Chasing the ammonia economy *p. 120*

Time invested matters for mice, rats, and humans *pp. 124 & 178*

Two spindles are better than one *pp. 128 & 189*

Science

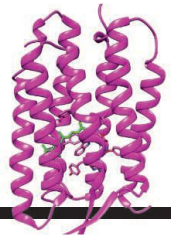
$15
13 JULY 2018
sciencemag.org

AAAS

NEUTRINOS FROM A BLAZAR

Multimessenger observations of an astrophysical neutrino source *pp. 115, 146, & 147*

CONTENTS



127 & 145
Responding to light

13 JULY 2018 • VOLUME 361 • ISSUE 6398

NEWS

IN BRIEF

IN DEPTH

FEATURE

INSIGHTS

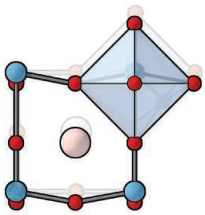
PERSPECTIVES

POLICY FORUM

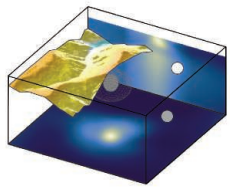
BOOKS *ET AL.*

CREDITS: (FROM TOP) P. NOGLY *ET AL.*; COAST PROTECTION BOARD, SOUTH AUSTRALIA

Published by AAAS



132 & 151
All-organic perovskites



166
Satellites locate nuclear test

ON THE COVER

Artist's impression of the IceCube Neutrino Observatory in Antarctica. Spherical digital optical modules (DOMs), each about 35 cm in diameter, are positioned up to 2.5 km deep in the ice. More than 5000 DOMs make up a cubic-kilometer detector weighing more than a billion tons. The DOMs detect the faint flash of light created when a high-energy neutrino interacts with the ice. See pages 115, 146, and 147. *Image: Jamie Yang and Savannah Guthrie/IceCube/NSF*

CREDITS: (CLOCKWISE FROM TOP LEFT) C. BICKEL/*SCIENCE*; T. WANG *ET AL.*; ISTOCK.COM/CMCNEILL17

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1808; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery): $89. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and others who use Copyright Clearance Center (CCC) Pay-Per-Use services provided that $35.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

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

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

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

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

Published by AAAS

EDITORIAL

Spain's good news

Last month finally brought good news for science in Spain. Newly elected prime minister Pedro Sánchez named Spanish astronaut and aeronautical engineer Pedro Duque as head of a new Ministry for Science, Innovation, and Universities. This will hopefully put Spanish research and development (R&D) back on the country's political agenda after two legislatures in which science languished under the Ministry of Economy. Given the minority position of the ruling party in both chambers of Congress, Duque will face extraordinary challenges to increase the budget. But the creation of this new ministry is an encouraging first step.

Over the past 7 years, Spain has suffered from insufficient R&D spending, setting its entire scientific enterprise back by two decades. In 2016, R&D represented only 1.2% of the gross domestic product (GDP), compared to 2% for the European Union; half of this was accounted for by the private sector. During this decline, most public research institutions were left nearly bankrupt [including the Spanish National Research Council (with 139 centers), the Spanish Institute of Oceanography, the Carlos III Health Institute, and the Research Centre for Energy, Environment and Technology (CIEMAT)]. Research projects became paralyzed by new bureaucratic processes; scientific talent thinned out dramatically; employment declined to precarious levels (plagued by irregular short-term contracts); and the average age of tenured academics increased to 53 years. From 2010 to 2014, Spain lost 27,358 R&D jobs—had funding been maintained at 2009 levels, about 61,940 such jobs would have been created.

Sadly, the previous Spanish government's vision of how R&D can change the economy was seriously flawed—it underestimated the importance of basic research, oversimplified knowledge transfer, minimized the time needed to bring ideas to market, and ignored how public investment attracts private investment. The new Sánchez government must move away from this ill-conceived model. Spain's prosperity will only come about through innovations based on the generation of new knowledge. Relying on the private sector to achieve the 2020 target of 3% GDP investment in R&D (proposed by the European Union in 2000) while drastically reducing public investment is unrealistic.

The creation of Spain's new Ministry for Science, Innovation, and Universities acknowledges the important role of basic research and the synergy that should exist between research and higher education in the sciences. It should be further supported by the inclusion of universities in Spain's Science, Technology and Innovation Law (to strengthen research at a national level) and by concrete measures to encourage mobility between research and educational institutions. The latter should include addressing the role of the National Agency for Quality Assessment and Accreditation in feeding academic inbreeding in Spanish universities (affecting 70% of professors; less than 2% are foreigners).

Madrid Innovation Campus, Villaverde district, Madrid, Spain.

"Spain's prosperity will only come about through innovations based on...new knowledge."

The new government certainly needs to invest more in R&D, but there are urgent measures that can be implemented quickly that require more political will than funding. These include ending bureaucratic processes imposed by the former Ministry of Finance on all research activities (contracts, procurements) that have paralyzed the system; spending all appropriated R&D funds (only 50% has been spent so far); ending the year-long delays on announcing calls for research; ending limitations on continually renewing contracts (unavoidable with the current 3-year grant system); exempting research-related contracts from lengthy competitive bidding processes; and ending the Prior Intervention Requirement that prevents research institutions from executing their budgets (some institutions have reported 50% losses). These practices were intended to sequester R&D funds to reduce spending.

The previous government compromised Spain's future by creating a deficit in innovation and discovery. Spain can no longer be trapped in this failed economic model.

– Amaya Moro-Martín

Amaya Moro-Martín is an astrophysicist and treasurer of the Spanish Association for the Advancement of Science.
amayamoromartin@gmail.com

10.1126/science.aau6630

> “A repudiation of established research … in what appears to be an effort to curry favor with commercial interests.”
>
> **Public health scholars** in a letter to the Trump administration, criticizing its attempts to alter a World Health Assembly resolution supporting breastfeeding to favor baby formula.

IN BRIEF

Edited by **Jeffrey Brainard**

CONSERVATION

Dams may threaten endangered apes

Flooding could jeopardize 1500 chimpanzees in a conservation preserve in Guinea.

Conservationists are protesting proposals to build two hydroelectric dams, one in West Africa and another in Indonesia, because both could reduce habitat for critically endangered nonhuman primates—western chimpanzees and Tapanuli orangutans, respectively. In Guinea, a management board was finalizing plans this week to construct a 294-megawatt facility in Moyen-Bafing National Park that wildlife experts say could kill up to 1500 western chimpanzees (*Pan troglodytes verus*), a subspecies whose population has fallen by more than 80% over the past 20 years. Flooding from the dam and associated human activity will force some of them to move, primatologists working with the International Union for Conservation of Nature report. Because chimpanzees are territorial, notes one of the scientists, Rebecca Kormos of the University of California, Berkeley, deadly warfare will likely break out when one group is pushed onto another’s turf. Kormos and colleagues have urged Guinea’s government to look for alternative sites for the dam or alternative energy sources. A dam planned for Sumatra in Indonesia faces similar criticism; the Tapanuli orangutan subspecies (*Pongo tapanuliensis*), discovered only last year, has a remaining population of just 800 apes.

Wellcome seeks long shots

RESEARCH FUNDING | The world’s second biggest private funder of medical research, the London-based charity Wellcome Trust, announced this week that it will provide £250 million for unconventional research that could transform science or health. The money is earmarked for ideas shut out from traditional funding because they are less likely to succeed, and for projects with researchers who lack a background in the life sciences but bring other expertise. Called the Wellcome Leap Fund, it will be set up as a subsidiary of the Wellcome Trust with an independent board; the search for a CEO will start shortly and its first research programs could start in late 2020. The fund, to be established over 5 years, will account for about 5% of Wellcome’s total spending during that period.

Death from nerve agent probed

CHEMICAL WEAPONS | A second incident involving a Novichok nerve agent in the United Kingdom has left one victim dead—and set abuzz the community of scientists who study countermeasures to chemical weapons. On 30 June, a couple fell ill in Amesbury, U.K., about 12 kilometers north of Salisbury, where, 4 months earlier, ex-spy Sergei Skripal and his daughter Yulia spent weeks in intensive care after exposure to a Novichok agent first developed by the Soviet Union decades ago. U.K. authorities say it’s too early to say how the incidents are related. Dawn Sturgess, 44, died on 8 July; her partner, Charlie Rowley, 45, remains critically ill. Investigators have not revealed whether the agent that poisoned them is the same compound used against the Skripals and if so, whether it is from the same batch. Novichok compounds stored in powder form are thought to be stable, but degrade quickly when exposed outdoors. The notion that the same agent might have lingered in the environment for months is “incredible, and deeply disturbing,” a U.S.-based chemical defense researcher says.

