test whether big-headed ants were associated with safety for zebra. Similarly, both density of whistling-thorn trees (and thus, visibility) and zebra density can vary because of a whole host of environmental variables (*30*–*32*), making it difficult to attribute changes in visibility solely to big-headed ants.

To test whether and how big-headed ant invasion shifted spatial variation in lion predation of zebra, we conducted nested path analysis (*33*, *34*). We tested 21 path models nested within our full model (Fig. 3A), representing a series of ecological linkages by which increased visibility could shift spatial variation in predation risk to zebra (figs. S8 and S9 and table S8). Nested path models represented different combinations of relationships among big-headed ant invasion, visibility, zebra density, and lion activity, as well as the influence of such relationships on zebra kill occurrence. We did not attempt to formulate and test an exhaustive set of nested path models within the full model; rather, our nested path models were based on our inferred understanding of this ecosystem. Our nested path analysis provided a test of our second prediction: that greater visibility, mediated by big-headed ant invasion, shapes the spatial distribution of lion-killed zebra.

Of the 21 path models nested within our full model, Fisher's C (a combined test of conditional independence among linkages in nested path models) (*33*) revealed that 14 were statistically viable, with correlation structures that represented the observed data (fig. S8 and table S8). The correlation structure proposed through seven nested path models (models 2, 5, 6, 8, 13, 14, and 16) differed from the observed data; each of these models included a linkage from visibility to zebra density, did not include a linkage from big-headed ant invasion to visibility, or both (fig. S8 and table S8). Through a model selection procedure on the remaining 15 path models (14 nested path models, plus the full model), there was statistical support for the hypothesis that big-headed ant invasion reduced the occurrence of zebra kills by increasing the visibility of lions to their prey (nested path models 17, 19, and 9) (Fig. 3 and tables S8 and S9) (*26*). Each of these three nested path models included linkages from big-headed ant invasion to visibility (Fig. 4A), from visibility to zebra kill occurrence (Fig. 4B), and from lion activity to zebra density, with path coefficients whose 95% confidence limits did not encompass zero for each linkage (Fig. 3B and table S9). Consequently, zebra kill occurrence was 2.87 times higher in uninvaded areas relative to areas invaded by big-headed ants (Fig. 4C). Nested path models 19 and 9 also included a linkage from zebra density to zebra kill occurrence, and nested path model 9 included a linkage from lion activity to zebra kill occurrence, but the 95% confidence limits for path coefficients from these linkages encompassed zero (Fig. 3B and table S9). Of the 10 remaining viable nested path models, those that did not include a linkage from big-headed ant invasion to visibility were >19 Akaike's information

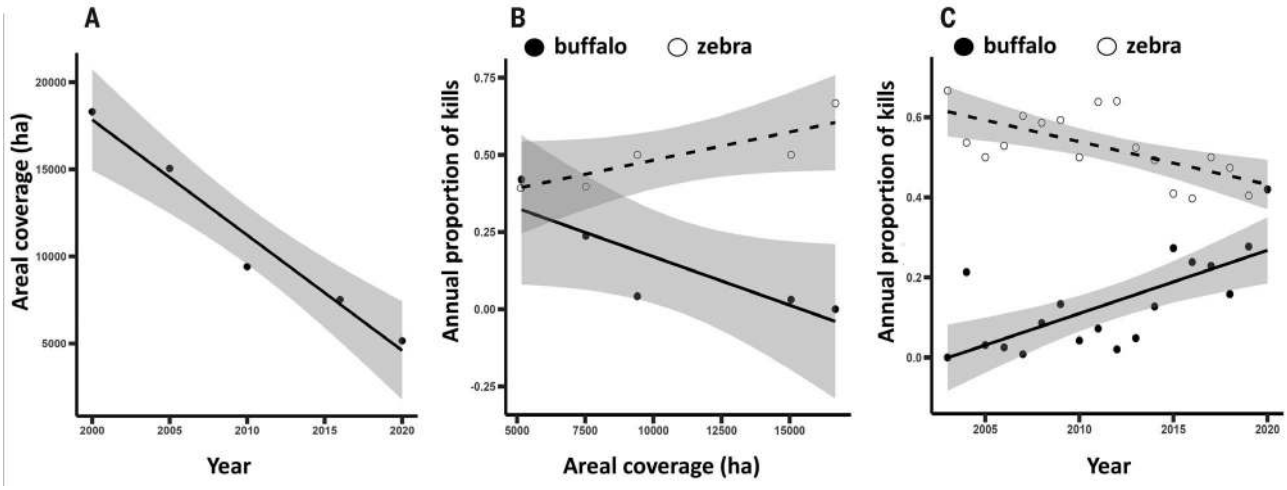


Fig. 5. Over 18 years, lion diets shifted toward buffalo as whistling-thorn tree cover declined. (**A**) Since the onset of big-headed ant invasion, areal coverage of habitats classified as whistling-thorn monoculture (in which >98% of stems were whistling-thorn trees) or whistling-thorn dominant (in which whistling-thorn trees were the most common woody species) (*26*) declined through time (R^2 = 0.86, P < 0.01). (**B**) The annual proportion of kills made by lions that were zebra tended to increase with increasing whistling-thorn tree cover in each of 5 years (2000, 2005, 2010, 2016, and 2020; R^2 = 0.72, P = 0.07), and the annual proportion of kills made by lions that were buffalo tended to decline with increasing whistling-thorn tree cover in the same 5 years (R^2 = 0.74, P = 0.06). For each relationship, areal coverage of whistling-thorn trees in 2003 was calculated as the midpoint of whistling-thorn tree coverage in 2000 and in 2005. (**C**) From 2003 to 2020, the annual proportion of kills made by lions that were zebra declined from 67 to 42% (R^2 = 0.42, P < 0.01), whereas the annual proportion of kills made by lions that were buffalo increased from 0 to 42% (R^2 = 0.47, P < 0.001). The shading on all panels represents 95% confidence limits. Kill proportions were calculated from kills discovered opportunistically by antipoaching patrols at Ol Pejeta Conservancy from 2003 to 2020 (fig. S2).

criterion (AIC) points higher than the best-supported model (table S8).

Can lions compensate for less-catchable zebra?

Invasion by big-headed ants is ongoing in Laikipia. Big-headed ants continue to expand at approximately 50 m per year (*23*), a rate comparable to that of other invasive ants (*35*). Although mutualism disruption shaped the spatial distribution of lion predation, we cannot know the extent to which lions will respond as big-headed ants continue to expand across this region. We consider two scenarios by which lions might compensate in a landscape with increasingly less-catchable zebra.

First, and by promoting visibility, big-headed ants may reduce the catchability of zebra in invaded areas, but not reduce the overall rate at which zebra are killed. This scenario requires that lions concentrate their hunting activity in uninvaded portions of their home ranges, an expectation that is not supported by our data. Mean lion activity was statistically indistinguishable between invaded and uninvaded areas (fig. S10). Further, the number of adult and subadult lions at Ol Pejeta Conservancy (53.31 ± 2.64 SE) has remained remarkably stable for the 13 years that it has been monitored (*26*), implying that this population is at or near carrying capacity (consistent with results from other fenced reserves across sub-Saharan Africa) (*36*). Taken together, this strongly suggests that lions are persisting despite losing habitat in which they can efficiently kill zebra.

The second scenario entails a functional response by lions, through which big-headed ant invasion leads to prey switching toward more catchable or energetically rewarding prey (*37*). Given that the distribution of big-headed ants is not at equilibrium on Ol Pejeta Conservancy, the progression of time since their introduction to Laikipia (estimated to be in the early 2000s) (*18*) should be a reasonable proxy for the area occupied and any corresponding impacts on zebra catchability. From 2000 to 2020, areal coverage of whistling-thorn trees declined (Fig. 5A), and whistling-thorn cover was marginally correlated with the proportion of lion kills that were zebra versus buffalo (*Syncerus caffer*, the second most commonly killed wild ungulate) (Fig. 5B) (*26*). From 2003 to 2020, the proportion of kills made by lions that were zebra declined from 67 to 42%, whereas the proportion of kills that were buffalo increased from 0 to 42% (Fig. 5C) (*26*). There were no directional changes in zebra or buffalo densities from 2014 to 2020 (no data were available on zebra or buffalo densities prior to 2014) (fig. S11).

Redirected trophic flows following mutualism disruption

Our results show that interactions between lions and their primary prey, the plains zebra, are mediated by a foundational ant-plant mutualism. Lions and other large carnivores use tree cover to conceal themselves, such that their success in hunting plains zebra was higher where visibility was lower (*11*, *21*). By disrupting the foundational mutualism between whistling-thorn trees and native acacia ants, invasion by big-headed ants renders trees more vulnerable to browsing by elephants, thereby reducing tree cover and increasing visibility. Contrary to our expectation, we found no evidence that higher visibility triggered by big-headed ant invasion changed zebra density, and zebra density itself was a weak predictor of zebra kill occurrence. Similarly, there was no evidence for a linkage between increased visibility and lion activity. Instead, big-headed ant invasion reduced the occurrence of zebra kills by increasing openness across the landscape, thereby limiting the frequency with which lions killed zebra.

Confronted with declining numbers or catchability of preferred prey, prey switching has long been recognized as theoretical basis for stabilizing populations (*38*). Yet, empirical evidence for prey switching among large mammals is scant, perhaps because prey are not uniformly vulnerable to predation (*39*). Dangerous prey tend to be avoided by predators, even when populations of preferred prey decline (*40*). Elsewhere in East Africa, larger groups (i.e., subsets of prides involved in hunts) of lions are required to kill buffalo, and male lions are significantly more likely to participate in buffalo kills than those of zebra (*41*, *42*), although large groups of lions still prefer zebra when zebra are abundant (perhaps to reduce injuries during hunts) (*42*, *43*). Although the invasion of big-headed ants has shaped the spatial distribution of zebra kills, and the frequency of zebra kills has declined over time, prey switching by lions to more formidable prey seems to

have (thus far) prevented any cascading effects on lion numbers. The role of behavioral adjustments (i.e., size and composition of hunting groups) in underlying the population stability of lions, plus the degree to which such stability can be maintained as big-headed ants advance across the landscape, remain open questions for future investigation.

Foundational mutualisms structure some of the most iconic environments on Earth, including coral reefs, kelp forests, and, as evidenced in this work, African savannas (*7*–*9*, *44*, *45*). When such mutualisms are disrupted, their effects can reverberate across landscapes, to the detriment of some species and to the benefit of others. We show that the spread of the big-headed ant, one of the globe's most widespread and ecologically impactful invaders (*35*, *46*), has sparked an ecological chain reaction that reduces the success by which lions can hunt their primary prey. The disruption of foundational mutualisms could be an underappreciated contributor to predator-prey dynamics and trophic restructuring of the world's ecosystems.

REFERENCES AND NOTES

1. J. L. Bronstein, *J. Ecol.* **97**, 1160–1170 (2009).
2. A. Traveset, D. M. Richardson, *Annu. Rev. Ecol. Evol. Syst.* **45**, 89–113 (2014).
3. E. Toby Kiers, T. M. Palmer, A. R. Ives, J. F. Bruno, J. L. Bronstein, *Ecol. Lett.* **13**, 1459–1474 (2010).
4. P. K. Dayton, in *Proceedings of the colloquium on conservation problems in Antarctica*, B. C. Parker, ed. (Allen Press, 1972), pp. 356.
5. J. F. Bruno, M. D. Bertness, *Marine Community Ecology* (Sinauer, 2001), p. 550.
6. A. M. Ellison, *iScience* **13**, 254–268 (2019).
7. L. Muscatine, R. R. Pool, R. K. Trench, *Trans. Am. Microsc. Soc.* **94**, 450–469 (1975).
8. T. van der Heide *et al*., *Science* **336**, 1432–1434 (2012).
9. J. R. Goheen, T. M. Palmer, *Curr. Biol.* **20**, 1768–1772 (2010).
10. J. W. Laundré, L. Hernández, K. B. Altendorf, *Can. J. Zool.* **79**, 1401–1409 (2001).
11. A. T. Ford *et al*., *Science* **346**, 346–349 (2014).
12. W. A. Mitchell, S. L. Lima, *Oikos* **99**, 249–259 (2002).
13. D. Tilman, P. Kareiva, *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions* (Princeton University Press, 1997), p. 368.
14. I. R. Dale, P. J. Greenway, *Kenya Trees and Shrubs* (University Press, 1961), p. 654.
15. J. Deckers, O. Spaargaren, F. Nachtergaele, *The sustainable management of vertisols*, J. K. Syers, F. W. T. Penning de Vries, P. Nyamudeza, Eds. (CABI, 2001), p. 304.
16. T. P. Young, C. H. Stubblefield, L. A. Isbell, *Oecologia* **109**, 98–107 (1996).
17. T. M. Palmer, A. K. Brody, *Ecology* **94**, 683–691 (2013).
18. C. Riginos, M. A. Karande, D. I. Rubenstein, T. M. Palmer, *Ecology* **96**, 654–661 (2015).
19. T. R. Kartzinel *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8019–8024 (2015).
20. C. Riginos, *J. Anim. Ecol.* **84**, 124–133 (2015).
21. J. G. C. Hopcraft, A. R. E. Sinclair, C. Packer, *J. Anim. Ecol.* **74**, 559–566 (2005).
22. A. Chen, L. Reperant, I. R. Fischhoff, D. I. Rubenstein, *Clim. Change Ecol.* **1**, 100001 (2021).
23. A. G. Pietrek, J. R. Goheen, C. Riginos, N. J. Maiyo, T. M. Palmer, *Oecologia* **195**, 667–676 (2021).
24. B. R. Hays *et al*., *Ecology* **103**, e3655 (2022).
25. R. D. Crego *et al*., *Biol. Conserv.* **242**, 108436 (2020).
26. See Supplementary Materials and Methods.
27. J. Berger, *Biol. Lett.* **3**, 620–623 (2007).
28. A. Oriol-Cotterill, D. W. MacDonald, M. Valeix, S. Ekwanga, L. G. Frank, *Anim. Behav.* **101**, 27–39 (2015).
29. C. C. Ng'weno, A. T. Ford, A. K. Kibungei, J. R. Goheen, *Ecology* **100**, e02698 (2019).
30. N. J. Georgiadis, M. Hack, K. Turpin, *J. Appl. Ecol.* **40**, 125–136 (2003).
31. J. E. Maclean, J. R. Goheen, D. F. Doak, T. M. Palmer, T. P. Young, *Ecology* **92**, 1626–1636 (2011).
32. W. O. Odadi, M. K. Karachi, S. A. Abdulrazak, T. P. Young, *Science* **333**, 1753–1755 (2011).
33. B. Shipley, *Ecology* **90**, 363–368 (2009).
34. R. Serrouya *et al*., *Proc. Biol. Sci.* **288**, 20202811 (2021).
35. D. A. Holway, L. Lach, A. V. Suarez, N. D. Tsutsui, T. J. Case, *Annu. Rev. Ecol. Syst.* **33**, 181–233 (2002).
36. C. Packer *et al*., *Ecol. Lett.* **16**, 635–641 (2013).
37. C. M. Prokopenko, T. Avgar, A. Ford, E. Vander Wal, *Ecology* **104**, e3928 (2023).
38. W. W. Murdoch, *Ecol. Monogr.* **39**, 335–354 (1969).
39. S. Mukherjee, M. R. Heithaus, *Biol. Rev. Camb. Philos. Soc.* **88**, 550–563 (2013).
40. A. Tallian *et al*., *Funct. Ecol.* **31**, 1418–1429 (2017).
41. D. Scheel, *Behav. Ecol.* **4**, 90–97 (1993).
42. C. Packer, *The Lion* (Princeton University Press, 2023), p. 356.
43. B. Van Valkenburgh, P. A. White, *PeerJ* **9**, e11313 (2021).
44. V. Tunnicliffe, *Am. Sci.* **80**, 336–349 (1992).
45. E. J. Carpenter *et al*., *Mar. Ecol. Prog. Ser.* **185**, 273–283 (1999).
46. S. Lowe, M. Browne, S. Boudjelas, M. De Poorter, *100 of the World's Worst Invasive Alien Species*, Invasive Species Specialist Group (Hollands Printing Ltd., 2000), pp. 12.

ACKNOWLEDGMENTS

We thank G. Busienei, S. Carpenter, M. Dyck, J. Ekedeli, S. Musila, S. Ngulu, K. Steinfield, D. Atkins, T. Avgar, K. Bandyopadhyay, J. Dolphin, K. Garrett, A. Helman, M. Kauffman, L. Khasoha, D. Laughlin, J. Merkle, F. Molina, D. Ngatia, R. Serrouya, S. Seville, B. Shipley, C. Tarwater, Fuse Consulting, and the Kenya Wildlife Service. Our work was conducted under the permission of the Kenyan National Commission for Science, Technology, and Innovation (NACOSTI/P/18/36141/25399) and with the permission of the Kenya Wildlife Service. **Funding:** This research was financially supported by grants from the US National Science Foundation (NSF DEB 1556905) to T.M.P., C.R., and J.R.G.; the Wyoming NASA Space Grant Consortium to J.R.G.; the American Society of Mammalogists African Research Fellowship to D.N.K.; the Rufford Foundation to D.N.K.; the University of Wyoming's Biodiversity Institute to D.N.K.; the University of Wyoming's College of Agriculture, Life Sciences, and Natural Resources Global Perspectives Grant Program to J.R.G.; the University of Wyoming's Global Engagement Office International Research Grant to J.R.G.; and the University of Wyoming's Department of Zoology & Physiology. **Author contributions:** D.N.K., T.M.P., C.R., R.M.C., J.M.G., and J.R.G. conceived of the study. D.N.K., T.M.P., C.R., B.C.G., P.D.M., S.M., C.C.N., M.O., A.G.P., and J.R.G. collected the data. D.N.K., C.R., A.K.K., C.C.N., B.T.W., and J.R.G. analyzed the data. C.T.L. wrote code for *d*-separation tests and nested path analysis. D.N.K., T.M.P., C.R., A.T.F., and J.R.G. wrote the manuscript. All authors assisted with edits and revisions. **Competing interests:** The authors declare no completing interests. **Data and materials availability:** The data reported in this paper and code used are available publicly at datadryad.org (https://doi.org/10.5061/dryad.np5hqbzzq). For review purposes, data are available privately at https://datadryad.org/stash/share/Ar6RSt6vT_klubm6yxNQQEG272CSrLner5HtM7ZycOY.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg1464
Materials and Methods
Figs. S1 to S11
Tables S1 to S9
References (*47*–*70*)
MDAR Reproducibility Checklist
Data S1 and S2
Code S1 and S2

Submitted 5 December 2022; accepted 1 December 2023
10.1126/science.adg1464

BIOCATALYSIS

Directed evolution of enzymatic silicon-carbon bond cleavage in siloxanes

Nicholas S. Sarai[1]†‡, Tyler J. Fulton[1]†, Ryen L. O'Meara[1]†, Kadina E. Johnston[2]§, Sabine Brinkmann-Chen[1], Ryan R. Maar[3], Ron E. Tecklenburg[3], John M. Roberts[3], Jordan C. T. Reddel[3], Dimitris E. Katsoulis[4]*, Frances H. Arnold[1]*

Volatile methylsiloxanes (VMS) are man-made, nonbiodegradable chemicals produced at a megaton-per-year scale, which leads to concern over their potential for environmental persistence, long-range transport, and bioaccumulation. We used directed evolution to engineer a variant of bacterial cytochrome $P450_{BM3}$ to break silicon-carbon bonds in linear and cyclic VMS. To accomplish silicon-carbon bond cleavage, the enzyme catalyzes two tandem oxidations of a siloxane methyl group, which is followed by putative [1,2]-Brook rearrangement and hydrolysis. Discovery of this so-called siloxane oxidase opens possibilities for the eventual biodegradation of VMS.

Linear and cyclic volatile methylsiloxanes (VMS) are anthropogenic compounds with material properties—such as high backbone flexibility and low surface tension—that make them useful in many consumer applications, from detergents and antifoaming agents to lotions, shampoos, and hair conditioners (*1*–*4*) (Fig. 1). Cyclic VMS are also important feedstocks for the synthesis of silicone polymers (*5*). To satisfy consumer and feedstock demand for siloxanes, production of VMS is on the order of megatons per year. However, the societal benefits of VMS must be balanced with their potential for environmental contamination, bioaccumulation, and toxicity (*6*–*12*). Regulations on VMS vary in different regions of the world (*13*). For example, octamethylcyclotetrasiloxane (**3**) is designated as a substance of very high concern (SVHC) by the European Chemicals Agency on the basis of persistence, bioaccumulation, and suspected reproductive toxicity (*14*, *15*). Given the prevalence, utility, and potential concerns of VMS, their degradation through Si–C bond cleavage is of substantial interest.

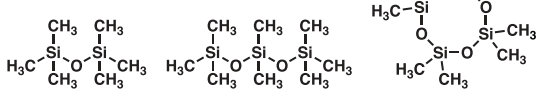
Degradation of VMS is nontrivial owing to their high thermal stability and lack of functional group handles. Hydrolysis of the Si–O bonds merely leads to speciation, producing silanols and siloxanediols, whereas complete degradation requires cleavage of the relatively inert Si–C bonds. Chemical means to accomplish VMS degradation are limited to a few examples, including TiO_2 photocatalysis, pyrolysis, and atmospheric oxidation by hydroxyl radicals (*4*, *16*, *17*). Generally, these Si–C bond cleavage reactions are initiated by one or more oxidations of the siloxane methyl group. Studies in higher organisms have found that VMS are metabolized to a bevy of products, including metabolites indicative of Si–C cleavage, which is proposed to occur after a C–H hydroxylation event (*18*–*20*). Enzymatic oxidation could unlock a mechanism for Si–C bond cleavage that is adaptable to a variety of environmental and process conditions (*21*–*23*). However, no enzyme capable of hydroxylating the siloxane C–H bond nor of facilitating breakage of a Si–C bond has been identified (*4*, *24*, *25*). We hypothesized that cytochromes P450, which can oxidize unactivated alkyl C–H bonds, might also be able to oxidize the similarly strong siloxane C–H bonds (*26*–*30*). Thus, we explored the potential of these enzymes as a starting point for enzymatic Si–C bond cleavage.

Discovery and directed evolution of siloxane Si–C bond cleavage activity

Cytochrome $P450_{BM3}$ is a self-sufficient enzyme made up of a heme domain fused to a reduced nicotinamide adenine dinucleotide phosphate (NADPH)–dependent reductase domain that contains flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) prosthetic groups. This soluble bacterial enzyme has been engineered to catalyze myriad non-native hyd[illegible] ylations (*29*, *31*). To test the hypothesis [illegible] enzymatic C–H hydroxylation would facilitate Si–C bond cleavage, we evaluated a panel of cytochrome $P450_{BM3}$ variants for their ability to oxidize the siloxane C–H bond in hexamethyldisiloxane (**1**) (table S9). Gas chromatography–mass spectrometry (GC-MS) analysis of enzymatic reactions in *Escherichia coli* lysate revealed detectable quantities of C–H hydroxylation product **4** and quantifiable production of Si–C bond cleavage and hydrolysis product **5** with several variants from our collection of cytochromes P450 that had been evolved in the laboratory for other functions. Of these, a previously unpublished $P450_{BM3}$ variant evolved for silane and siloxane Si–H hydroxylation designated *L*SilOx1 (linear siloxane oxidase, generation 1) was chosen as a starting point for evolution of the Si–C bond cleavage activity (Fig. 2A) (*32*). *L*SilOx1 has 13 amino acid substitutions with respect to wild-type $P450_{BM3}$, which has no activity for siloxane hydroxylation or Si–C bond cleavage (table S9).

The Si–C bond cleavage activity was enhanced over several rounds of directed evolution in *E. coli* lysate, where activity was defined as the ratio of silanol **5** concentration to enzyme concentration (Fig. 2A). Directed evolution with siloxane **1** was particularly challenging because of its volatility, low aqueous solubility, and incompatibility with the 96-well polypropylene plates that are typically used for biocatalytic reactions. This inspired the adoption of a centrifuge-compatible 96-well plate composed of individual glass shells for subsequent evolution campaigns with siloxane substrates (supplementary materials). To expand the scope of enzymatic VMS Si–C bond cleavage to more complex substrates, we also investigated linear siloxane **2**. Random mutagenesis on *L*SilOx4 and performing enzymatic reactions in glass 96-well plates enabled identification of improved variant *L*SilOx5 (Fig. 2B). Although C–H hydroxylation and Si–C bond cleavage can occur at either the internal or terminal silicon groups of siloxane **2**, we only observe carbinol **6** and silanol **7** (supplementary materials). In the siloxane **2** reaction manifold, we can also quantify silanol **5** as a hydrolysis

volatile methylsiloxanes
thermally stable, unreactive
open environmental use
emerging regulations on the basis of persistence, bioaccumulation, and toxicity
degradation requires Si–C bond cleavage



	1	***2***	***3***
solubility in distilled water at 23 °C (ppb):	930.7 ± 0.7	34.5 ± 1.0	56.2 ± 2.5
hydrolysis half life at 25 °C, pH 7 (d):	4.8	13.7	4.5
vapor pressure at 25 °C (kPa):	5.5	0.5	1.3

Fig. 1. Physicochemical properties and structures of selected volatile siloxanes. ppb, parts per billion; d, days.

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA. [2]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA. [3]Core R&D, Dow Inc., Midland, MI 48674, USA. [4]Dow Silicones Corporation, Auburn, MI 48611, USA.
*Corresponding author. Email: dimi.katsoulis@dow.com (D.E.K.); frances@cheme.caltech.edu (F.H.A.)
†These authors contributed equally to this work.
‡Present address: Huminly, Inc., San Carlos, CA 94070, USA.
§Present address: Merck, South San Francisco, CA 94080, USA.