PHOTO: KALYANEE MAM

Published by AAAS

A high-fat diet may have helped Ötzi the Iceman survive at high altitudes.

PHOTO: SOUTH TYROL ARCHAEOLOGY MUSEUM\EURAC\M. SAMADELLI

BIOANTHROPOLOGY

Iceman's last meal analyzed

Researchers have provided a detailed chemical analysis of the last meal eaten by Ötzi the Iceman—the 5300-year-old man found frozen and naturally mummified in the Italian Alps in 1991—and found it was rich in fat. As described in previous accounts, Ötzi had a completely full stomach when he was shot with an arrow and died. In the new study, published 12 July in *Current Biology*, scientists quantified the ancient DNA, proteins, and other chemicals preserved in his stomach contents. Fat residues, which came from the red deer and ibex he ate, made up a remarkably high share of the total, about 50%, they determined; although one meal doesn't reveal a lifetime's diet, a high-fat diet may have given Ötzi the energy he needed to survive at high altitudes. The iceman's last meal was balanced with grains from einkorn wheat and traces of a genus of toxic ferns called bracken. Researchers say he may have consumed the bracken as medicine to treat parasites previously identified in his intestines.

Puerto Rico's stats agency saved

SCIENCE POLICY | Puerto Rico has backed away from a controversial plan to eliminate the Puerto Rican Institute of Statistics (PRIS), the agency that vets data and statistical analyses about the island. Governor Ricardo Rosselló had proposed bringing PRIS under the control of the island's Department of Economic Development and Commerce and ultimately outsourcing its responsibilities to the private sector, as part of a broad reorganization plan. Critics—including AAAS, which publishes *Science*—argued that the bill would undermine PRIS's political independence and the trustworthiness of Puerto Rico's statistics. On 30 June, the island's legislature approved the reorganization of the economic department but preserved PRIS in its current form. The move won praise from Lisa LaVange, president of the American Statistical Association in Alexandria, Virginia. "Having independent, strong, and transparent statistics will serve the health, well-being, and economy of Puerto Rico and its people for decades to come," she says.

Global science body launched

SCIENTIFIC SOCIETIES | A new international body aiming to be the "global voice for science" was launched at a meeting in Paris last week. The inauguration of the International Science Council (ISC) follows a decision last year to merge two existing bodies representing the social and natural sciences—the International Social Science Council and the International Council for Science—to pool resources and break down barriers between disciplines. The ISC, which will focus on supporting research related to international issues such as poverty and climate change, counts more than 180 organizations as members, including the U.S. National Academy of Sciences and The Royal Society of the United Kingdom. At the 4 July meeting, South African mathematician Daya Reddy was elected ISC's first president.

BY THE NUMBERS

40

Countries in which people posted on social media during the first week of July celebrating the first International Day of LGBTQ+ People in Science, Technology, Engineering, and Math on 5 July.

500 million

Tons of insects consumed by birds per year, according to an estimate in a study that examined habitats around the world and noted the importance of bird populations for keeping insect pests in check (*The Science of Nature*).

Ancient tools found in China

ARCHAEOLOGY | Hominins lived in China more than 2 million years ago, according to new evidence that may push back the earliest dates for hominins living outside Africa by 250,000 years. If the dates hold, they will boost a minority view that a key human ancestor, *Homo erectus*, could have arisen in Asia, not Africa. In this week's *Nature,* researchers report that they recovered and analyzed 96 stone points, flakes, and cores dropped in a gully at Shangchen in north-central China's windswept Loess Plateau. Until now, the oldest known site for hominins outside of Africa was Dmanisi, Georgia, where fossil hominins and stone tools date to 1.85 million years. The Shangchen site lacks volcanic minerals, which provide the gold standard for radiometric dating methods. Instead, the scientists used paleomagnetic dating—which detects known reversals in Earth's magnetic field that are recorded in ancient rock—to date the layers of sediment containing the stone tools. The oldest layers date from 2.1 million years ago.

Church backs climate action

CLIMATE SCIENCE | The Church of England will, by 2023, begin to divest from companies that aren't taking enough action against climate change, leaders agreed this week. The decision follows other steps by the church to influence the energy industry by using its £12 billion investment fund, which contains about £125 million in shares of large oil and gas companies. In 2015, the church divested £12 million of its assets from tar sands and coal projects. Last year, the church teamed up with the London School of Economics to create the Transition Pathway Initiative, which has been evaluating efforts by companies to help reach the Paris climate agreement goal of keeping global warming to less than 2°C. The church originally planned to start divesting from companies by 2020 if they didn't pass muster, but leaders this week voted to delay that by 3 years to allow more time to pressure companies to change their ways. The climate divestment would be the largest by any church fund.

Carrier ends monkey flights

ANIMAL RESEARCH | The Russian carrier AirBridgeCargo Airlines will no longer transport nonhuman primates for research purposes, leaving Air France as the only major airline still carrying the animals. A representative of the Moscow-based company disclosed the decision in a 3 July email to People for the Ethical Treatment of Animals (PETA) in Norfolk, Virginia. In late June, PETA urged its supporters to ask AirBridgeCargo to stop transporting primates, mostly macaques, bound for research labs. More than 188,000 people emailed the airline. It had been shipping animals from China to Chicago O'Hare International Airport. Airline officials did not respond to emails requesting comment. Most primates imported to the United States for research originate in China; others arrive from Vietnam and Mauritius. In 2014, the last year for which the U.S. Fish and Wildlife Service made such information available, 23,426 research primates were imported.

New leader at EPA

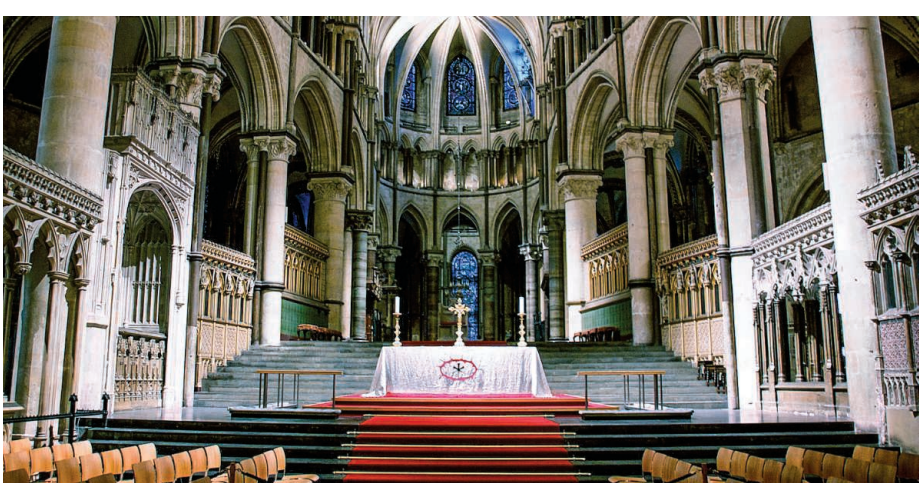
SCIENCE POLICY | Andrew Wheeler, a former energy industry lobbyist and U.S. Senate aide, became acting head of the Environmental Protection Agency (EPA) this week, replacing Scott Pruitt, who resigned on 5 July after becoming engulfed in ethics scandals. Wheeler, a polished Washington insider who was confirmed as EPA deputy administrator in April, is expected to forge ahead with many of Pruitt's initiatives, including efforts to roll back climate and wetlands protection rules. But his arrival likely marks the demise of one dream of EPA foes: undermining the so-called endangerment finding, a 2009 legal opinion that obligates EPA to regulate carbon dioxide emissions. Wheeler has called the finding "settled law."

Newborn screening list amended

PUBLIC HEALTH | Health and Human Services (HHS) Secretary Alex Azar last week added spinal muscular atrophy (SMA) to the list of disorders for which the government recommends that newborns be screened. States implement the screening and generally follow the federal government's lead; with the addition of the fatal, inherited disease, which can be detected by a genetic test, HHS now suggests screening for 35 such conditions. SMA destroys motor neurons in the brain stem and spinal cord. In its most common, most severe form, children die in the first few years of life. In December 2016, the Food and Drug Administration (FDA) approved the first therapy, nusinersen (Spinraza), made by Biogen of Cambridge, Massachusetts. In February, an advisory committee had voted 8 to 5 to recommend that Azar add the disease to the screening panel, but dissenters worried about the lack of long-term data on the drug's effectiveness and side effects (*Science*, 29 June, p. 1385).