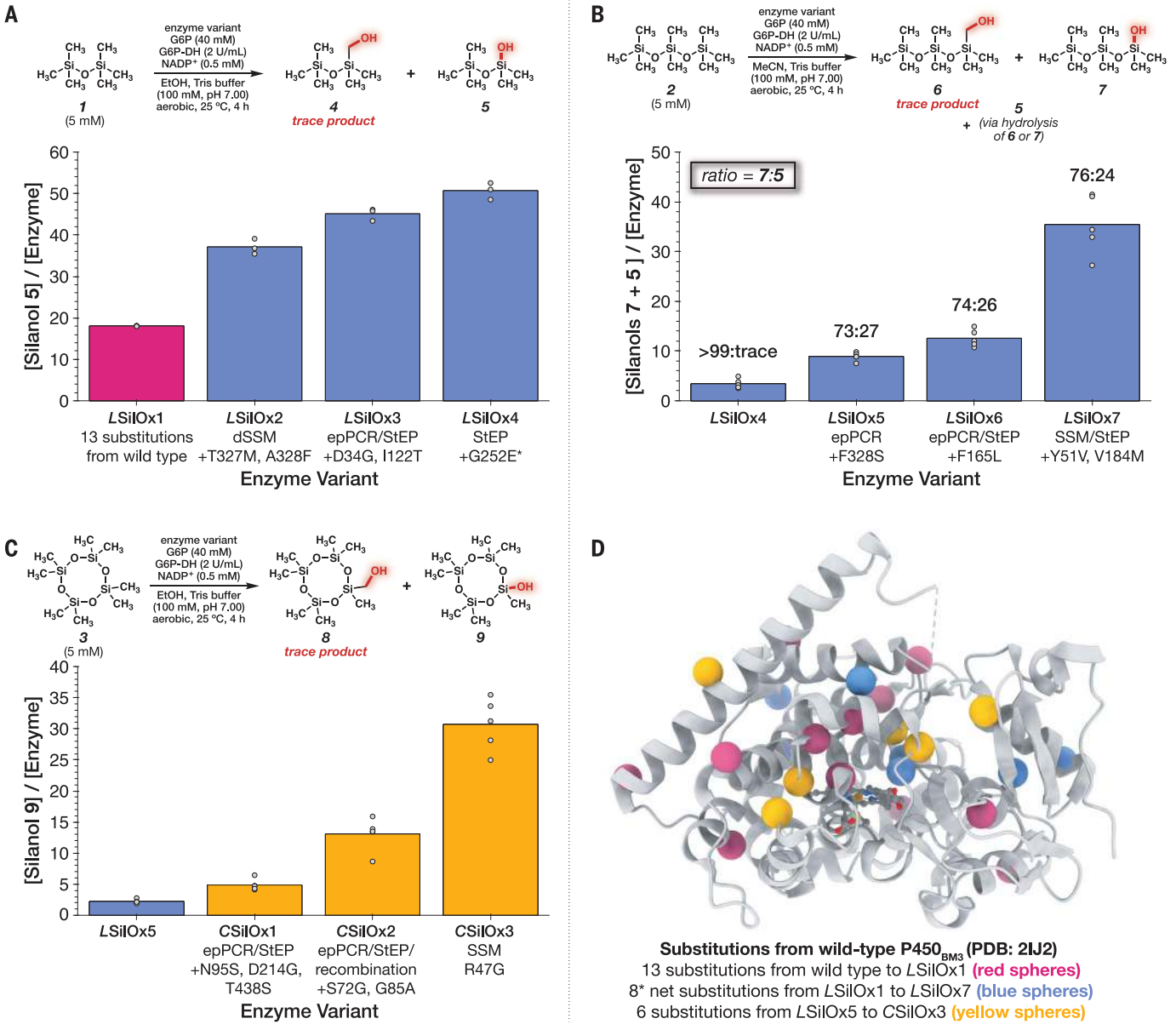


Fig. 2. Directed evolution with hexamethyldisiloxane (1), octamethyltrisiloxane (2), and octamethylcyclotetrasiloxane (3) in *E. coli* lysate. (**A**) Directed evolution of the *L*SilOx lineage on hexamethyldisiloxane (**1**). The G252E* mutation is a reversion to wild type [not shown in (D)]. (**B**) Directed evolution of the *L*SilOx lineage on octamethyltrisiloxane (**2**). (Inset) Ratio of silanol **7**/silanol **5** as determined by GC-MS. (**C**) Directed evolution of the *C*SilOx lineage on octamethylcyclotetrasiloxane (**3**). (**D**) Structure of wild-type $P450_{BM3}$ (Protein Data Bank ID 2IJ2) (*33*) showing amino acid substitutions accumulated during directed evolution. Substitutions present in the initial variant, *L*SilOx1, are shown as red spheres. Substitutions accrued throughout evolution in the *L*SilOx and *C*SilOx lineages are highlighted as blue and yellow spheres, respectively. Reactions were performed with enzyme in *E. coli* lysate with 5 mM substrate, 3.6% organic cosolvent [ethanol (EtOH) or acetonitrile (MeCN)], 0.5 mM $NADP^+$, 2 U/mL G6PDH, and 40 mM G6P, in 100 mM Tris buffer at pH 7.00. G6PDH, glucose-6-phosphate dehydrogenase; G6P, D-glucose-6-phosphate; Tris, tris(hydroxymethyl)aminomethane buffer; dSSM, double site-saturation mutagenesis; epPCR, error-prone polymerase chain reaction; StEP, staggered extension process recombination. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

product of carbinol **6** and silanol **7**. Variant *L*SilOx5 also exhibits activity on cyclic siloxane **3**, thus serving as a suitable starting point for directed evolution (Fig. 2C). Evolution on siloxanes **2** and **3** as representatives of linear and cyclic siloxanes resulted in divergent lineages of *L*SilOx and *C*SilOx (cyclic siloxane oxidase) variants, demonstrating the ability to enhance Si–C bond cleavage activity for different scaffolds. The amino acid substitutions accumulated during directed evolution are denoted below each individual variant relative to the previous generation variant in Fig. 2, A to C. The locations of the accumulated mutations are depicted with spheres superimposed on a crystal structure of wild-type $P450_{BM3}$ (*33*) in Fig. 2D. The three final variants, *L*SilOx4, *L*SilOx7, and *C*SilOx3, retain >50% of their activity when formulated as freeze-dried lyophilized lysate (tables S20, S22, and S24, respectively). The supplementary materials

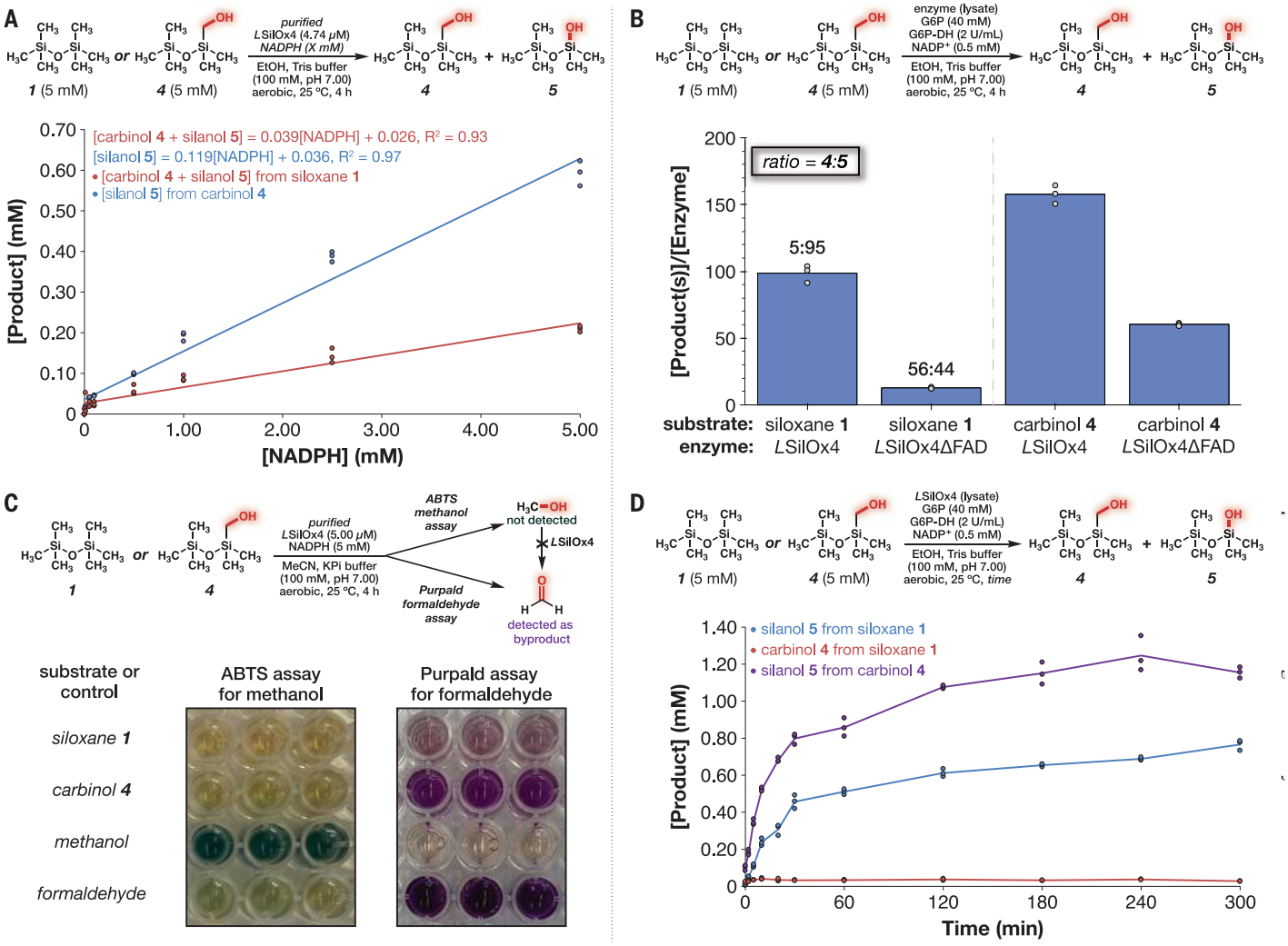


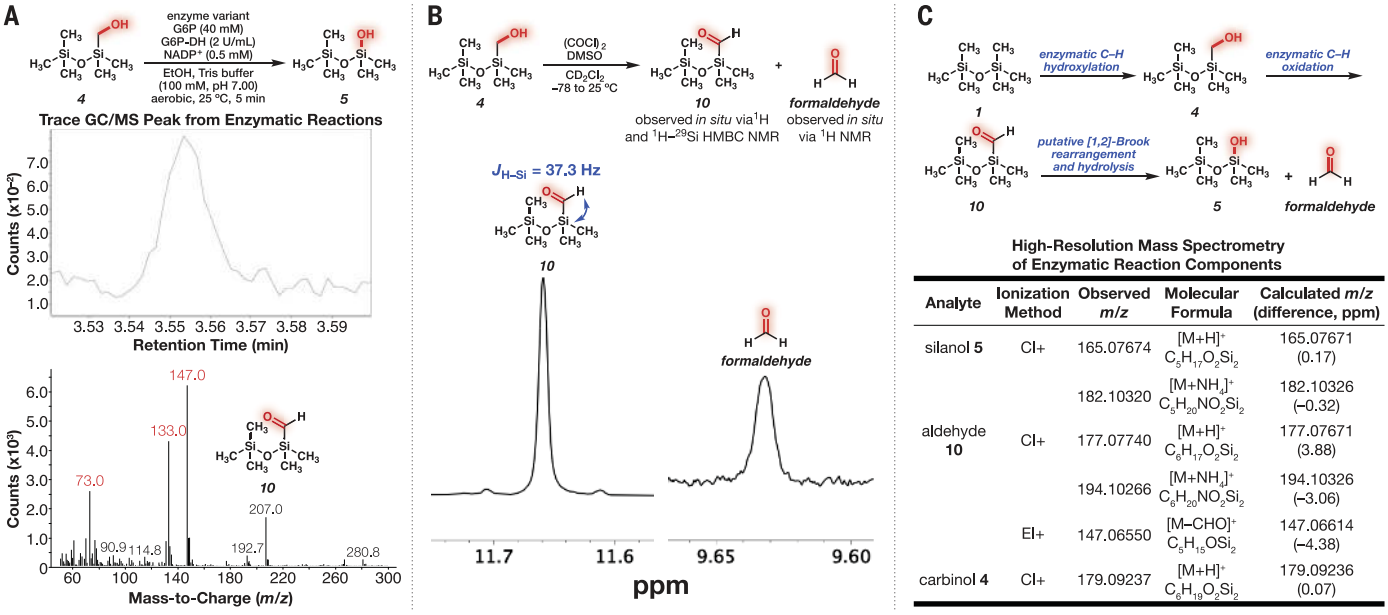
Fig. 3. Mechanistic investigation of enzymatic Si–C bond cleavage. (**A**) Si–C bond cleavage activity depends on NADPH concentration. (**B**) Truncation of the FAD domain in variant *L*SilOx4ΔFAD results in a 2.6-fold decrease in activity for Si–C bond cleavage from carbinol **4**. (**C**) Reaction analysis using the ABTS methanol assay and purpald formaldehyde assay demonstrates that formaldehyde is generated as a by-product of the enzymatic reaction. Control reactions use methanol or formaldehyde as a substrate instead of siloxane **1** or carbinol **4**. Methanol is not oxidized to formaldehyde by *L*SilOx4, and the cleaved carbon is lost as formaldehyde. See the supplementary materials for experimental details and quantitative analysis (figs. S28 and S30). The assays were performed in KPi buffer because of the high background observed in Tris. See fig. S44 for a comparison of enzymatic reactions in Tris versus KPi buffer. (**D**) Enzymatic reaction time courses with siloxane **1** or carbinol **4** as a substrate. See the supplementary materials for experimental details and controls.

contain GC-MS chromatograms and mass spectra for enzymatic reactions with siloxanes **1** to **3** (figs. S21, S23, and S25). These are compared with overlaid traces of reactions with the wild-type enzyme, which has no activity on siloxanes **1** to **3**. (figs. S22, S24, and S26). These data demonstrate that enzymes can cleave Si–C bonds under mild conditions—an activity not possible with any previously reported chemocatalysts and not previously reported for an enzyme (*4*, *25*). Although the overall Si–C bond cleavage activity is modest at this point, it demonstrates that biological activity on this non-natural substrate is both possible and can be enhanced. Further engineering and investigation of the prevalence of this activity will yield more powerful biocatalysts for siloxane degradation. The current level of activity also enabled investigation of the mechanism of Si–C bond cleavage.

The nature of enzymatic Si–C bond cleavage

In reactions of siloxanes with the SilOx variants, the silanols resulting from Si–C bond cleavage and hydrolysis dominate the product manifolds, whereas only trace quantities of the initial C–H hydroxylation carbinol products are detected. We performed a series of experiments to interrogate the mechanism responsible for Si–C bond cleavage from the intermediate carbinols, using siloxane **1** and carbinol **4** as a model system. First, we established that conversion of carbinol **4** to silanol **5** does not occur in the absence of enzyme as the result of a buffer-mediated [1,2]-Brook rearrangement in *E. coli* lysate (without expressed engineered enzyme variants; fig. S32) or in enzyme-free buffer tested at various pH conditions ranging from 6.25 to 9.00 (fig. S33). Instead, carbinol **4** undergoes pH-dependent decay through hydrolysis to trimethylsilanol and hydroxymethyldimethylsilanol. Conversion of carbinol **4** to silanol **5** occurs only through enzymatic catalysis with activity directly related to enzyme concentration (fig. S34). Furthermore, denaturation of *L*SilOx4 by heat treatment results in no formation of silanol **5**.

With the knowledge that engineered P450s catalyze both the siloxane C–H hydroxylation and subsequent Si–C bond cleavage, we investigated the parameters required for the enzymatic Si–C bond cleavage. With purified



Analyte	Ionization Method	Observed *m/z*	Molecular Formula	Calculated *m/z* (difference, ppm)
silanol **5**	CI+	165.07674	$[M+H]^+$ $C_5H_{17}O_2Si_2$	165.07671 (0.17)
		182.10320	$[M+NH_4]^+$ $C_5H_{20}NO_2Si_2$	182.10326 (–0.32)
aldehyde **10**	CI+	177.07740	$[M+H]^+$ $C_6H_{17}O_2Si_2$	177.07671 (3.88)
		194.10266	$[M+NH_4]^+$ $C_6H_{20}NO_2Si_2$	194.10326 (–3.06)
	EI+	147.06550	$[M–CHO]^+$ $C_5H_{15}OSi_2$	147.06614 (–4.38)
carbinol **4**	CI+	179.09237	$[M+H]^+$ $C_6H_{19}O_2Si_2$	179.09236 (0.07)

Fig. 4. Tandem double enzymatic oxidation of the siloxane methyl group results in silicon-carbon bond cleavage. (**A**) Identification of a trace peak found in GC-MS traces from enzymatic reactions with carbinol **4** with an example from a 5-min reaction. (**B**) ^{1}H NMR characterization of formylsiloxane **10** prepared by Swern oxidation of carbinol **4** in situ. Full ^{1}H and ^{1}H–^{29}Si NMR are available in the supplementary materials (figs. S40 to S42). ppm, parts per million. (**C**) Plausible mechanism for enzymatic Si–C bond cleavage through tandem enzymatic C–H hydroxylation and carbinol oxidation to formylsiloxane **10** with high-resolution mass spectrometry data acquired from 5-min enzymatic reactions using carbinol **4** as a substrate. Additional details are provided in table S41.