Ontario head shrinks science

SCIENCE POLICY | Researchers in Ontario, Canada's most populous province, are concerned by a sudden shift in the government's approach to science and evidence-based policy following the election of populist Doug Ford as the province's premier. Since taking power on 29 June, Ford, the brother of the late Toronto Mayor Rob Ford, has canceled the province's carbon cap-and-trade system; eliminated the position of minister of science and innovation in his Cabinet; and fired the province's chief scientist, who had been on the job only since November 2017. "It's unfortunate to see Premier Ford shift robust evidence and science-based policy down his list of priorities," said Kathleen Walsh, director of policy at the science campaign group Evidence for Democracy in Ottawa.



Canterbury Cathedral is the seat of the Church of England, which is mulling divesting funds over climate change.

PHOTO: BIRGER HOPPE/FLICKR/CC BY-NC-ND

SCIENCEMAG.ORG/NEWS
Read more news from *Science* online.

IN DEPTH

The IceCube Laboratory at the South Pole operates a neutrino detector 1.5 kilometers beneath the ice.

ASTROPHYSICS

Ice reveals a messenger from a blazing galaxy

Neutrino astronomy is born as IceCube and other instruments find likely home of particle

By **Daniel Clery**

If astronomers are right, a ghostly particle that lit up an instrumented swathe of ice beneath the South Pole on 22 September last year was a messenger from a distant galaxy. The particle was a neutrino, electrically neutral and almost massless, which means its path could be traced back to the extragalactic event that created it. Cued by IceCube, the Antarctic detector, the orbiting Fermi Gamma-ray Space Telescope found that the neutrino likely came from a far off blazar, a hugely bright source of radiation powered by a supermassive black hole.

Astronomers have long been tantalized by the prospect of using neutrinos, which move at almost the speed of light and rarely interact with other matter, to learn about violent cosmic events. The new finding, reported on pages 146 and 147, could mark the founding event of neutrino astronomy. The detection also triggered a powerful example of another new trend, multimessenger astronomy, in which telescopes and other instruments studied the flaring blazar in all parts of the electromagnetic spectrum, from gamma rays to radio waves.

A neutrino-producing blazar could also help solve a decades-old mystery in astronomy: Where do the extremely high energy protons and other nuclei that occasionally bombard Earth come from? Known as ultrahigh-energy cosmic rays, these particles have a million times more energy than has ever been produced in an earthbound particle accelerator, but what boosts them to such colossal energies is unknown. Suspects have included neutron stars, gamma-ray bursts, hypernovae, and the radiation-spewing black holes at the center of some galaxies, but whatever the source, high energy neutrinos are a likely byproduct (*Science*, 31 July 2015, p. 465). If the IceCube team is right, blazars could be the first confirmed source of these cosmic rays.

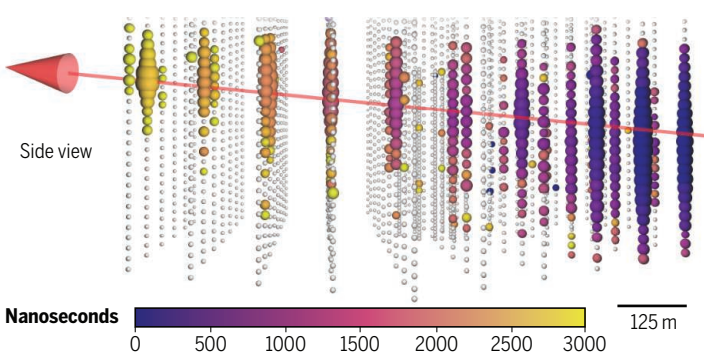
Researchers note, however, that the link between the neutrino and the blazar isn't rock-solid. "It's a very mouthwatering observation and I very much hope it will be confirmed," says Pierre Sokolsky of the University of Utah in Salt Lake City. "If their interpretation of those observations is correct it will be revolutionary, extraordinary," says Eli Waxman of the Weizmann Institute of Science in Rehovot, Israel. But, he adds, "an extraordinary result needs extraordinary support, and the support is not quite extraordinary yet."

Completed in 2010, the IceCube neutrino detector snares these elusive particles in a cubic kilometer of Antarctic ice. When a neutrino hits a nucleus in the frozen water molecules, other particles fly off in recoil; as they decelerate, they emit light called Cherenkov radiation, which some of IceCube's 5160 light detectors may pick up. Based on the location, timing, and brightness of the detected light, researchers can reconstruct the path and energy of the neutrino.

Most of the neutrinos detected by IceCube originated nearby, spawned by cosmic rays hitting Earth's upper atmosphere. IceCube researchers eliminate those using a variety of methods, leaving the very few, very high energy neutrinos, above 30 trillion electron volts (TeV). In 2013, the IceCube team first revealed a handful of such events, arguing that their high energies and other properties showed they must have come from outside of our galaxy (*Science*, 22 November 2013, p. 920). The detector continues to bag about a dozen high energy neutrinos a year; when it gets a clean track with a well-defined direction, other telescopes scramble to see if

A polar light show with cosmic origins

The detection, and computed path, of neutrino IceCube-170922A. Each circle represents one of IceCube's spherical light detectors in the Antarctic ice: size indicates the brightness detected, from earliest (in dark blue), to latest (in yellow).



CREDITS: (PHOTO) MARTIN WOLF, ICECUBE/NSF; (GRAPHIC) GRANT *ET AL.*, *SCIENCE*, VOL. 361, 147 (2018); ADAPTED BY C. BICKEL/*SCIENCE*

there is an obvious cosmic source—until now, without success.

In 2016 IceCube's operators set up an alert service, with the hope of getting more telescopes at different wavelengths involved in the hunt. Then, last September, IceCube got lucky. A detected neutrino, dubbed IceCube-170922A and calculated to have an energy of 290 TeV, offered a relatively clear track back into space. An automatic alert went out less than a minute later.

Several observatories initially didn't see anything unusual. Six days later, the Fermi team reported the satellite had found that a blazar, known as TXS 0506+056 and just 0.1° away from the neutrino track suggested by IceCube, was especially bright, having started flaring a few months earlier. Soon, more than a dozen telescopes had studied the blazar. Blazars, like quasars, are distant cosmic beacons powered by supermassive black holes, which generate intense radiation and fire jets of particles from their poles. Blazars are exceptionally bright, astronomers believe, because their jets happen to be aimed straight at Earth.

IceCube and the other observers estimate the probability that the neutrino path and the blazar coincided by chance is roughly one in 740. Physicists and astronomers, however, aren't usually convinced that two phenomena are connected until there's no more than a one in 3.5 million, or 5-sigma, probability of a coincidence.

IceCube researchers also went back through almost a decade of data to see whether an excess of high-energy neutrinos had streamed from the same location before. They found a period of 150 days in late 2014 and early 2015 when IceCube detected around 13 more neutrinos than normal from that spot. It's not yet clear whether TXS 0506+056 was flaring at that time, but "the archival event was much more interesting" than the recent detection, says IceCube Principal Investigator Francis Halzen of the University of Wisconsin in Madison.

Sokolsky and Waxman agree that IceCube's September 2017 detection should strengthen the project's longstanding bid to massively increase the size of the instrument, which would also increase how many neutrinos it can detect and improve its pointing accuracy. Since IceCube was built, the team has found the ice is clearer than previously thought, so they believe they can make IceCube 10 times bigger while only doubling the number of light detectors, matching the $280 million cost of its original construction. The team is about to start experiments to test that. "With a 10-times bigger detector, the answer [to where high energy neutrons come from] would be clear and obvious," Waxman says. ■

IMAGE: INSTITUTE OF HIGH ENERGY PHYSICS, CAS

SPACE

China announces new flotilla of space science missions

Probes will watch for colliding black holes, image Earth's magnetosphere, and monitor the sun

By **Dennis Normile**

China's ambitious human space missions get most of the headlines, but its fledgling space science program is quietly gaining strength. The Chinese Academy of Sciences (CAS) last week confirmed plans to launch four new scientific satellites beginning in 2020. Coming on the heels of four successful missions, including one devoted to x-ray astronomy and another that demonstrated quantum entanglement over a record-setting 1200 kilometers, these "phase 2" projects will examine areas including solar physics and the hunt for electromagnetic signals associated with gravitational waves.

Given that China's space science program only started about 10 years ago, the lengthening track record "is impressive, but there are still not many missions given that it's a big country with a big science community," says Xin Wu, a China-born physicist at the University of Geneva in Switzerland who collaborates on China's astrophysics missions. "There is pent-up demand" among Chinese space scientists, he says.

CAS broke with tradition for one of the new missions, the Gravitational Wave High-energy Electromagnetic Counterpart All-sky Monitor (GECAM). It fast-tracked selection and development to take advantage of a new scientific opportunity, which Xiong Shaolin, an astrophysicist at CAS's Institute of High Energy Physics in Beijing, and his colleagues identified a month after the U.S. Laser Interferometer Gravitational-Wave Observatory announced its historic detection of gravitational waves in February 2016. They proposed putting two satellites into orbit on opposite sides of Earth that together could watch the entire sky for gamma rays emanating from the events that generate gravitational waves. Funding for technical studies arrived a few months later, and the mission has jumped to the front of the launch queue, with a date of 2020. "When you have this kind of opportunity you can't handle it like a normal mission, with selection and review taking 10 or 20 years," Xiong says.

So far, gamma rays and other electromagnetic signals have only been detected from one kind of gravitational-wave source, a neutron star merger, but they yielded a trove of detail about the enigmatic event. Astrophysicists are still debating whether black hole mergers, the other confirmed source of gravitational waves, also produce electromagnetic emissions. The GECAM team is betting that they do—and that much can be learned from the signals. "I think probably we will find something," Xiong says.

GECAM's observations will complement those of another phase 2 mission, the Einstein Probe (EP), which will survey the sky

Two satellites, launching in 2020, will watch for gamma rays from the violent birth of gravitational waves.

for the low energy x-rays associated with violent phenomena such as gamma ray bursts and black hole collisions. Combining GECAM, EP, and gravitational wave observations "will allow us to better understand gamma ray burst astrophysics," says Ik Siong Heng, an astrophysicist at the University of Glasgow in the United Kingdom.

China's space scientists have long targeted another area: solar physics. Only the United States produces more papers in the field than China. "But [China's] papers used data from missions developed by Japan and the U.S. and elsewhere," says Gan Weiqun, a solar physicist at CAS's Purple Mountain Observatory in Nanjing. He says China's solar scientists have been pushing for their own mission for 40 years; they've finally gotten the nod for the Advanced Space-based Solar Observatory (ASO-S). "It's very important for us to make original contributions in terms of hardware and data," Gan says. He explains that ASO-S will be the first space observatory to monitor the sun's magnetic field while watching for solar flares and the titanic blasts known as coronal mass ejections. The simultaneous observations could yield clues to how those eruptions are triggered.

The last mission included in the phase 2 list was identified as a priority years ago (*Science*, 22 July 2016, p. 342). The Solar wind Magnetosphere Ionosphere Link Explorer (SMILE), a joint CAS and European Space Agency mission, will pioneer a new technique for imaging Earth's magnetosphere. Previous satellites have made point measurements as they traveled through the magnetosphere. But scientists recently learned that collisions between particles of the solar wind and stray particles of Earth's atmosphere produce low energy x-rays that light up the magnetosphere. By watching these x-rays, SMILE will capture its dynamic behavior.

Last week's announcement bodes well for China's space science program beyond the next four missions. The 4 billion yuan ($605 million) phase 2 budget includes support for development on future missions, particularly the enhanced X-ray Timing and Polarimetry mission (*Science*, 9 March, p. 1085), an ambitious international project led by Chinese scientists to study black holes, neutron stars, and magnetars.

China's planetary exploration and astronaut programs will continue to make history; later this year or early next, for example, it plans to land the first probe on the far side of the moon. But the future of its space science efforts also seems assured. ■

TOXICOLOGY

Digital chemical test impresses

Giant database shows promise for replacing animal studies

By **Vanessa Zainzinger**

Toxicologists this week unveiled a digital chemical safety screening tool that could greatly reduce the need for six common animal tests. Those tests account for nearly 60% of the estimated 3 million to 4 million animals used annually in risk testing worldwide.

The computerized tool—built on a massive database of molecular structures and existing safety data—appears to match, and sometimes improve upon, the results of animal tests for properties such as skin sensitization and eye irritation, the researchers reported in the 11 July issue of *Toxicological Sciences*. But it also has limitations; for instance, the method can't reliably evaluate a chemical's risk of causing cancer. And it's not clear how open regulatory agencies will be to adopting a nonanimal approach.

> ***"This won't be the end of animal testing."***
> **Mike Rasenberg**, European Chemicals Agency

Still, "We're really excited about the potential of this model," says toxicologist Nicole Kleinstreuer, the deputy director of a center that evaluates alternatives to animal testing at the National Institute of Environmental Health Sciences (NIEHS) in Durham, North Carolina. Kleinstreuer, who was not involved in the work, adds that using "big data … to build predictive models is an extremely promising avenue for reducing and replacing animal testing."

Most developed nations require new chemicals that enter commerce to undergo at least some safety testing. But the long-standing practice of exposing rabbits, rats, and other animals to chemicals to evaluate risks is facing growing public objections and cost concerns, helping spur a hunt for alternatives.

One approach is to use what is known about the safety of existing chemical compounds to predict the risks posed by new chemicals with similar molecular structures. Two years ago, a team led by Thomas Hartung of the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, took a step toward that goal by assembling test data on 9800 chemicals regulated by the European Chemicals Agency (ECHA) in Helsinki. They then showed that chemicals with similar structures can have similar health effects, such as being an irritant (*Science*, 12 February 2016, p. 651).

In their new paper, Hartung's team goes bigger. First, the researchers expanded their database to 10 million chemical structures by adding information from the public database PubChem and the U.S. National Toxicology Program. Next, they compared the structures and toxicological properties of every possible pair of compounds in their database—a total of 50 trillion comparisons—creating a vast similarity map that groups compounds by structure and effect. Finally, they tested the model: They asked it to predict a randomly chosen chemical's toxicological profile by linking it to similar "neighbors" on the map, and compared the results to six actual animal tests of the compound.

On average, the computational tool reproduced the animal test results 87% of the time. That's better than animal tests themselves can do, Hartung says: In reviewing the literature, his group found that repeated animal tests replicated past results just 81% of the time, on average. "This is an important finding," Hartung says, because regulators often expect alternative methods to animal testing to be reproducible at the 95% threshold—a standard even the animal tests aren't meeting.

The screening method has weaknesses. Although it can predict simple effects such as irritation, more complex endpoints such as cancer are out of its reach, says Mike Rasenberg, who heads the ECHA's Computational Assessment & Dissemination unit. "This won't be the end of animal testing," he predicts. Still, Rasenberg thinks European regulators will accept the method for simple endpoints.

In the United States, the NIEHS center is validating the method, and Kleinstreuer says if it passes, she could see U.S. regulators adopting it for some purposes. In the meantime, the researchers are moving to make the tool available to companies that might want to screen products before submitting them for regulatory review. ■

Vanessa Zainzinger is a science journalist in the United Kingdom.

The genome editor CRISPR can be used to engineer female lab mice that have increased odds of passing down a specific gene to offspring.

GENOME EDITING

A 'gene drive' makes its debut in mammals

CRISPR-based gene-spreading strategy promises to speed development of engineered mice

By **Jon Cohen**

Researchers have used CRISPR, the genome editing tool, to speed the inheritance of specific genes in mammals for the first time. Demonstrated in lab-reared insects several years ago, this controversial "gene-drive" strategy offers the ability to quickly spread a gene throughout an entire species. It has sparked dreams of deploying lethal genes to eradicate pests such as malaria-carrying mosquitoes—and now, perhaps, crop-damaging, disease-causing mammals such as rabbits, mice, and rats.

The new research aims to create novel strains of lab mice, not wipe out wild populations, and it shows that gene drives work less efficiently in rodents than in insects. Still, Paul Thomas, a molecular geneticist at the University of Adelaide in Australia who is working on similar experiments, calls it an "important first step towards development of gene drive technology in mammals."

The study was posted 4 July on bioRxiv, an online site for preprints, by a team at the University of California, San Diego (UCSD), led by geneticist Kimberly Cooper. The researchers, among them Ethan Bier and Valentino Gantz, who 3 years ago showed that CRISPR could create an efficient gene drive in flies (*Science*, 24 April 2015, p. 442), have submitted the study to a peer-reviewed journal that asked them not to speak with media. But it has already triggered plenty of scientific discussion. "It's a very good study and it's of pretty high significance," says Gaétan Burgio, a mouse geneticist at the John Curtin School of Medical Research in Canberra, Australia, who tweeted a series of comments about the report. "Nothing is really known about gene drives in rodents. We all assumed the efficiency would be the same as in flies, but it turns out to be very different."