*L*SilOx4, we determined that enzymatic conversion of carbinol **4** to silanol **5** is NADPH dependent (Fig. 3A and see the supplementary materials for additional controls) and oxygen dependent (figs. S35 and S36), which indicates that oxidation by the Fe-heme is involved in the Si–C bond cleavage. Removal of the FAD domain of *L*SilOx4 led to a 2.6-fold loss of Si–C bond cleavage activity with variant *L*SilOx4ΔFAD (Fig. 3B). These results are consistent with the involvement of enzymatic oxidation in the Si–C bond cleavage. Thus, we turned our attention to uncovering the fate of the cleaved methyl group. We used the colorimetric ABTS [2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt] and purpald assays with purified *L*SilOx4 in potassium phosphate (KPi) buffer for the detection of methanol and formaldehyde, respectively (*34*, *35*). Methanol is not detected as a direct by-product of enzymatic reactions with either siloxane **1** or carbinol **4** by the ABTS assay, nor is *L*SilOx4 capable of oxidizing methanol (Fig. 3C and fig. S28). These results suggest that carbinol **4** does not undergo direct [1,2]-Brook rearrangement and subsequent hydrolysis to silanol **5** and methanol. Instead, we detect formaldehyde as a by-product in enzymatic reactions with siloxane **1** and carbinol **4**, in concentrations consistent with the concentration of silanol **5** quantified with parallel GC-MS analysis (table S32).

Enzymatic reaction time courses with carbinol **4** (Fig. 3D) revealed trace amounts of a GC-MS peak in 2- to 30-min time points that we hypothesized to be formylsiloxane **10** (Fig. 4 and fig. S37). To investigate this possibility, we performed a Swern oxidation of carbinol **4**, which enabled in situ synthesis and characterization of formylsiloxane **10** by ^{1}H–^{29}Si heteronuclear multiple bond correlation (HMBC) and proton nuclear magnetic resonance (^{1}H NMR) (*36*). Notably, formaldehyde was also observed in the NMR spectra, presumably owing to decomposition of formylsiloxane **10** with trace amounts of water. The Swern oxidation sample of formylsiloxane **10** matched the peak discovered in enzymatic reactions in both retention time and mass fragmentation with GC-MS (figs. S38 and S43). Attempts to isolate this labile intermediate were unsuccessful, with GC-MS analysis of samples showing complete decay of this signal over several hours (fig. S39). To further support our assignment, we acquired high-resolution electron impact (EI+) fragmentation and field ionization (FI+) soft ionization mass spectra of formylsiloxane **10** from an enzymatic reaction with carbinol **4** after extraction with ethyl acetate (EtOAc) after 5 min (table S41). With this evidence, we hypothesize that the initial C–H hydroxylation of the parent siloxane **1** is followed by a second oxidation to formyl siloxane **10**, which is converted to silanol **5**, ostensibly through a [1,2]-Brook rearrangement and hydrolysis (Fig. 4C) (*4*). Taken together, the tandem oxidation of the siloxane methyl group unlocks a mechanism for the enzymatic cleavage of Si–C bonds in VMS and serves as an entryway into future biodegradation efforts.

Unlocking degradation activities not known in nature

The large body of directed evolution literature demonstrates that even trace activities can be amplified to yield powerful biocatalysts for transformations not yet known in the biological world (*21*, *37*), including enzymes that forge (*38*) and break Si–C bonds. Siloxanes are produced at a megaton-per-year scale and are used in applications open to the environment (*4*). The cleavage of a siloxane Si–C bond catalyzed by the enzymes reported here is a proof of principle and represents a first step toward biodegradation of siloxanes that are currently not considered biodegradable (*25*). This situation is reminiscent of the discovery of *Ideonella sakaiensis*—a microorganism collected from sediment near a recycling plant rich in polyethylene terephthalate (PET)—and its evolved PET-degrading enzymes that enable it to grow on PET as the sole carbon source (*39*, *40*). This discovery inspired a slew of innovations, both enzymatic and microbial, that presented viable routes to PET biodegradation, which are just now coming to fruition (*41*). Legislation is already appearing to limit the use of VMS, including octamethylcyclotetrasiloxane (**3**), on which we demonstrate activity (*14*, *15*). By engineering enzymes that can cleave Si–C bonds, we take a key step forward toward the eventual biodegradation of VMS.

REFERENCES AND NOTES

1. J. Chelikowsky, in *Silicon: Evolution and Future of a Technology*, P. Siffert, E. F. Krimmel, Eds. (Springer, 2004), pp. 1–22.
2. W. Noll, Ed., in *Chemistry and Technology of Silicones* (Academic Press, 1968), pp. 531–638.
3. A. J. O'Lenick, *Silicones for Personal Care* (Allured Publishing, ed. 2, 2008).
4. C. Rücker, K. Kümmerer, *Chem. Rev.* **115**, 466–524 (2015).
5. W. Noll, Ed., in *Chemistry and Technology of Silicones* (Academic Press, 1968), pp. 190–245.
6. K. Mojsiewicz-Pieńkowska, D. Krenczkowska, *Chemosphere* **191**, 204–217 (2018).
7. S. Genualdi *et al*., *Environ. Sci. Technol.* **45**, 3349–3354 (2011).
8. D.-G. Wang, W. Norwood, M. Alaee, J. D. Byer, S. Brimble, *Chemosphere* **93**, 711–725 (2013).
9. X. Xiang, N. Liu, L. Xu, Y. Cai, *Ecotoxicol. Environ. Saf.* **224**, 112631 (2021).
10. N. A. Warner *et al*., *Environ. Sci. Technol.* **44**, 7705–7710 (2010).
11. O. L. Flaningam, *J. Chem. Eng. Data* **31**, 266–272 (1986).
12. Y. D. Lei, F. Wania, D. Mathers, *J. Chem. Eng. Data* **55**, 5868–5873 (2010).
13. Global Silicones Council, "Regulatory Information"; https://globalsilicones.org/regulation/.
14. ECHA (European Chemicals Agency), "Inclusion of substances of very high concern in the Candidate List for eventual inclusion in Annex XIV" (2018); https://echa.europa.eu/documents/10162/2be7bcbf-f797-c28c-2c67-939664155c7c.
15. ECHA, "Support Document for Identification of Octamethylcyclotetrasiloxane (D_4) as a Substance of Very High Concern because of its PBT and vPvB properties" (2018); https://echa.europa.eu/documents/10162/115f70a9-a387-1525-d49f-b715e84996e4.
16. E. C. Tuazon, S. M. Aschmann, R. Atkinson, *Environ. Sci. Technol.* **34**, 1970–1976 (2000).
17. M. W. Alton, E. C. Browne, *ACS Environ. Au* **2**, 263–274 (2022).
18. S. Varaprath, J. M. McMahon, K. P. Plotzke, *Drug Metab. Dispos.* **31**, 206–214 (2003).
19. S. Varaprath, K. L. Salyers, K. P. Plotzke, S. Nanavati, *Drug Metab. Dispos.* **27**, 1267–1273 (1999).
20. M. B. Reddy *et al*., *Toxicol. Sci.* **72**, 3–18 (2003).
21. F. H. Arnold, *Angew. Chem. Int. Ed.* **57**, 4143–4148 (2018).
22. R. A. Sheldon, J. M. Woodley, *Chem. Rev.* **118**, 801–838 (2018).
23. T. D. Sutherland *et al*., *Clin. Exp. Pharmacol. Physiol.* **31**, 817–821 (2004).
24. N. S. Sarai, B. J. Levin, J. M. Roberts, D. E. Katsoulis, F. H. Arnold, *ACS Cent. Sci.* **7**, 944–953 (2021).
25. C. Rücker, E. Grabitz, K. Kümmerer, *Chemosphere* **321**, 137858 (2023).
26. N. P. Dunham, F. H. Arnold, *ACS Catal.* **10**, 12239–12255 (2020).
27. K. Mittra, M. T. Green, *J. Am. Chem. Soc.* **141**, 5504–5510 (2019).
28. J. Münch, P. Püllmann, W. Zhang, M. J. Weissenborn, *ACS Catal.* **11**, 9168–9203 (2021).
29. C. J. C. Whitehouse, S. G. Bell, L.-L. Wong, *Chem. Soc. Rev.* **41**, 1218–1260 (2012).
30. A. M. Doncaster, R. Walsh, *J. Chem. Soc., Faraday Trans. I* **72**, 2908–2916 (1976).
31. S. T. Jung, R. Lauchli, F. H. Arnold, *Curr. Opin. Biotechnol.* **22**, 809–817 (2011).
32. S. Bähr *et al*., *Angew. Chem. Int. Ed.* **59**, 15507–15511 (2020).
33. H. M. Girvan *et al*., *J. Biol. Chem.* **282**, 564–572 (2007).
34. R. E. Childs, W. G. Bardsley, *Biochem. J.* **145**, 93–103 (1975).
35. M. S. Quesenberry, Y. C. Lee, *Anal. Biochem.* **234**, 50–55 (1996).
36. R. E. Ireland, D. W. Norbeck, *J. Org. Chem.* **50**, 2198–2200 (1985).
37. K. Chen, F. H. Arnold, *Nat. Catal.* **3**, 203–213 (2020).
38. S. B. J. Kan, R. D. Lewis, K. Chen, F. H. Arnold, *Science* **354**, 1048–1051 (2016).
39. I. Taniguchi *et al*., *ACS Catal.* **9**, 4089–4105 (2019).
40. S. Yoshida *et al*., *Science* **351**, 1196–1199 (2016).
41. S. Miri *et al*., *Chemosphere* **286**, 131670 (2022).

ACKNOWLEDGMENTS

We thank S. Athavale, S. Bähr, L. F. Caldera, and B. J. Levin for productive discussions and experimental assistance and S. Gao and J. C. Reisenbauer for critical review of the manuscript and data. We thank P. J. Almhjell, A. Das, Y. Long, K. Plotzke, G. Kozerski, D. Eldred, E. Joffre, and D. McNett for productive discussions. We thank N. W. Goldberg for the initial design of custom reaction apparatus. We thank D. VanderVelde for NMR assistance and maintenance of the Caltech Liquids NMR facility and M. Shahgoli for insightful discussions, mass spectrometry assistance, and maintenance of the Caltech CCE Multiuser Mass Spectrometry Laboratory. We thank the Dow Next Generation Instrument Grant. We thank E. Hicks and B. Stoltz for access to distillation equipment. **Funding:** This work was supported by the Dow University Partnership Initiative (grants 227027AO and 227027AU). N.S.S. and R.L.O. are each supported by grants from the National Science Foundation Graduate Research Fellowship Program (NSF grant DGE-1745301). **Author contributions:** Conceptualization: N.S.S., T.J.F., R.L.O., K.E.J., S.B.-C., R.R.M., J.M.R., J.C.T.R., D.E.K., and F.H.A. Methodology: N.S.S., T.J.F., R.L.O., K.E.J., S.B.-C., and R.E.T. Investigation: N.S.S., T.J.F., R.L.O., K.E.J., S.B.-C., R.E.T., and J.M.R. Visualization: N.S.S., T.J.F., and R.L.O. Funding acquisition: F.H.A. Project administration: S.B.-C., R.R.M., J.M.R., J.C.T.R., D.E.K., and F.H.A. Supervision: S.B.-C. and F.H.A. Writing – original draft: N.S.S., T.J.F., R.L.O., D.E.K., and F.H.A. Writing – review and editing: N.S.S., T.J.F., R.L.O., K.E.J., S.B.-C., R.R.M., J.M.R., D.E.K., and F.H.A. **Competing interests:** N.S.S., T.J.F., R.L.O., F.H.A., S.B.-C., R.R.M., J.M.R., D.E.K., and J.C.T.R. are inventors on a patent application describing enzymatic siloxane degradation filed by the Dow Chemical Company and the California Institute of Technology. The authors declare no other competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Plasmids encoding the enzymes reported in this study are available for research purposes from F.H.A. under a material transfer agreement with the California Institute of Technology.

SUPPLEMENTARY MATERIALS

Submitted 3 May 2023; resubmitted 7 November 2023
Accepted 18 December 2023
10.1126/science.adi5554

PLANT SYMBIOSIS

Receptor-associated kinases control the lipid provisioning program in plant–fungal symbiosis

Sergey Ivanov and Maria J. Harrison*

The mutualistic association between plants and arbuscular mycorrhizal (AM) fungi requires intracellular accommodation of the fungal symbiont and maintenance by means of lipid provisioning. Symbiosis signaling through lysin motif (LysM) receptor-like kinases and a leucine-rich repeat receptor-like kinase DOES NOT MAKE INFECTIONS 2 (DMI2) activates transcriptional programs that underlie fungal passage through the epidermis and accommodation in cortical cells. We show that two *Medicago truncatula* cortical cell–specific, membrane-bound proteins of a CYCLIN-DEPENDENT KINASE-LIKE (CKL) family associate with, and are phosphorylation substrates of, DMI2 and a subset of the LysM receptor kinases. *CKL1* and *CKL2* are required for AM symbiosis and control expression of transcription factors that regulate part of the lipid provisioning program. Onset of lipid provisioning is coupled with arbuscule branching and with the REDUCED ARBUSCULAR MYCORRHIZA 1 (RAM1) regulon for complete endosymbiont accommodation.