In a variation of the earlier fly work, the UCSD researchers built their gene drive by engineering female mice to carry the gene for the DNA-cutting enzyme Cas9, one of CRISPR's two components; they engineered males to carry a gene for the other component, the guide RNA (gRNA) that shuttles Cas9 to a specific target on a genome, plus a gene that modifies coat color. Breeding the altered mice created pups that had the genes for both CRISPR components on different chromosomes.

After Cas9 makes its cut, a cell repairs the damage—and how it does so is key to the success of a gene drive. The cell can either reconnect the severed strands of DNA or bridge the gap by slotting in chunks of new DNA, a process called homology directed repair (HDR). A gene drive harnesses HDR to insert a new gene—in this case, a copy of the coat-color modifier gene. But cells naturally prefer to simply reconnect the severed DNA, which thwarts a gene drive.

The UCSD team exploited a fundamental biological phenomenon to force cells toward HDR. They manipulated Cas9 to turn on during meiosis, the cell division process that helps creates sperm or eggs. The chromosomes naturally swap DNA during meiosis, and during those exchanges the cell only allows HDR.

The experiment did not work in males, likely because spermatogonia go through mitotic cell division before meiosis, stymying HDR. But in females, the gene drive succeeded. It copied the coat color modifying gene to the partner chromosome in many eggs, which would significantly raise the odds of offspring inheriting it. In one mouse, 79% of her eggs ended up with the color modifying gene on both chromosomes. If she mated with a male without the gene, about 90% of her pups would inherit the gene. (Flies have a different process of embryogenesis, one that boosts the efficiency of HDR and allows it to work in both sexes.)

Cooper and her colleagues write that this system of "active genetic elements" could speed the creation of mice that have several introduced or crippled genes. Michael Wiles, who heads tech evaluation and development at the Jackson Laboratory in Bar Harbor, Maine—one of the world's largest commercial producers of engineered mice—says the method could be "very useful," as many human diseases are caused by aberrations in several genes, and producing mice that genetically mimic those diseases is slow and laborious. "I'm now getting requests to make mice with six modifications, and the breeding time becomes phenomenal," Wiles says. With a gene drive like this, what takes 5 years could be done in one, he says.

Even though this new work aims only to engineer lab mice, Kevin Esvelt, an evolutionary biologist at the Massachusetts Institute of Technology Media Lab in Cambridge, says it concerns him. He believes a gene drive that could be released into the wild should include a kill switch to shut it down and restore animals to their natural state. "It's troubling to see that the study does not explicitly mention safeguards," Esvelt says.

The new UCSD gene drive, however, likely would stop spreading in a mouse population after a few generations. Because the genes for Cas9 and the gRNA ride on different chromosomes, they would gradually become separated and the drive would lose effectiveness. In the preprint, Cooper and her colleagues stress the continuing challenges of creating an efficient gene drive for wild mammals. "[T]he optimism and concern that gene drives may soon be used to reduce invasive rodent populations in the wild is likely premature," they conclude. ■

PHOTO: ISTOCK.COM/GORKEMDEMIR

ARCHAEOLOGY

Molecular 'barcodes' reveal lost whale hunts

DNA method can rapidly identify species, shows toll of ancient cultures on biodiversity

By **Elizabeth Pennisi** *and* **Michael Price**

According to myth, the Maori people arrived in New Zealand riding the back of a whale; today, whales figure prominently in Maori art and stories. Now it appears that Maori ancestors may also have systematically hunted the animals.

Archaeological sites in New Zealand hold few recognizable whale bones, but a new technique for detecting species from short snippets of DNA in bone scraps has enabled researchers to identify the remains of five whale species in early New Zealand settlements, including smaller or slow-moving species, the likely targets for hunters. The same "DNA metabarcode" method showed that the early New Zealanders may also have decimated populations of a native parrot. And on the opposite side of the world, a team used a simpler barcode method to suggest that the Romans hunted whales too, potentially pushing the first known commercial whaling industry back by about a thousand years.

By revealing species from once-meaningless bone fragments, barcodes are "opening new windows into historical ecosystems" says Ana Rodrigues, an ecologist at the University of Montpellier in France. And the new data are changing views. "For a long time, we thought that ancient cultures were technologically not advanced enough to significantly influence animal populations, but more and more evidence is accumulating that they actually did," says evolutionary biologist Michael Hofreiter of the University of Potsdam in Germany.

At Curtin University in Perth, Australia, DNA researcher Michael Bunce, graduate student Frederik Seersholm, and colleagues have been assessing the impact of human arrival in New Zealand in about 750 B.C.E. They developed a way to cheaply and easily determine the DNA barcodes of hundreds of bones by high throughput sequencing of blends of ground-up bones. Barcodes, named after the ubiquitous labels on items in stores, are short stretches of DNA that uniquely identify a species. Bunce's team applied their method to 5000 bone fragments from 38 sites in New Zealand spanning the past 20,000 years. "It's bones that never would have been identified," says Rodrigues, who was not part of this study.

Seersholm pinpointed 110 near-perfect matches to known species, as the group reported this week in the *Proceedings of the National Academy of Sciences*. Until now, archaeologists usually assumed that whale bones at sites in New Zealand were from pilot whales that had been scavenged after they beached. But the DNA barcodes identified orca, true dolphins, Cuvier's beaked whales, fin whales, and southern right whales. "We think they were hunting the smaller whales," driving the mammals into shallow water with boats and then harpooning them, Seersholm says.

Scientists retrieved the DNA signatures of whales and other species from bone fragments like these in New Zealand.

PHOTO: PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES

The barcodes also suggested the early New Zealanders hunted fur seals, sea lions, and elephant seals. And the method identified several terrestrial species including the kākāpō, a now-endangered ground-dwelling New Zealand parrot. Further analysis of kākāpō DNA showed that its genetic diversity plunged once New Zealand was settled. "It really shows that high impact humans had on the biodiversity," Seersholm says.

In the Gibraltar region by the Mediterranean Sea, Rodrigues and her colleagues took a focused approach, analyzing 11 putative whale bones from five sites. There, the Romans fished tuna from villages dating to between 400 B.C.E. and 425 C.E. Cryptic texts referring to "sea monsters" or "ram fishes," plus large bones suspected to be whale (and often shaped into tools), suggested the Romans also hunted coast-hugging gray and right whales. But these behemoths don't ply those waters today.

Rodrigues's team combined tried-and-true barcoding with collagen fingerprinting, a technique that matches amino acids in samples' connective tissue to known animal families. They found that 10 of the 11 bones belonged to whales—three gray, three right, one fin, one pilot, one sperm, and one dolphin. The preponderance of gray and right whales, which are reliably found close to shore and thus could have been hunted, raises the possibility of a forgotten 2000-year-old Roman whaling industry, the researchers suggest this week in the *Proceedings of the Royal Society B*. That's about 1000 years before Basque whalers, long considered the first commercial whalers, emerged.

If the Romans did hunt gray and right whales from Gibraltar, the authors note, they may have played a role in the species' disappearance from the eastern North Atlantic. But other scholars say it's more likely that the Romans scavenged beached whales.

The identification of the 11th presumed whale bone threw the researchers for a loop: It belonged to an elephant. That's "why it's good to use these methods," Rodrigues says. "Who knows how many other bones out there have been misidentified?"

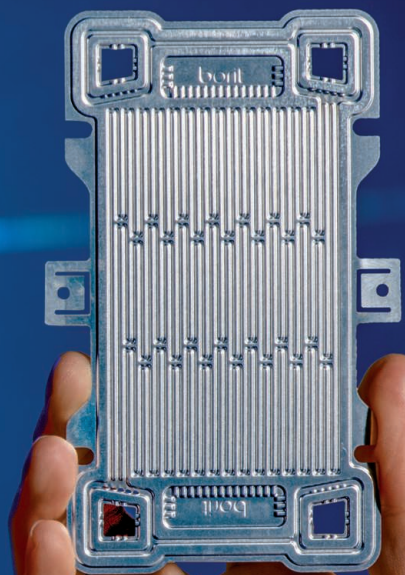
Barcoding has limitations. It can't identify a species if its sequence isn't included in the barcode database, which is rapidly expanding but still incomplete. And barcodes can't yet show how abundant a species was—only that it was present.

But Rodrigues, for one, is in awe of barcoding's power. "It's like a miracle," she says. "You send them a little bit of dust and they send you back results." ■

Michael Price is a freelance journalist and a magazine editor at San Diego State University in California.