Arbuscular mycorrhizal (AM) symbioses are widespread in terrestrial environments; they influence plant mineral nutrition and carbon allocation below ground and, consequently, ecosystem productivity (*1*). The endosymbiosis develops in the roots, where differentiated hyphae called arbuscules are accommodated in membrane-bound apoplastic compartments generated de novo within the cortical cells (*2*, *3*). These elaborate interfaces are the sites of nutrient exchange. The mutualism requires substantial modifications to root cell metabolism and transport to enable the cell to provision the lipid auxotrophic endosymbiont and to access the phospha delivers (*4*). These modifications are achi through alterations to the root cortical cell transcriptome, but knowledge of the signaling that underlies these responses is still incomplete.

AM symbiosis arose early in the plant lineage (*5*), and AM symbiosis–competent host plants share several genes conserved in hosts and missing from nonhosts (*6*, *7*). These so-called AM symbiosis–conserved genes include a family of kinases that cluster within a larger family referred to as *CYCLIN-DEPENDENT KINASE-LIKE* (*CKL*) (*6*, *8*). The genome of model plant *Medicago truncatula* (*Medicago*) contains two AM symbiosis–conserved *CKL* genes, which we refer to as *CKL1* and *CKL2*. The *CKL* family is related to but distinct from the canonical cyclin-dependent kinases (fig. S1, A and B). To date, functions for *CKL* family members have not been reported.

CKL1 and *CKL2* are required for expression of part of the lipid provisioning program

Expression of *CKL1* and *CKL2* is induced during AM symbiosis (fig. S1C), and the promoters of

Boyce Thompson Institute, Ithaca, NY 14853, USA.
*Corresponding author. Email: mjh78@cornell.edu

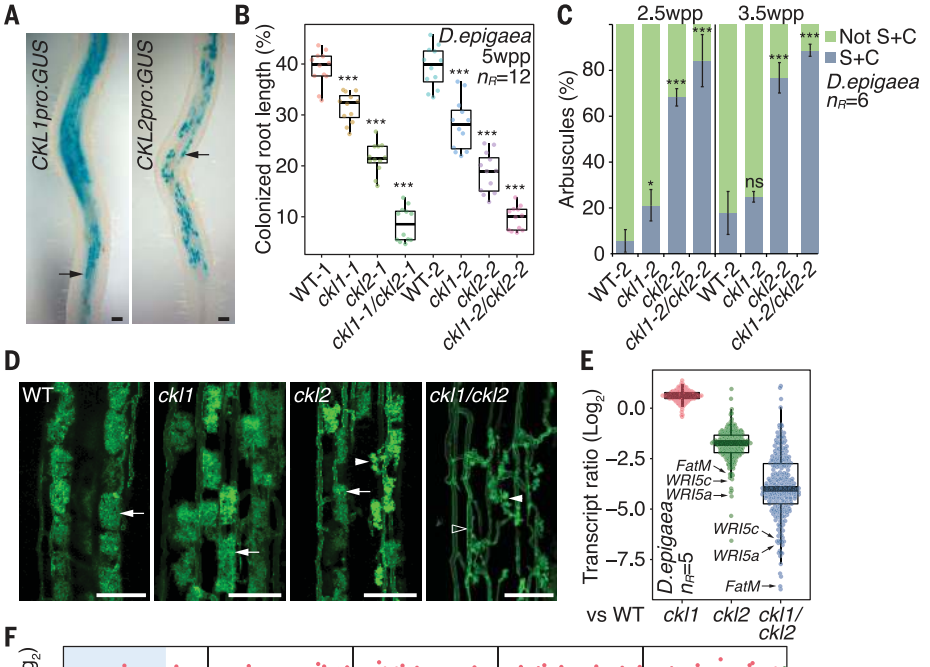


Fig. 1. *CKL1* and *CKL2* are required for AM symbiosis. (**A**) *CKL1* and *CKL2* promoter activity in *M. truncatula* roots colonized by AM fungus *Rhizophagus irregularis* as assessed through β-glucuronidase (GUS) activity. Arrows indicate arbuscules. Scale bars, 100 μm. (**B** to **F**) Effect of mutations in *CKL1* and *CKL2*. (B) The percentage of root length with intraradical colonization in WT segregant (WT), *ckl1*, *ckl2*, and *ckl1/ckl2* mutants colonized by AM fungus *D. epigaea* 5 weeks postplanting (wpp). n_R = 12, number of root systems evaluated for each mutant allele (***P < 0.001; ns, not significant; t test). (C) Proportion of senescing and collapsed arbuscules (S+C) and not senescent or collapsed arbuscules (Not S+C) in WT segregants, *ckl1*, *ckl2*, and *ckl1/ckl2* (*P < 0.01; ***P < 0.001; t test). (D) Arbuscules in WT, *ckl1*, *ckl2*, and *ckl1/ckl2*. Confocal microscopy images of WT and mutant roots stained with wheat germ agglutinin (WGA)–Alexa488 to visualize fungal structures (z-stack projection, n = 10 optical slices, 1-μm intervals). Arrows indicate mature arbuscules; solid arrowheads indicate senescing and collapsed arbuscules; open arrowhead indicates fungal septa. Scale bars, 100 μm. (E) Box plot shows transcript ratios (log_2) in colonized roots *ckl1*:WT, *ckl2*:WT, and *ckl1/ckl2*:WT. Box plot shows median, upper, and lower quartiles, and whiskers show 1.5 interquartile range. (F) Transcript ratios, *ckl1*:WT and *ckl2*:WT, from colonized roots in rank order based on *ckl2*:WT data. Blue shading shows the 15th percentile. Lipid provisioning–related genes and AM marker genes are indicated. Data in (E) and (F) are the top 250 genes induced in *D. epigaea*–colonized WT relative to mock-inoculated control (fig. S5A). Experiments were replicated as shown in table S8.

both genes are active in the cortex of roots colonized by AM fungi (Fig. 1A and fig. S1D). *CKL2* promoter activity is restricted to colonized cells containing arbuscules, whereas the *CKL1* promoter is active in colonized and adjacent non-colonized cells (Fig. 1A and fig. S1D). Constitutive expression of gain-of-function variants of the major regulators *DOES NOT MAKE INFECTIONS 3* (*DMI3*) (*9*) and *DELLA* (*10*), but not *REDUCED ARBUSCULAR MYCORRHIZA 1* (*RAM1*) (*11*), drives 10- to 100-fold increases in *CKL1* and *CKL2* transcripts in uninoculated roots (fig. S1, E to G). These data indicate differences in the control of *CKL* expression relative to genes of the accommodation program (*12*). Public transcriptome data indicate almost AM symbiosis–specific expression of *CKL2*, whereas *CKL1* expression is detected in non-AM symbiotic conditions (fig. S2).

To assess *CKL1* and *CKL2* functions during AM symbiosis, we generated *Medicago* mutants using CRISPR-Cas9 genome editing and obtained two independent alleles for each gene and two corresponding double mutants. The mutations resulted in premature stop codons (fig. S3A and table S1). The *ckl* mutants did not exhibit visible differences in general root or shoot growth relative to the wild-type (WT) segregants (fig. S3B). After inoculation with AM fungus *Diversispora epigaea*, initial hyphal penetration of the epidermis was equivalent in mutants and WT (fig. S3, C and D), but intraradical colonization was impaired in the mutants (Fig. 1B). Relative to WT, *ckl1* alleles showed, on average, a 24% reduction in fungal colonized root length, whereas *ckl2* and the *ckl1/ckl2* alleles showed, on average, a 49 and 77% reduction in fungal colonized root length, respectively (Fig. 1B). In addition, *ckl2* and *ckl1/ckl2* showed more than 60 and 80% senescing and collapsed arbuscules, respectively (Fig. 1, C and D, and fig. S3E). Periarbuscular membrane (PAM)–resident proteins PHOSPHATE TRANSPORTER 4 (PT4) (*13*) and BLUE COPPER BINDING PROTEIN 1 (BCP1) (*14*), as well as members of symbiotic exocytotic membrane fusion machinery, could be detected in *ckl2* and *ckl1/ckl2* mutants, indicating that PAM development and trafficking of proteins occur in the mutants before arbuscule collapse (fig. S3, F and G). Endoreduplication occur in AM roots and results in measurable increases in nuclei size in colonized cortical cells (*15*). The size of nuclei in WT and *ckl1/ckl2* colonized cells did not differ, thus falsifying an initial hypothesis that *CKL1* and *CKL2* are required for cortical cell endoreduplication during symbiosis (fig. S4).

We examined the transcript profiles of *ckl* mutant roots with and without colonization by *D. epigaea* (Fig. 1, E and F; fig. S5; and data S1). Transcript profiles of *ckl1* and WT colonized roots were similar to each other and showed the expected marker gene expression (*16*). More

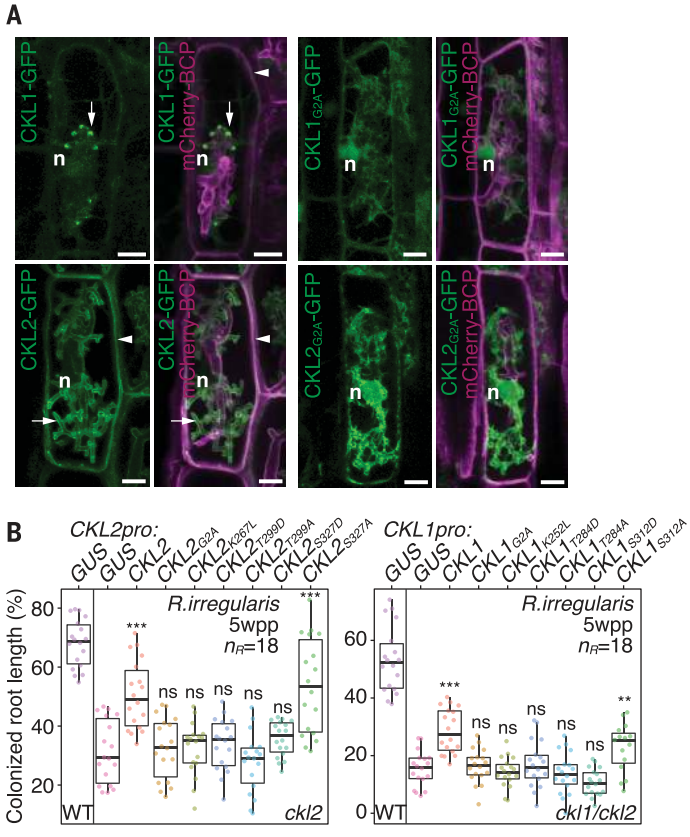


Fig. 2. CKL1 and CKL2 functions require membrane location and kinase activity. (**A**) CKL1 and CKL2 fusion proteins expressed from their native promoters. CKL1-GFP detected on the PAM in the region around the tips of arbuscule branches (arrow). CKL2-GFP detected on the PAM (arrow) and PM (arrowhead). Myristoylation mutants $CKL1_{G2A}$-GFP and $CKL2_{G2A}$-GFP lost their membrane association and located in the cytoplasm and nucleoplasm. Proteins were coexpressed with a PM and PAM marker (mCherry-BCP). Confocal microscopy images (z-stack projection, n = 10 optical slices, 0.5-μm intervals). n, nucleus. Scale bars, 10 μm. (**B**) Complementation of colonization in *ckl2* or *ckl1/ckl2* with different CKL1 and CKL2 variants, including myristoylation mutant ($CKL1_{G2A}$ and $CKL2_{G2A}$), kinase active site mutant ($CKL1_{K252L}$ and $CKL2_{K267L}$; fig. S1B), phosphoablative ($CKL1_{T284A}$ and $CKL2_{T299A}$) and phosphomimetic ($CKL1_{T284D}$ and $CKL2_{T299D}$) mutants of the conserved threonine in kinase domain activation segment (fig. S1B), and phosphoablative ($CKL1_{S312A}$ and $CKL2_{S327A}$) and phosphomimetic ($CKL1_{S312A}$ and $CKL2_{S327A}$) mutants of a conserved serine. Percentage of root length with intraradical colonization by *R. irregularis* at 5 wpp, n_R = 12, number of evaluated root systems for each mutant. WT included for colonization reference. Asterisks indicate the level of statistical significance (**P < 0.01; ***P < 0.001; t test) in comparison to mutants expressing a GUS gene from the CKL1 or CKL2 promoters. The *ckl1/ckl2* mutant provided a sensitized background against which to assess the *CKL1* variants. Experiments were replicated as shown in table S8.

than 250 genes showed a $\log_2$-fold increase of >2, and 165 genes showed a $\log_2$-fold increase of >5 in transcript level in colonized roots relative to mock-inoculated controls. Transcriptional increases in marker genes in colonized *ckl2* and *ckl1/ckl2* roots were lower than those of WT and *ckl1*, which is consistent with lower levels of fungal colonization (fig. S5B). Examination of *ckl2*:WT transcript ratios from colonized roots revealed a small group of genes expressed at very low levels in *ckl2* (Fig. 1, E and F, and data S1). This group included genes with essential roles in AM symbiosis including *WRINKLED* transcription factors *WRI5a* and *WRI5c* (*17*, *18*), as well as *FATTY ACYL-ACP THIOESTERASE M* (*FatM*) and *STUNTED ARBUSCULE* (*STR*) (*19*, *20*), which are regulated by WRI5a and WRI5c (*18*). The *ckl1/ckl2*:WT transcript ratios showed a similar pattern to that observed for *ckl2*:WT, with a slightly different rank order (fig. S5C). FatM is required for fatty acid biosynthesis, and STR, a half-size ABC transporter (*20*), is involved in lipid export to the fungus (*17*, *19*–*22*); their loss-of-function phenotypes are similar to those of *ckl2* and *ckl1/ckl2* mutants. At least two other genes in the *ckl2* low-expression group have predicted lipid-related roles, including a lipid transfer protein (*LTP*) and a ceramidase (*CER*) (Fig. 1F and data S1); however, *RAM2*, a glyceraldehyde-3-phosphate acyltransferase, also essential in the lipid biosynthesis program but directly regulated by *RAM1* (*11*, *17*), was not in this group. Although any of the genes in the low-expression cluster could contribute to the *ckl2* mycorrhizal phenotype, reduced expression of *WRI5a*, *WRI5c*, *FatM*, and *STR* alone is sufficient to explain the *ckl2* and *ckl1/ckl2* mutant phenotypes. Thus, we conclude that *CKL1* and *CKL2* functions are necessary for expression of several genes

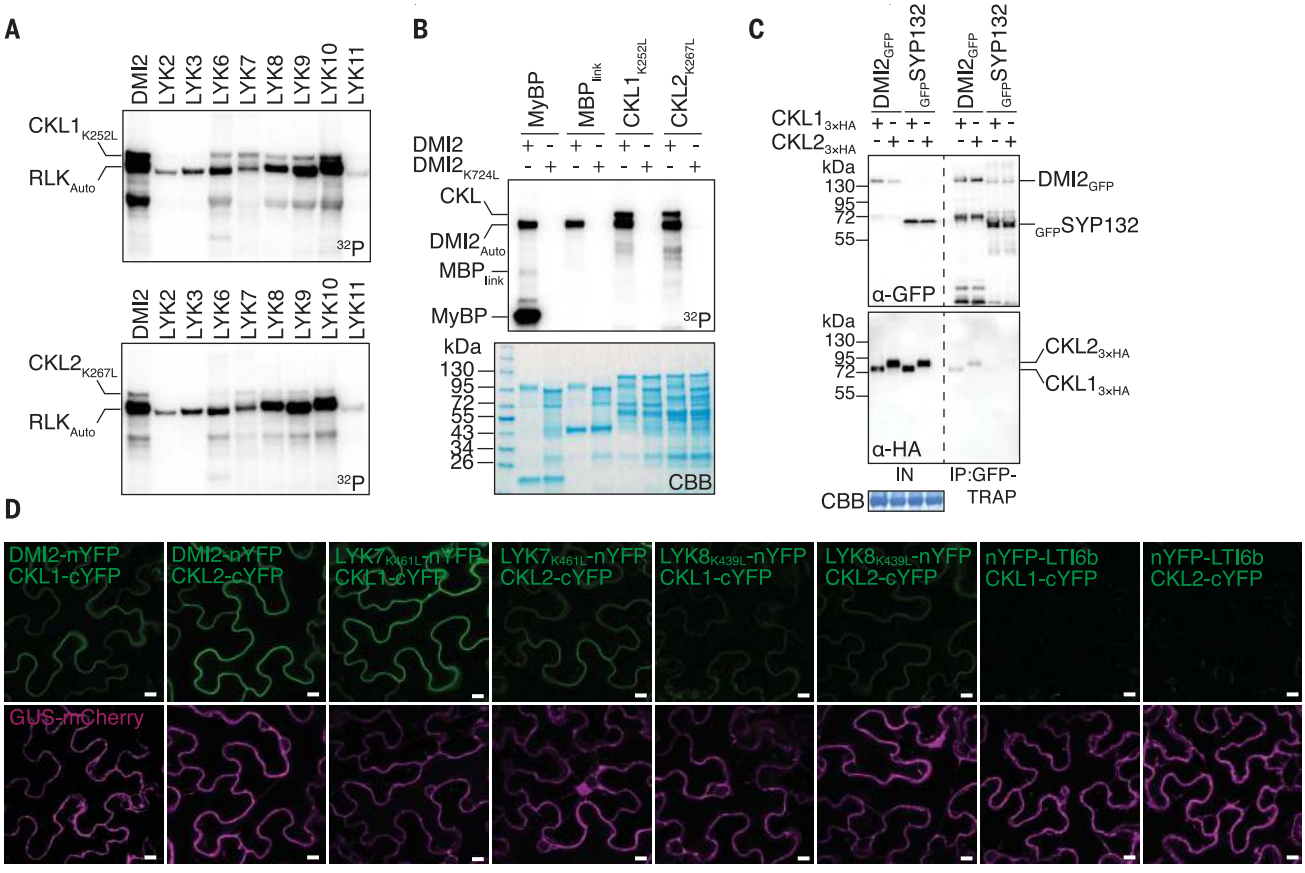


Fig. 3. DMI2 and a subset of the LysM-RLKs interact with, and phosphorylate, CKL1 and CKL2. (**A** and **B**) ^{32}P radiographs showing in vitro transphosphorylation of kinase-dead $CKL1_{K252L}$ and $CKL2_{K267L}$ by RLKs (endodomains) (A); phosphorylation is abolished when kinase-dead $DMI2_{K724L}$ is used (B). Myelin basic protein (MyBP) was used as generic phosphorylation substrate. The kinases were purified as maltose binding protein (6×His-MBP) fusions, so free maltose binding protein with recombination linker (MBP_{link}) was included as a negative control. RLK_{Auto}, autophosphorylation of RLKs; CBB, Coomassie brilliant blue stained gel corresponds to radiograph in (B). (**C**) Coimmunoprecipitation assay of $CKL1_{3\times HA}$ or $CKL2_{3\times HA}$ with DMI2-GFP. GFP-SYP132 was used as a membrane protein control for immunoprecipitation. Western blots were probed with antibodies α-GFP and α-HA. IN, input; IP, immunoprecipitation. (**D**) BiFC assays of CKL1-cYFP or CKL2-cYFP and RLKs-nYFP in *N. benthamiana* leaves. YFP complementation is visualized in transformed cells that exhibit fluorescence from transformation marker GUS-mCherry. nYFP-LTI6b serves as a membrane protein negative control. *CKL1-cYFP* or *CKL2-cYFP* (fusions to C-terminal fragment of YFP) and *DMI2-nYFP* or kinase-dead variants *LYK7$_{K461L}$-nYFP*, *LYK8$_{K439L}$-nYFP*, or *nYFP-LTI6b* (fused to N-terminal fragment of YFP) were expressed from a single transfer DNA with *GUS-mCherry*. Confocal microscopy images. Scale bars, 10 μm. Experiments were replicated as shown in table S8.

involved in the production and provisioning of lipids to the fungus.

Membrane association and kinase activity are required for CKL1 and CKL2 functions

The subcellular location of each CKL protein was assessed through visualization of translational fusions with fluorescent proteins expressed from the native *CKL1* or *CKL2* promoters. CKL1-GFP (green fluorescent protein) was visible at the PAM in the region around the tips of arbuscule branches and in the cytoplasm and nucleoplasm, whereas CKL2-GFP showed an even distribution exclusively across the entire plasma membrane (PM) and PAM (Fig. 2A and fig. S6, A to C). Both proteins were detected throughout the arbuscule lifetime (fig. S7). The CKL proteins lack transmembrane domains, but both proteins are predicted to be myristoylated. Mutation of the myristoylation motif (G2A) abolished membrane association of each CKL protein and resulted in their accumulation in the cytoplasm and nucleoplasm (Fig. 2A and fig. S6D).

Using in vitro phosphorylation assays, we found that CKL1 and CKL2 were capable of autophosphorylation and transphosphorylation of a generic substrate, myelin basic protein. Phosphorylation activity was abolished by mutation of the predicted kinase active site, $CKL1_{K252L}$ and $CKL2_{K267L}$ (K, Lys; L, Leu; figs. S1B and S8A). *CKL* genes encoding kinase-dead mutant proteins expressed from their native promoters failed to complement *ckl2* or *ckl1/ckl2* mutants (Fig. 2B). Likewise, *CKL1* and *CKL2* myristoylation motif mutants were also unable to restore colonization in *ckl1/ckl2* (Fig. 2B and fig. S9). Thus, CKL1 and CKL2 kinase activities and membrane location are essential for their functions.

CKL proteins are phosphorylated by two classes of receptor-like kinases that function during AM symbiosis

Given their locations at the PM and PAM and their downstream impact on transcription, we hypothesized that CKL1 and CKL2 might be involved in signal transduction from membrane receptor kinases. Members of two classes of receptor-like kinases (RLKs) are central to AM symbiosis signaling: the lysin motif (LysM) receptor-like kinases (*23*–*25*) and a malectin-like domain and leucine-rich repeat (MLD-LRR) receptor-like kinase, DMI2 (*Lotus japonicus SYMRK*) (*26*, *27*). During AM symbiosis, chitooligosaccharides are perceived by LysM-RLKs, and, in the presence of DMI2, symbiosis signaling is initiated. This leads to nuclear calcium oscillations and expression of genes for intracellular accommodation of the AM fungus (*23*, *28*, *29*). The mechanisms

underlying signal transduction from membrane-located RLKs to the nucleus are unclear but may involve a metabolite, mevalonate (*30*). Chitooligosaccharide signaling has been studied mostly in epidermal cells but is proposed to occur also in cortical cells coincident with fungal colonization of the root cortex (*31*). Consistent with this proposal, we detected promoter activity of *DMI2* and several *LysM-RLKs* in cortical cells, although DMI2-GFP and LysM-RLK-GFP protein fusions were below our limits of detection (fig. S10). To determine whether DMI2 or LysM-RLKs could phosphorylate CKL1 or CKL2, the endodomains of DMI2 and 10 *Medicago* LysM-RLKs (LYK2 to LYK11) (table S1) were purified, and their kinase activity was evaluated using in vitro phosphorylation assays (fig. S8, C and D). $CKL1_{K252L}$ and $CKL2_{K267L}$ protein variants, which lack kinase activity, were provided as substrates to enable an assessment of transphosphorylation by the RLKs. DMI2 and five LysM-RLKs (LYK6 to LYK10) phosphorylated CKL1 and CKL2, whereas phosphorylation was not detected for three LysM-RLKs (LYK2, LYK3, and LYK11) (Fig. 3, A and B, and fig. S11, A and B). LYK3, an RLK generally considered specific for Nod factor signaling (*32*), shows autophosphorylation activity comparable to, if not higher than, that of LYK7 but no phosphorylation of $CKL1_{K252L}$ or $CKL2_{K267L}$. Thus, these data suggest some selectivity among these RLKs for CKL substrates. The CKLs each contain many potential phosphorylation sites; in an initial assay with fragments of the CKL proteins, DMI2 preferentially phosphorylated the C-terminal regions of $CKL1_{K252L}$ and $CKL2_{K267L}$ (fig. S11C). CKL1 and CKL2 were evaluated for their ability to phosphorylate kinase-dead variants of DMI2, LYK2 to LYK11, and LYK-related (LYR) co-receptors, LYR1, LYR4, LYR8, and NOD FACTOR PERCEPTION (NFP) (fig. S8B). Neither CKL1 nor CKL2 phosphorylated these RLK endodomains (fig. S12). Thus, data from in vitro kinase assays support the hypothesis that CKL1 and CKL2 are phosphorylation substrates of DMI2 and a subset of the LysM-RLKs.

To assess association of CKL1 and CKL2 with the RLKs, genes encoding full-length *CKL1* or *CKL2* epitope-tagged fusions were transiently coexpressed with full-length *DMI2-GFP*, *LYK3-GFP*, *LYK9-GFP*, *LYK10-GFP*, *NFP-GFP*, or *LYR1-GFP* in *Nicotiana benthamiana* leaves (fig. S13A) and interactions assessed using coimmunoprecipitation assays from cell extracts. $CKL1_{3\times HA}$ and $CKL2_{3\times HA}$ coimmunoprecipitated with DMI2-GFP, indicating their associations in planta (HA, hemagglutinin; Fig. 3C and fig. S13B). In addition, $CKL1_{3\times HA}$ and $CKL2_{3\times HA}$ coimmunoprecipitated with LYR1-GFP, an inactive LysM-RLK (fig. S13C), and to a lesser extent with LYK3, LYK9, and LYK10 (fig. S13C). Expression of *LYK9-GFP* in *N. benthamiana* leaves induces cell death (*23*), which compromises the coimmunoprecipitation assays. A similar response, even stronger than that elicited by LYK9, was observed for *LYK7-GFP* and *LYK8-GFP* (fig. S13D). Cell death was not induced by the kinase-dead variants $LYK7_{K461L}$, $LYK8_{K439L}$, or $LYK9_{K443L}$ (fig. S13D). Both $LYK7_{K461L}$-GFP and $LYK8_{K439L}$-GFP accumulated in the PM (fig. S13E); however, $LYK9_{K443L}$-GFP lost its membrane location and accumulated in the cytosol (fig. S13E). Kinase-dead LYK7 and LYK8 variants offer an opportunity to assess interactions with CKL proteins. In coimmunoprecipitation assays, $CKL1_{3\times HA}$ and $CKL2_{3\times HA}$ coimmunoprecipitated with $LYK7_{K461L}$-GFP and $LYK8_{K439L}$-GFP (fig. S13F). Trace amounts of CKL were visible in some of the MtLTI6b and MtSYP132 coimmunoprecipitation negative controls, albeit at levels lower than the RLK coimmunoprecipitations (Fig. 3C and fig. S13, B and F). Therefore, to further evaluate CKL–RLK interactions, we implemented bimolecular fluorescence complementation (BiFC) assays using proteins tagged with N-terminal (nYFP) or C-terminal (cYFP) fragments of yellow fluorescent protein. DMI2-nYFP interacted with CKL1-cYFP and CKL2-cYFP in BiFC assays and did not interact with the LTI6b membrane protein control (Fig. 3D). Interactions of CKL proteins with $LYK7_{K461L}$-nYFP and $LYK8_{K439L}$-nYFP were also detected (Fig. 3D), whereas interactions with the cytosol-located $LYK9_{K443L}$ were not detected (fig. S14A). Interactions between CKL proteins and LYK10 or LYR1 were not visible; however, a weak fluorescence complementation signal for LYK10 with CKL2 was detected when coexpressed with LYR1 (fig. S14B), suggesting that the predicted co-receptor may stabilize this interaction. As anticipated, given the cell death responses, interactions between active LYK7, LYK8, or LYK9 and CKLs were not detected by BiFC assays (fig. S14C). On the basis of these data, we conclude that CKL1 and CKL2 each interact with DMI2 and at least two LysM-RLKs, LYK7 and LYK8. Together, the interaction data and the finding that DMI2 and LYK6 to LYK10 phosphorylate CKL1 and CKL2 support the hypothesis that CKL1 and CKL2 are substrates for RLK signal transduction.