LIQUID SUNSHINE

Ammonia made from sun, air, and water could turn Australia into a renewable energy superpower



By **Robert F. Service**
in Sydney, Brisbane, and Melbourne, Australia

The ancient, arid landscapes of Australia are fertile ground for new growth, says Douglas MacFarlane, a chemist at Monash University in suburban Melbourne: vast forests of windmills and solar panels. More sunlight per square meter strikes the country than just about any other, and powerful winds buffet its south and west coasts. All told, Australia boasts a renewable energy potential of 25,000 gigawatts, one of the highest in the world and about four times the planet's installed electricity production capacity. Yet with a small population and few ways to store or export the energy, its renewable bounty is largely untapped.

That's where MacFarlane comes in. For the past 4 years, he has been working on a fuel cell that can convert renewable electricity into a carbon-free fuel: ammonia. Fuel cells typically use the energy stored in chemical bonds to make electricity; MacFarlane's operates in reverse. In his third-floor laboratory, he shows off one of the devices, about the size of a hockey puck and clad in stainless steel. Two plastic tubes on its backside feed it nitrogen gas and water, and a power cord supplies electricity. Through a third tube on its front, it silently exhales gaseous ammonia, all without the heat, pressure, and carbon emissions normally needed to make the chemical. "This is breathing nitrogen in and breathing ammonia out," MacFarlane says, beaming like a proud father.

Companies around the world already produce $60 billion worth of ammonia every year, primarily as fertilizer, and MacFarlane's gizmo may allow them to make it more efficiently and cleanly. But he has ambitions to do much more than help farmers. By converting renewable electricity into an energy-rich gas that can easily be cooled and squeezed into a liquid fuel, MacFarlane's fuel cell effectively bottles sunshine and wind, turning them into a commodity that can be shipped anywhere in the world and converted back into electricity or hydrogen gas to power fuel cell vehicles. The gas bubbling out of the fuel cell is colorless, but environmentally,

PHOTO: STEVEN MORTON/FELLOW OF THE ROYAL PHOTOGRAPHIC SOCIETY

A component in a reverse fuel cell uses renewable power to knit together water and nitrogen to make ammonia.

MacFarlane says, ammonia is as green as can be. "Liquid ammonia is liquid energy," he says. "It's the sustainable technology we need."

Ammonia—one nitrogen atom bonded to three hydrogen atoms—may not seem like an ideal fuel: The chemical, used in household cleaners, smells foul and is toxic. But its energy density by volume is nearly double that of liquid hydrogen—its primary competitor as a green alternative fuel—and it is easier to ship and distribute. "You can store it, ship it, burn it, and convert it back into hydrogen and nitrogen," says Tim Hughes, an energy storage researcher with manufacturing giant Siemens in Oxford, U.K. "In many ways, it's ideal."

Researchers around the globe are chasing the same vision of an "ammonia economy," and Australia is positioning itself to lead it. "It's just beginning," says Alan Finkel, Australia's chief scientist who is based in Canberra. Federal politicians have yet to offer any major legislation in support of renewable ammonia, Finkel says, perhaps understandable in a country long wedded to exporting coal and natural gas. But last year, the Australian Renewable Energy Agency declared that creating an export economy for renewables is one of its priorities. This year, the agency announced AU$20 million in initial funds to support renewable export technologies, including shipping ammonia.

In Australia's states, politicians see renewable ammonia as a potential source of local jobs and tax revenues, says Brett Cooper, chairman of Renewable Hydrogen, a renewable fuels consulting firm in Sydney. In Queensland, officials are discussing creating an ammonia export terminal in the port city of Gladstone, already a hub for shipping liquefied natural gas to Asia. In February, the state of South Australia awarded AU$12 million in grants and loans to a renewable ammonia project. And last year, an international consortium announced plans to build a US$10 billion combined wind and solar plant known as the Asian Renewable Energy Hub in Western Australia state. Although most of the project's 9000 megawatts of electricity would flow through an undersea cable to power millions of homes in Indonesia, some of that power could be used to generate ammonia for long-distance export. "Ammonia is the key enabler for exporting renewables," says David Harris, research director for low-emissions technologies at Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO) Energy in Pullenvale. "It's the bridge to a whole new world."

First, however, the evangelists for renewable ammonia will have to displace one of the modern world's biggest, dirtiest, and most time-honored industrial processes: something called Haber–Bosch.

THE AMMONIA FACTORY, a metallic metropolis of pipes and tanks, sits where the red rocks of Western Australia's Pilbara Desert meet the ocean. Owned by Yara, the world's biggest producer of ammonia, and completed in 2006, the plant is still gleaming. It is at the technological vanguard and is one of the largest ammonia plants in the world. Yet at its core are steel reactors that still use a century-old recipe for making ammonia.

Until 1909, nitrogen-fixing bacteria made most of the ammonia on the planet. But that year, German scientist Fritz Haber found a reaction that, with the aid of iron catalysts, could split the tough chemical bond that holds together molecules of nitrogen, N_2, and combine the atoms with hydrogen to make ammonia. The reaction takes brute force—up to 250 atmospheres of pressure in the tall, narrow steel reactors—a process first industrialized by German chemist Carl Bosch. The process is fairly efficient; about 60% of the energy put into the plant ends up being stored in the ammonia's bonds. Scaled up to factories the size of Yara's, the process can produce vast amounts of ammonia. Today, the facility makes and ships 850,000 metric tons of ammonia per year—more than double the weight of the Empire State Building.

Most is used as fertilizer. Plants crave nitrogen, used in building proteins and DNA, and ammonia delivers it in a biologically available form. Haber–Bosch reactors can churn out ammonia much faster than natural processes can, and in recent decades the technology has enabled farmers to feed the world's exploding population. It's estimated that at least half the nitrogen in the human body today comes from a synthetic ammonia plant.

Haber–Bosch led to the Green Revolution, but the process is anything but green. It requires a source of hydrogen gas (H_2), which is stripped away from natural gas or coal in a reaction using pressurized, superheated steam. Carbon dioxide (CO_2) is left behind, accounting for about half the emissions from the overall process. The second feedstock, N_2, is easily separated from air, which is 78% nitrogen. But generating the pressure needed to meld hydrogen and nitrogen in the reactors consumes more fossil fuels, which means more CO_2. The emissions add up: Ammonia production consumes about 2% of the world's energy and generates 1% of its CO_2.

Yara is taking a first step toward greening that process with a pilot plant, set to open in 2019, that will sit next to the existing Pilbara factory. Instead of relying on natural gas to make H_2, the new add-on will feed power from a 2.5-megawatt solar array into a bank of electrolyzers, which split water into H_2 and O_2. The facility will still rely on the Haber–Bosch reaction to combine the hydrogen with nitrogen to make ammonia. But the solar-powered hydrogen source

Australia's windy coasts offer a bounty of energy, which it might one day export as a carbon-free fuel.

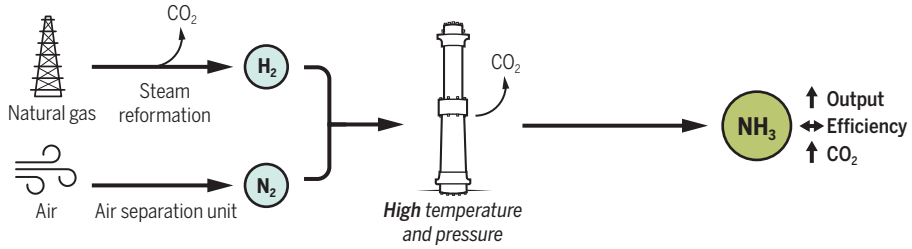
PHOTO: COAST PROTECTION BOARD, SOUTH AUSTRALIA

A green way to make ammonia

Reverse fuel cells can use renewable power to make ammonia from air and water, a far more environmentally friendly technique than the industrial Haber–Bosch process. Renewable ammonia could serve as fertilizer—ammonia's traditional role—or as an energy-dense fuel.