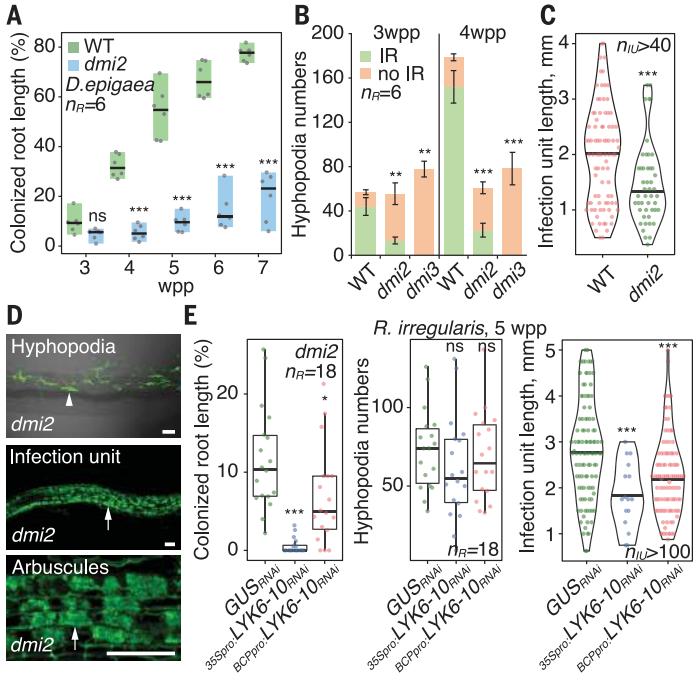


Fig. 4. DMI2 and LysM-RLKs contribute to intraradical colonization. (**A**) Limited intraradical colonization in *dmi2-7* relative to WT. n_R = 6 root systems per genotype at each time point. (**B**) Hyphopodia that penetrate the epidermis and result in intraradical colonization (IR) or fail to penetrate and lack intraradical colonization (no IR) in WT A17, *dmi2-7*, and *dmi3* (TRV25) at 3 and 4 wpp. Error bars are standard errors. (**C**) Infection unit (IU) length is shorter in *dmi2-7* relative to WT. n_{IU} > 40, number of infection units measured at 3 wpp (**P < 0.01; ***P < 0.001; t test). (**D**) Most hyphopodia fail to penetrate the epidermis (as indicated by the arrowhead). A few successful penetrations lead to intraradical colonization with arbuscules (arrow) in *dmi2-7*. Confocal microscopy images of roots stained with WGA-Alexa488 to visualize fungal structures (z-stack projection, n = 10 of optical slices, 1-µm interval). Scale bars, 100 µm. (**E**) A subset of the LysM-RLKs are required for expansion of intraradical infection units in *dmi2-7*. Effect of an RNAi construct ($LYK6\text{-}10_{RNAi}$) targeting LYK6, LYK7, LYK8, LYK9, and LYK10 expressed from constitutive *CaMV35S* promoter (*35Spro*) or symbiosis-induced, cortical cell–specific BCP1 promoter (*BCP1pro*) on the percentage of root length with intraradical colonization, number of hyphopodia, and length of infection units in *dmi2-7* at 5 wpp. Comparison made to *dmi2-7* expressing a GUS-RNAi vector control (*P < 0.05; ***P < 0.001; t test). Experiments were replicated as shown in table S8.

DMI2 and LysM-RLKs are required to support fungal colonization in the cortex

The *Medicago* LysM-RLK gene family is large, and it is likely that its members show some

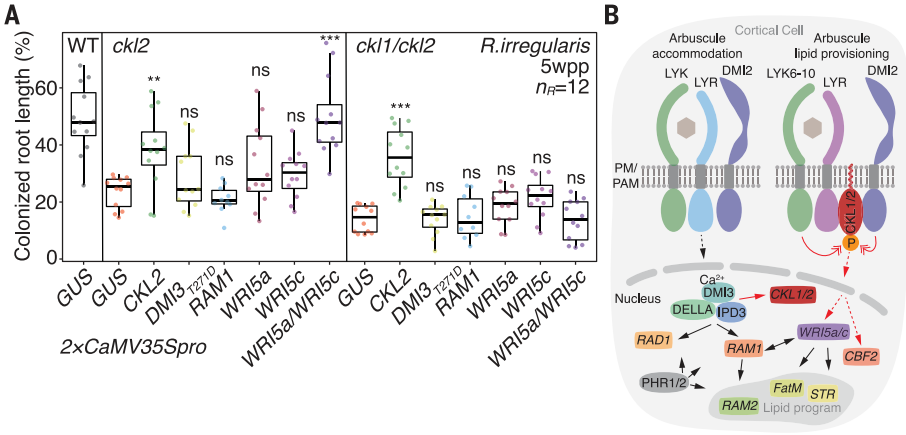


Fig. 5. CKL1 and CKL2 are components of a signaling pathway regulating the lipid provisioning program. (**A**) Simultaneous overexpression (double *CaMV35S* promoter, *2×CaMV35Spro*) of *WRI5a* and *WRI5c*, but not gain-of-function $DMI3_{T271D}$, *RAM1*, *WRI5a*, or *WRI5c*, increases intraradical colonization in *ckl2* but not in *ckl1/ckl2*. *2×CaMV35Spro:CKL2* serves as a positive complementation control and *2×CaMV35Spro:GUS* as a vector control. n_R = 12, number of evaluated transgenic root systems for each transformation (**P < 0.01; ***P < 0.001; t test). (**B**) Proposed model of CKL1 and CKL2 action. In cortical cells, initial signaling by LysM-RLKs (LYK-LYR complex) in concert with DMI2 leads to the activation of a nuclear-located DMI3-IPD3 complex, which along with DELLA induces expression of *CKL1* and *CKL2* and the transcriptional regulators *RAM1* and *RAD1*. The RAM1 regulon includes *RAM2* and, either directly or indirectly, genes involved in intracellular accommodation functions including generation of the PAM. Upon location at the PM and PAM, CKL1 and CKL2 proteins are phosphorylated by DMI2 and a subset of LysM-RLKs and transduce signaling, which ultimately leads to the transcriptional activation of several transcriptional regulators, including *WRI5a* and *WRI5c* and downstream target genes *FatM* and *STR*. The CKL1/2 signaling pathway controls the onset of the lipid provisioning program. Feedback regulation of *RAM1* by WRI5s (*16*) may further amplify the cellular accommodation pathway, including expression of *RAM2*. PHOSPHATE STARVATION RESPONSE 1 and 2 (PHR1/2) promote expression of most symbiosis-associated genes (*43*, *44*). Black solid arrows and red solid arrows indicate transcriptional regulation and refer to data published previously (black) (*9*, *12*, *27*, *40*–*44*) or presented here (red). Red double-headed arrow indicates phosphorylation. Red dashed arrow indicates unknown intermediate components.

level of functional redundancy (*33*). *Medicago lyk9* displays a modest quantitative reduction in AM fungal colonization relative to WT, and a further reduction is observed in a *lyk9/nfp* double mutant (*23*, *34*). By contrast, *dmi2* shows a strong AM symbiosis phenotype. *DMI2* and *Lotus japonicus* ortholog, *SYMRK*, are required for efficient hyphal entry through the epidermis (*27*, *35*), and most studies have focused on this epidermal function. However, occasional colonization of the root cortex has been reported in *symrk* alleles (*36*, *37*). In Medicago, *DMI2* is also required to support intraradical colonization of the cortex; colonized root lengths of *dmi2-7* and WT are similar at 3 weeks after planting, but by 7 weeks after planting, colonization levels in *dmi2-7* are <25% those of WT (Fig. 4, A and B). At 3 weeks after planting, individual infection units in *dmi2-7* are 37.5% shorter than those in WT, indicating that loss of DMI2 slows fungal growth in the cortex (Fig. 4C). Arbuscules show a WT appearance, and PT4 is present on PAM, indicating normal trafficking and PAM development in *dmi2-7* (Fig. 4D and fig. S15A), which suggest that the slower intraradical fungal growth is not the result of an impaired symbiotic interface. Slower growth could, however, result from a partial reduction in lipid provisioning. Expression of *DMI2* from an AM symbiosis–induced, cortical cell–specific promoter (*BCP1*) promotes fungal colonization in *dmi2-7* roots, further supporting a role for *DMI2* in the cortex (fig. S15B). *DMI2* is also required for symbiosis with nitrogen-fixing bacteria, where, similar to our current observations, the protein functions in both epidermal and cortical cells (*38*).

The mycorrhizal phenotypes of *ckl2* and *ckl1/ckl2* are considerably more severe than those of the RLK mutants; the differences in phenotypic severity could be explained if the CKLs integrate signals from both LysM-RLKs and DMI2. Expression of an RNA interference (RNAi) construct targeting LysM-RLKs *LYK6*, *LYK7*, *LYK8*, *LYK9*, and *LYK10* simultaneously from either the *CaMV35S* or the *BCP1* promoters did not affect hyphopodia numbers on the epidermis but did decrease total colonization and infection unit length in *dmi2-7*, indicating an additive effect of these RLKs on cortical colonization (Fig. 4E and fig. S15C). These data are consistent with the hypothesis of signal integration potentially at the CKLs. To provide further support that CKLs act downstream of DMI2 and the LysM-RLKs, we attempted to generate constitutively active CKL kinases. However, for both CKL1 and CKL2, mutation of the conserved threonine in the activation segment ($CKL1_{T284D}$ and $CKL2_{T299D}$; T, Thr; D, Asp), which in some cases causes constitutive activation, resulted in proteins unable to complement the respective *ckl* mutants (Fig. 2B). Mutation of a serine that is conserved in CKL1 and CKL2 proteins from AM symbiosis host species ($CKL1_{S312}$ and $CKL2_{S327}$; S, Ser) resulted in CKL proteins capable of complementing their respective mutants only when the mutation was phosphoablative ($CKL1_{S312A}$ and $CKL2_{S327A}$; A, Ala). The corresponding phosphomimetic ($CKL1_{S312D}$ and $CKL2_{S327D}$) mutant proteins did not complement the *ckl* mutants, suggesting potential negative regulation of the CKLs through phosphorylation of this residue (Fig. 2B). Overexpression of both *CKL1* and *CKL2*, $CKL1_{S312A}$ and $CKL2_{S327A}$, or $CKL1_{S312D}$ and $CKL2_{S327D}$ proteins did not rescue the *dmi2-7* low-colonization phenotype (fig. S15D), consistent with the hypothesis that DMI2 activates CKL1 and CKL2. However, the absence of gain-of-function CKL1 and CKL2 variants precludes a final test of this hypothesis.

Simultaneous overexpression of two WRI5 transcription factors suppresses the *ckl2* phenotype

To determine whether signaling downstream of CKLs involves proteins of the symbiosis signaling pathway, we focused on central regulators, a calcium- and calmodulin-dependent protein kinase DMI3 (CCaMK in *L. japonicus*) (*9*), INTERACTING PROTEIN of DMI3 (IPD3; CYCLOPS in *L. japonicus*) (*39*, *40*), and transcriptional regulators RAM1 (*11*, *12*) and the WRI5s (*17*, *18*), which together regulate transcription of many cellular accommodation and lipid provisioning genes. Simultaneous overexpression of *WRI5a* and *WRI5c* suppressed *ckl2*, resulting in colonization comparable to that obtained by expressing *CKL2*, whereas overexpression of a constitutively active version of $DMI3_{T271D}$ (*41*) or *RAM1* did not suppress the *ckl2* phenotype (Fig. 5A). These data, along with the *ckl2* transcript profiles, support the hypothesis that *WRI5a* and *WRI5c* act downstream of *CKL2*. However, simultaneous overexpression of *WRI5a* and *WRI5c* was insufficient to suppress *ckl1/ckl2*, likely because additional transcriptional regulators (Fig. 1F) are required to obtain full gene expression in this double-mutant background.

In summary (Fig. 5B), we propose that initial symbiosis signaling, involving the LysM-RLKs and DMI2 proteins, leads to the expression of arbuscule accommodation genes regulated through RAM1 as described (*42*) and to the expression of the *CKL* genes via DMI3, IPD3, and DELLA. CKL proteins locate at the PM and PAM, where they associate with DMI2 and

LysM-RLKs and redirect RLK signaling through a pathway independent of DMI3 and RAM1. This ultimately leads to the transcriptional activation of several regulators, including *WRI5a* and *WRI5c* and their target genes, *FatM* and *STR* (*18*), for fatty acid biosynthesis and export. As shown previously, WRI5s also increase *RAM1* expression (*18*), and, consequently, CKL signaling has the potential to further amplify the RAM1 regulon and therefore to increase the expression of *RAM2*, another central component of the lipid provisioning pathway. Thus, the data identify roles of CKL1 and CKL2 and uncover two gene modules within the lipid provisioning program whose expression is directed by two distinct but interconnected signaling pathways. As configured, a functional lipid program, which requires *FatM*, *RAM2*, and *STR* (*17*, *19*, *21*, *22*), will be expressed only coincident with arbuscule branching and not during the cellular accommodation associated with hyphal passage through cortical cells. Such mechanisms may extend to monocots, as shown by the single AM-conserved *CKL* in *Brachypodium distachyon*, whose loss-of-function phenotype, protein location at the PAM and PM, and interaction with BdDMI2 (fig. S16) mirror that of the *M. truncatula ckl1/ckl2* double mutant.

REFERENCES AND NOTES

1. A. Genre, L. Lanfranco, S. Perotto, P. Bonfante, *Nat. Rev. Microbiol.* **18**, 649–660 (2020).
2. M. Parniske, *Curr. Opin. Plant Biol.* **3**, 320–328 (2000).
3. M. J. Harrison, S. Ivanov, *Curr. Opin. Plant Biol.* **38**, 101–108 (2017).
4. A. Keymer, C. Gutjahr, *Curr. Opin. Plant Biol.* **44**, 137–144 (2018).
5. P. M. Delaux, S. Schornack, *Science* **371**, eaba6605 (2021).
6. A. Bravo, T. York, N. Pumplin, L. A. Mueller, M. J. Harrison, *Nat. Plants* **2**, 15208 (2016).
7. G. V. Radhakrishnan *et al*., *Nat. Plants* **6**, 280–289 (2020).
8. M. Menges, S. M. de Jager, W. Gruissem, J. A. H. Murray, *Plant J.* **41**, 546–566 (2005).
9. J. Lévy *et al*., *Science* **303**, 1361–1364 (2004).
10. D. S. Floss, J. G. Levy, V. Lévesque-Tremblay, N. Pumplin, M. J. Harrison, *Proc. Natl. Acad. Sci. U.S.A.* **110**, E5025–E5034 (2013).
11. E. Gobbato *et al*., *Curr. Biol.* **22**, 2236–2241 (2012).
12. P. Pimprikar *et al*., *Curr. Biol.* **26**, 987–998 (2016).
13. M. J. Harrison, G. R. Dewbre, J. Liu, *Plant Cell* **14**, 2413–2429 (2002).
14. N. Pumplin, M. J. Harrison, *Plant Physiol.* **151**, 809–819 (2009).
15. G. Carotenuto *et al*., *New Phytol.* **223**, 430–446 (2019).
16. N. Hohnjec, L. F. Czaja-Hasse, C. Hogekamp, H. Küster, *BMC Genomics* **16**, 994 (2015).
17. L. H. Luginbuehl *et al*., *Science* **356**, 1175–1178 (2017).
18. Y. Jiang *et al*., *Mol. Plant* **11**, 1344–1359 (2018).
19. A. Bravo, M. Brands, V. Wewer, P. Dörmann, M. J. Harrison, *New Phytol.* **214**, 1631–1645 (2017).
20. Q. Zhang, L. A. Blaylock, M. J. Harrison, *Plant Cell* **22**, 1483–1497 (2010).
21. A. Keymer *et al*., *eLife* **6**, e29107 (2017).
22. Y. Jiang *et al*., *Science* **356**, 1172–1175 (2017).
23. F. Feng *et al*., *Nat. Commun.* **10**, 5047 (2019).
24. C. H. Chiu, U. Paszkowski, *Plant Physiol.* **182**, 1597–1612 (2020).
25. L. Buendia, T. Wang, A. Girardin, B. Lefebvre, *New Phytol.* **210**, 184–195 (2016).
26. G. Endre *et al*., *Nature* **417**, 962–966 (2002).
27. S. Stracke *et al*., *Nature* **417**, 959–962 (2002).
28. F. Maillet *et al*., *Nature* **469**, 58–63 (2011).
29. A. Genre *et al*., *New Phytol.* **198**, 190–202 (2013).
30. M. Venkateshwaran *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 9781–9786 (2015).
31. B. J. Sieberer, M. Chabaud, J. Fournier, A. C. Timmers, D. G. Barker, *Plant J.* **69**, 822–830 (2012).
32. E. Limpens *et al*., *Science* **302**, 630–633 (2003).
33. L. Buendia, A. Girardin, T. Wang, L. Cottret, B. Lefebvre, *Front. Plant Sci.* **9**, 1531 (2018).
34. X. Zhang *et al*., *Plant J.* **81**, 258–267 (2015).
35. C. Calantzis, D. Morandi, C. Arnould, V. Gianinazzi-Pearson, *Symbiosis* **30**, 97–108 (2001).
36. C. Kistner *et al*., *Plant Cell* **17**, 2217–2229 (2005).
37. K. Demchenko, T. Winzer, J. Stougaard, M. Parniske, K. Pawlowski, *New Phytol.* **163**, 381–392 (2004).
38. E. Limpens *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10375–10380 (2005).
39. B. Horváth *et al*., *Mol. Plant Microbe Interact.* **24**, 1345–1358 (2011).
40. K. Yano *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20540–20545 (2008).
41. C. Gleason *et al*., *Nature* **441**, 1149–1152 (2006).
42. G. E. D. Oldroyd, *Nat. Rev. Microbiol.* **11**, 252–263 (2013).
43. D. Das *et al*., *Nat. Commun.* **13**, 477 (2022).
44. J. Shi *et al*., *Cell* **184**, 5527–5540.e18 (2021).

ACKNOWLEDGMENTS

The authors thank P. Lindsay [Boyce Thompson Institute (BTI)], F. Giska (BTI), and F. Schwanke (C. Zipfel lab, University of Zurich) for helpful discussions regarding in vitro transphosphorylation assays; P. Lindsay for advice on protein purification and coimmunoprecipitation techniques; and D. A. Daniels (BTI) for guidance on *Medicago* transformation. The BTI Computational Biology Center led by S. Strickler assembled 3′ RNA-seq and differential expression data and shared an R script for the heatmap. Confocal images were acquired using instrumentation at the BTI Plant Cell Imaging Center. The BTI Center for Plant Biotechnology Research generated *B. distachyon* transgenic plants. The authors thank D. Cook (University of California, Davis) for providing initial seed of *M. truncatula dmi2-7* and J. Vogel (Joint Genome Institute) for providing T3 seed of *B. distachyon* mutant line T-28188E. M.J.H. is affiliated with the Plant Pathology and Plant Microbe Biology Section of the School of Integrative Plant Sciences at Cornell University. **Funding:** US National Science Foundation Plant Genome Research IOS #2139351 (M.J.H.) and US National Science Foundation, Major Research Instrumentation NSF DBI-0618969 (M.J.H.). **Author contributions:** Conceptualization, Investigation, Visualization, and Data analysis: S.I. and M.J.H. Funding acquisition: M.J.H. Writing: S.I. and M.J.H. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. *M. truncatula* mutants and constructs generated in this study are all available under a material transfer agreement from the Boyce Thompson Institute.

SUPPLEMENTARY MATERIALS

Submitted 27 July 2022; resubmitted 16 August 2023
Accepted 19 December 2023
10.1126/science.ade1124

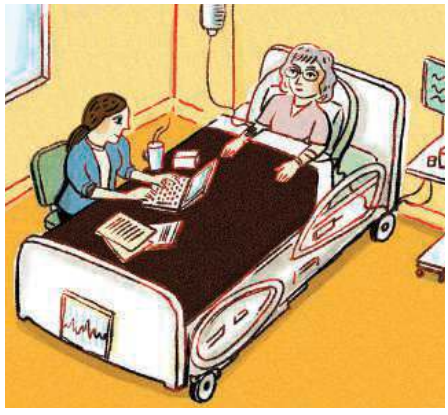
WORKING LIFE

By **Rebecca Lengnick-Hall**

In the right place

My mom, 3 months into her cancer diagnosis, was being admitted to the intensive care unit (ICU). I got the news just 10 minutes before I was scheduled to meet with my program officer about a grant to help me transition to a faculty position. It was unprofessionally last minute to cancel. But as I frantically threw things into a suitcase, I knew I wouldn't be able to give it my full attention. So, I emailed the program officer to explain the situation. She wrote back immediately, urging me to focus on my family. She reiterated her message later, when my mom was out of the ICU but I was still too distracted to meet. "The overall goal over the coming weeks is to just be a good daughter," she said. "The research can wait." Her words became my personal guidepost.

Since my mother's diagnosis in mid-2020 she has blasted through every prognosis, for which we are incredibly grateful. But as the months and years passed and I juggled my caregiving role with my professional responsibilities, I found myself increasingly falling short at work. I was missing events because I was out of town with my mother for treatments or was too exhausted to get out of bed. I didn't apply to training and funding opportunities because I couldn't focus or didn't care enough to try. I was making sloppy mistakes, forgetting things, asking for another extension, please. I feared I was squandering the career I had worked for years to achieve, letting down my mentors, collaborators, and myself.



"I was more engaged than ever with my scientific questions."

But in time, I came to feel I was in the right place—personally and professionally. As it happens, my research is in implementation science, a field that seeks to assess and improve health care delivery. With my mother's illness, I suddenly went from studying it in a sterile, removed way to being completely, heartbreakingly in the middle of it. Amid the helplessness, stress, and sadness, I found I just couldn't turn off my research brain, studying her care and texting colleagues with ideas. In a weird way, this reassured me that even though I was currently only giving my career about 50% of my attention, I was more engaged than ever with my scientific questions and excited to pursue them.

I also learned that some good could come from allowing the boundaries between my personal and work lives to blur. Before my mom's illness, I would have been mortified to cry in front my dean when she asked how I was doing. But her compassionate response, as well as the kindness I received from my program officer, were just two examples of how being authentic and vulnerable often brought out the best in people. As I became more proactive about explaining my situation, including to students and journal editors, I received empathy, understanding, and stories that reminded me I was not alone. Talking about my grief and caregiver stress helped me keep going.

The effects also helped me be a more sensitive and thoughtful colleague, mentor, and instructor. When I check in with someone or ask "how are you?" I remember that this can sometimes be a very difficult question to answer. When I experience potentially negative comments or interactions at work, I remind myself that I have no idea what this person is carrying and feeling today.

In a turn of events I can only see as comical now, my application for the grant I was supposed to discuss with my program officer that day got rejected twice. The second time happened the same day in March 2023 when my mom's doctor said the immunotherapy wasn't working and she had just weeks to live. Months later and entering 2024, my mom met her first grandchild and we had another holiday season together. My research ideas found a new funding home and I somehow got a promotion.

I no longer try to predict what will happen next. Instead, I try to focus on my newfound appreciation for the things that transcend grants and publications—such as the friendship, connection, and sometimes unexpected support that come out of being a part of a scientific community. And although I sometimes miss the clean boundaries I used to have, I've learned to embrace the blurriness. ■

Rebecca Lengnick-Hall is an assistant professor at the Brown School at Washington University in St. Louis. Send your career story to SciCareerEditor@aaas.org.

2023 AAAS Kavli Science Journalism Award winners

International awards program drew entries from a record 74 nations

By **Earl Lane**

Stories about troubling aspects of science's past as well as some hopeful signs for its future are among the winners of the 2023 AAAS Kavli Science Journalism Awards.

Presenter Adam Rutherford and producer Ilan Goodman won a Gold Award in the Audio category for a BBC series on the eugenics movement and its continuing repercussions in the modern age. Ashley Smart of Undark magazine won the Gold Award in the Science Reporting In-Depth category for a piece on the lingering impact of scientific racism, including the appropriation of legitimate genetics research for extremist ends.

On a more optimistic note, a NOVA documentary from Terra Mater Studios for PBS won a Gold Award in the Video In-Depth category for tracing the heritage and future of African astronomy through the eyes of a visionary Senegalese astronomer trying to spur the establishment of a space agency in his home country.

The Silver Award in the same category went to the "Wild Hope" series for PBS Nature from HHMI Tangled Bank Studios. The winning entry looked at a variety of habitat restoration and species recovery efforts, emphasizing the resilience of nature when given a chance.

The international awards program, endowed by The Kavli Foundation, drew entries from a record 74 countries. Among the winners were entrants from Australia, Austria, Brazil, India, Nigeria, and the United Kingdom. The winners will receive their awards at a 16 February ceremony held in conjunction with the 2024 AAAS Annual Meeting in Denver. For a description of the winning entries, with comments from the judges and the winners, go to https://sjawards.aaas.org.

Science Reporting – Large Outlet

GOLD AWARD

Lauren Sommer, Ryan Kellman, Rebecca Hersher, Connie Hanzhang Jin, and Daniel Wood, NPR

"Beyond the Poles: The far-reaching dangers of melting ice" (series)

"Why Texans need to know how fast Antarctica is melting"

"The surprising connection between Arctic ice and Western wildfires"

"The unexpected link between imperiled whales and Greenland's melting ice"
19 April 2023

SILVER AWARD

Sarah Kaplan, Simon Ducroquet, Bonnie Jo Mount, Frank Hulley-Jones, and Emily Wright, *The Washington Post*

"Hidden beneath the surface,"
20 June 2023

Science Reporting – Small Outlet

GOLD AWARD

Christine Peterson, WyoFile

"Euthanize or release? The quandaries of handling captive animals,"
8 October 2022

"Wolf killing and the consequences of disturbing pack dynamics,"
6 April 2023

"Study: Deer's lifelong fate is affected by mother's health at birth,"
24 January 2023

SILVER AWARD

Duda Menegassi, Associação O Eco (Brazil)

"A Shrinking Home: a monkey cornered by deforestation,"
22 December 2022

Science Reporting – In-Depth (More than 5,000 words)

GOLD AWARD

Ashley Smart, *Undark*

"A Field at a Crossroads: Genetics and Racial Mythmaking,"
12 December 2022

SILVER AWARD

Kemi Busari, Dubawa/ *Premium Times* (Nigeria)

"INVESTIGATION: Baba Aisha, Nigeria's fake 'doctor' cashing out on deadly concoction that cures nothing,"
10 June 2023

"NAFDAC confirms arrest of Baba Aisha's producer after PREMIUM TIMES' investigation," 14 June 2023

"Baba Aisha: NAFDAC commences nationwide mop-up of harmful concoction," 19 June 2023

Children's Science News

GOLD AWARD

Laura Allen, *Science News Explores*

"For a better brick, just add poop,"
23January 2023

SILVER AWARD

Stephen Ornes, *Science News Explores*

"Some ecologists value parasites—and now want a plan to save them,"
22 September 2022

Magazine

GOLD AWARD

Lauren Fuge, *Cosmos* (Australia)

"Point of view," March 2023

SILVER AWARD

Paul Tullis, *Bulletin of the Atomic Scientists*

"Is the next pandemic brewing on the Netherlands' poultry farms?",
26 September 2022

Audio

GOLD AWARD

Adam Rutherford and Ilan Goodman, BBC Radio 4/BBC World Service/BBC Sounds Podcast

"Bad Blood: The Story of Eugenics" (series)

"You Will Not Replace Us," 28 November 2022

"The Curse of Mendel," 19 December 2022

"Newgenics," 27 December 2022

SILVER AWARD

Wendy Zukerman, Rose Rimler, Meryl Horn, Blythe Terrell, and Michelle Dang, Science Vs on Spotify

"Superbugs: Apocalypse...Now?",
13 April 2023

Video Spot News/Feature Reporting (20 minutes or less)

GOLD AWARD

Emily Driscoll and Jeffery DelViscio, *Scientific American*

"Quest to Save the Parasites,"
13 March 2023

SILVER AWARD

Bahar Dutt, Samreen Farooqui, Vijay Bedi, Anmol Chavan, and Ajay Bedi, Roundglass Sustain (India)

"Saving the Bhimanama: Ayushi Jain and a Giant Turtle," 27 April 2023

Video In-Depth Reporting (more than 20 minutes)

GOLD AWARD

Ruth Berry and Christian Stoppacher, A NOVA/GBH production by Terra Mater Studios (Austria) for PBS

"Star Chasers of Senegal,"
8 February 2023

SILVER AWARD

Jared Lipworth, Geoff Luck, Whitney Beer-Kerr, and Matt Hill, PBS Nature from HHMI Tangled Bank Studios

"Wild Hope" (series)

"Wild Hope: The Big Oyster"

"Wild Hope: The Beautiful Undammed"

"Wild Hope: Woodpecker Wars"
12 June 2023