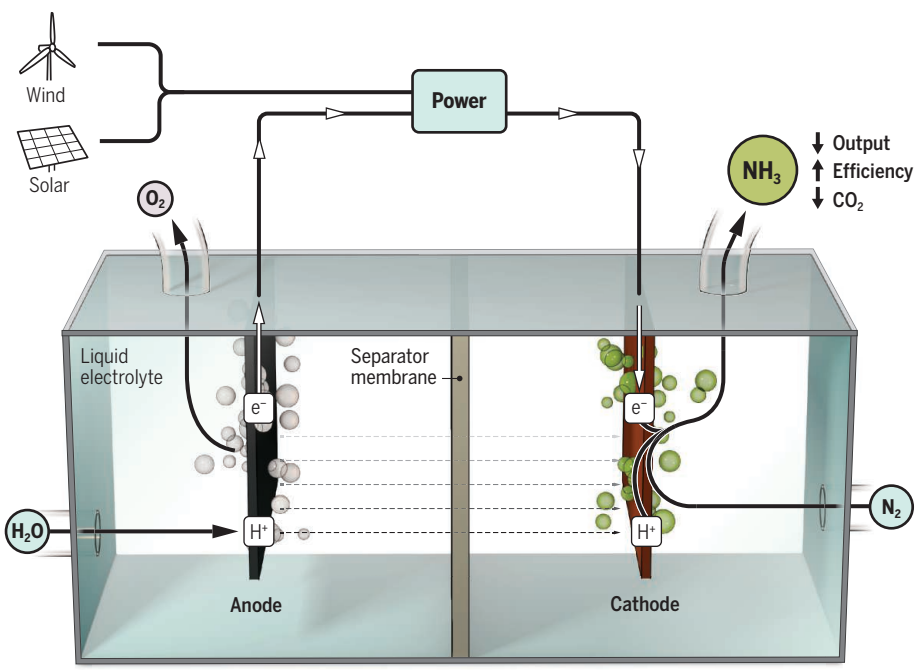
Industrial ammonia

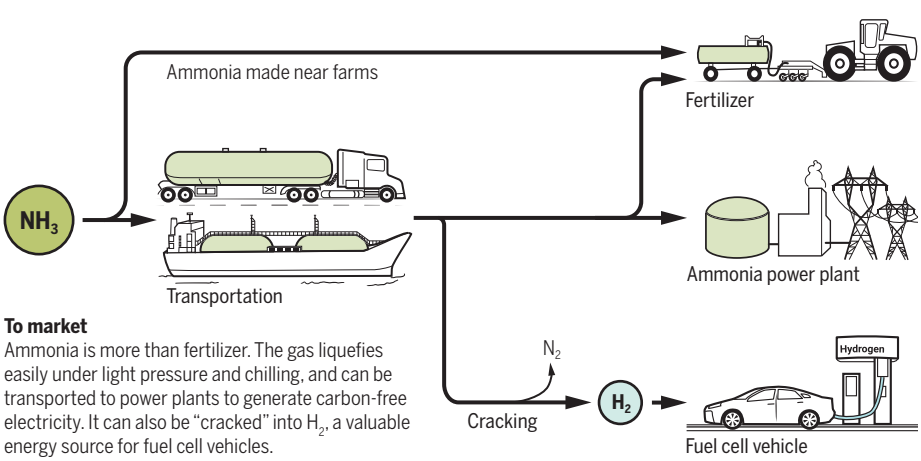
Most of the world's ammonia is synthesized using Haber–Bosch, a century-old process that is fast and fairly efficient. But the factories emit vast amounts of carbon dioxide (CO_2).



Gentler reactions

A reverse fuel cell uses renewable electricity to drive a chemical reaction that makes ammonia. Water reacts at the anode to make hydrogen ions (H^+), which migrate to the cathode where they react with nitrogen (N_2) to form ammonia. The reaction is efficient, but slow.





To market

Ammonia is more than fertilizer. The gas liquefies easily under light pressure and chilling, and can be transported to power plants to generate carbon-free electricity. It can also be "cracked" into H_2, a valuable energy source for fuel cell vehicles.

cuts total CO_2 emissions from the process roughly in half.

Other projects are following suit. The state of South Australia announced plans in February to build a AU$180 million ammonia plant, again relying on electrolyzers powered by renewable energy. Slated to open in 2020, the plant would be a regional source of fertilizer and liquid ammonia, which can be burned in a turbine or run through a fuel cell to make electricity. The supply of liquid energy will help stabilize the grid in South Australia, which suffered a debilitating blackout in 2016.

Ammonia made this way should attract buyers in places such as the European Union and California, which have created incentives to buy greener fuels. And as the market grows, so will the distribution routes for importing ammonia and the technologies for using it, Harris says. By then, fuel cells like MacFarlane's could be ready to displace Haber–Bosch itself—and the half-green approach to ammonia production could become fully green.

INSTEAD OF APPLYING fearsome heat and pressure, reverse fuel cells make ammonia by deftly wrangling ions and electrons. As in a battery being charged, charged ions flow between two electrodes supplied with electricity. The anode, covered with a catalyst, splits water molecules into O_2, hydrogen ions, and electrons. The protons flow through an electrolyte and a proton-permeable membrane to the cathode, while the electrons make the journey through a wire. At the cathode, catalysts split N_2 molecules and prompt the hydrogen ions and electrons to react with nitrogen and make ammonia.

At present, the yields are modest. At room temperature and pressure, the fuel cell reactions generally have efficiencies of between 1% and 15%, and the throughput is a trickle. But MacFarlane has found a way to boost efficiencies by changing the electrolyte. In the water-based electrolyte that many groups use, water molecules sometimes react with electrons at the cathode, stealing electrons that would otherwise go into making ammonia. "We're constantly fighting having the electrons going into hydrogen," MacFarlane says.

To minimize that competition, he opted for what's called an ionic liquid electrolyte. That approach allows more N_2 and less water to sit near the catalysts on the cathode, boosting the ammonia production. As a result, the efficiency of the fuel cell skyrocketed from below 15% to 60%, he and his colleagues reported last year in *Energy & Environmental Science*. The result has since improved to 70%, MacFarlane

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

says—but with a tradeoff. The ionic liquid in his fuel cell is goopy, 10 times more viscous than water. Protons have to slog their way to the cathode, slowing the rate of ammonia production. "That hurts us," MacFarlane says.

To speed things up, MacFarlane and his colleagues are toying with their ionic liquids. In a study published in April in *ACS Energy Letters*, they report devising one rich in fluorine, which helps protons pass more easily and speeds ammonia production by a factor of 10. But the production rate still needs to rise by orders of magnitude before his cells can meet targets, set for the field by the U.S. Department of Energy (DOE), that would begin to challenge Haber–Bosch.

Next to Monash University, Sarb Giddey and his colleagues at the Clayton offices of CSIRO Energy are making ammonia with their "membrane reactor." It relies on high temperatures and modest pressures—far less than those in a Haber-Bosch reactor—that, compared to MacFarlane's cell, boost throughput while sacrificing efficiency. The reactor designs call for a pair of concentric long metallic tubes, heated to 450°C. Into the narrow gap between the tubes flows H_2, which could be made by a solar- or wind-powered electrolyzer. Catalysts lining the gap split the H_2 molecules into individual hydrogen atoms, which modest pressures then force through the atomic lattice of the inner tube wall to its hollow core, where piped-in N_2 molecules await. A catalytically active metal such as palladium lines the inner surface, splitting the N_2 and coaxing the hydrogen and nitrogen to combine into ammonia—much faster than in MacFarlane's cell. So far only a small fraction of the input H_2 reacts in any given pass—another knock to the reactor's efficiency.

Other approaches are in the works. At the Colorado School of Mines in Golden, researchers led by Ryan O'Hayre are developing button-size reverse fuel cells. Made from ceramics to withstand high operating temperatures, the cell can synthesize ammonia at record rates—about 500 times faster than MacFarlane's fuel cell. Like Giddey's membrane reactors, the ceramic fuel cells sacrifice some efficiency for output. Even so, O'Hayre says, they still need to improve production rates by another factor of 70 to meet the DOE targets. "We have a lot of ideas," O'Hayre says.

Whether any of those approaches will wind up being both efficient and fast is still unknown. "The community is still trying to figure out what direction to go," says Lauren Greenlee, a chemical engineer at the University of Arkansas in Fayetteville. Grigorii Soloveichik, a manager in Washington, D.C., for the DOE's Advanced Research Projects Agency-Energy program on making renewable fuels, agrees. "To make [green] ammonia is not hard," he says. "Making it economically on a large scale is hard."

HOWEVER DISTANT, the prospect of Asia-bound tankers, full of green Australian ammonia, raises the next question. "Once you get ammonia to market, how do you get the energy out of it?" asks Michael Dolan, a chemist at CSIRO Energy in Brisbane.

The simplest option, Dolan says, is to use the green ammonia as fertilizer, like today's ammonia but without the carbon penalty. Beyond that, ammonia could be converted into electricity in a power plant customized to burn ammonia, or in a traditional fuel cell, as the South Australia plant plans to do. But currently, ammonia's highest value is as a rich source of hydrogen, used to power fuel cell vehicles. Whereas ammonia fertilizer sells for about $750 a ton, hydrogen for fuel cell vehicles can go for more than 10 times that amount.

In the United States, fuel cell cars seem all but dead, vanquished by battery-powered vehicles. But Japan is still backing fuel cells heavily. The country has spent more than US$12 billion on hydrogen technology as part of its strategy to reduce fossil fuel imports and meet its commitment to reduce CO_2 emissions under the Paris climate accord. Today the country has only about 2500 fuel cell vehicles on the road. But by 2030 Japanese officials expect 800,000. And the nation is eyeing ammonia as a way to fuel them.

> ***"It looks like there's enough interest to get this industry started."***
>
> **David Harris**, CSIRO Energy

Converting hydrogen into ammonia only to convert it back again might seem strange. But hydrogen is hard to ship: It has to be liquefied by chilling it to temperatures below –253°C, using up a third of its energy content. Ammonia, by contrast, liquefies at –10°C under a bit of pressure. The energy penalty of converting the hydrogen to ammonia and back is roughly the same as chilling hydrogen, Dolan says—and because far more infrastructure already exists for handling and transporting ammonia, he says, ammonia is the safer bet.

That last step—stripping hydrogen off ammonia molecules—is what Dolan and his colleagues are working on. In a cavernous metal warehouse on the CSIRO campus that has long been used to study coal combustion, two of Dolan's colleagues are assembling a 2-meter-tall reactor that is dwarfed by a nearby coal reactor. When switched on, the reactor will "crack" ammonia into its two constituents: H_2, to be gathered up for sale, and N_2, to waft back into the air.

That reactor is basically a larger version of Giddey's membrane reactor, operating in reverse. Only here, gaseous ammonia is piped into the space between two concentric metal tubes. Heat, pressure, and metal catalysts break apart ammonia molecules and push hydrogen atoms toward the tube's hollow core, where they combine to make H_2 that's sucked out and stored.

Ultimately, Dolan says, the reactor will produce 15 kilograms per day of 99.9999% pure hydrogen, enough to power a few fuel cell cars. Next month, he plans to demonstrate the reactor to automakers, using it to fill tanks in a Toyota Mirai and Hyundai Nexo, two fuel cell cars. He says his team is in late-stage discussions with a company to build a commercial pilot plant around the technology. "This is a very important piece of the jigsaw puzzle," Cooper says.

Beyond 2030, Japan will likely import between $10 billion and $20 billion of hydrogen each year, according to a renewable energy roadmap recently published by Japan's Ministry of Economy, Trade and Industry. Japan, Singapore, and South Korea have all begun discussions with Australian officials about setting up ports for importing renewably produced hydrogen or ammonia. "How it all comes together economically, I don't know," Harris says. "But it looks like there's enough interest to get this industry started."

Cooper knows how he wants it to end. Over coffee on a rainy morning in Sydney, he describes his futuristic vision for renewable ammonia. When he squints, he can see, maybe 30 years down the road, Australia's coast dotted with supertankers, docked at offshore rigs. But they wouldn't be filling up with oil. Seafloor powerlines would carry renewable electricity to the rigs from wind and solar farms on shore. On board, one device would use the electricity to desalinate seawater and pass the fresh water to electrolyzers to produce hydrogen. Another device would filter nitrogen from the sky. Reverse fuel cells would knit the two together into ammonia for loading on the tankers—a bounty of energy from the sun, air, and sea.

It's the dream that nuclear fusion never reached, he says: inexhaustible carbon-free power, only this time from ammonia. "It can never run out, and there is no carbon in the system." ■

COGNITION

When persistence doesn't pay

Rats, mice, and humans all invest more time in a foraging task than is in their interest

By **Sarah F. Brosnan**[1,2]

People routinely make bad decisions. Far from being random, however, these bad decisions are often predictable, occurring reliably in specific contexts (*1*–*3*). One set of such suboptimal decisions are cognitive biases, wherein individuals make decisions that predictably violate rationality or their own best interests without a logical reason for doing so (*4*). One such bias is the sunk cost bias, in which individuals invest more time or resources in an outcome than it is worth given the potential gain, presumably because they focus on hindsight evaluation of irrecoverable costs rather than prospective gains (*5*). On page 178 of this issue, Sweis *et al.* (*6*) report a clever experiment demonstrating the sunk cost bias in rats, mice, and humans. The findings also provide insight into why results from previous studies have been so variable.

Cognitive biases can be substantial impediments to optimal decision-making and are costly to individuals and society. Comparing responses across different species is key to understanding how these biases evolved and will help to both predict when they will manifest and offer suggestions as to how to offset them. However, determining how the sunk cost effect evolved has not been easy, because the bias is not consistently expressed in other species (*7*–*9*).

Studying the sunk cost bias is challenging because it is difficult to separate investment from future expectations, especially in field studies. Lab experiments can more easily distinguish between these factors, but lab experiment setups can differ substantially across species, particularly between humans and other species, making comparisons difficult.

[1]*Neuroscience Institute and Language Research Center, Georgia State University, Atlanta, GA 30302, USA.* [2]*National Center for Chimpanzee Care, MD Anderson Cancer Center, Bastrop, TX 78602, USA. Email: sbrosnan@gsu.edu*

PHOTO: JESUS GIRALDO GUTIERREZ/SHUTTERSTOCK

When foraging, animals must decide whether to continue in the current patch or move on. Staying too long is an example of the sunk cost bias, which Sweis *et al.* demonstrate in rats, mice, and humans.

Sweis *et al.* overcome the latter problem by using the same basic procedure, based on a foraging task, across all three species.

In their experiment, subjects had a limited time budget during which they rotated through four reward options that differed in how well the subjects liked them (subjects showed consistent preferences across the study). At each transition, subjects first had to decide whether to commit to the available option or move to the next option. If they committed, they could not get the reward until they had waited for a predetermined period of time; the time remaining was indicated to the subject throughout the waiting period. Subjects could leave at any point during the waiting period, without penalty, and move to the next option in the rotation.

A subject not showing the sunk cost bias should leave quickly if a less preferred reward is paired with a long wait time. Nonetheless, subjects of all species were more committed to sticking with their choices than their preferences warranted, thus demonstrating a sunk cost bias. Indeed, this bias grew stronger the longer they waited; that is, it increased with increasing investment. However, it was unaffected by the time spent making the choice initially. Supplemental work indicates that there is a distinct decision phase during which sunk costs do not accumulate, even when this phase is not explicitly included as a separate period in the experimental design. The authors argue that this is evidence for two independent valuation processes, one assessing whether to accept an offer and the other—the one that is susceptible to sunk costs—assessing whether to continue investing in the choice.

The evidence for two different valuation systems working together offers the possibility of a comprehensive explanation for the sunk cost effect that integrates cognition, behavior, and neural architecture across species and contexts. Moreover, these results hint at some of the reasons for the previous disparate results. Perhaps wait times were not long enough to get past the assessment period, or, more speculatively, decisions may change if subjects do not know how long they must wait or if they are otherwise engaged in an active task (such as lever pressing) during the waiting period.

A key reason for Sweis *et al.*'s success is the degree to which they worked to provide a relevant, comparable methodology. They chose a context, foraging, that had ecological relevance to these species in the context of sunk costs and came up with a clever way of providing information about how much future investment remained to minimize information uncertainty. Importantly, they provided the humans and the rodents with the same procedures and information.

Of course, there were differences among species. Differences in training between the rodent labs required control tests to verify that behavior was truly the same. Humans, unlike rodents, are not consistently food motivated, and so they "foraged" for entertainment, in the form of movie clips, instead. Such adaptations to account for species-typical behavior and differing environments are nearly unavoidable in comparative work, but it is critical that the fundamental procedure is the same and that the limitations of one's study are acknowledged (*10*), as in Sweis *et al.* For instance, the authors chose film clips because, like the differently flavored food, they differ in value but not handling time, which allows for a direct comparison across the four choice options.

Why are we susceptible to the sunk cost bias? In some situations, it may be difficult to evaluate future costs; in these cases, past efforts can be a reasonable proxy. If this heuristic worked well enough most of the time, there would not have been strong selection pressure to evolve a more sophisticated strategy. Indeed, many human decision-making biases likely result from a trade-off between the costs of acquiring and processing information and the costs of getting it wrong (*11*). Understanding this evolutionary trade-off can help us to better understand why humans make the decisions that we do and, ultimately, suggest ways of improving human decision-making (*12*). Sweis *et al.*'s adaptable design is likely to prove helpful in future experiments investigating sunk cost biases across a variety of species and contexts. ■

REFERENCES

1. G. Gigerenzer, in *Bounded Rationality: The Adaptive Toolbox*, G. Gigerenzer, R. Selten, Eds. (MIT Press, 1999), pp. 37–50.
2. D. Kahneman, P. Slovic, A. Tversky, Eds., *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge Univ. Press, ed. 1, 1982).
3. R. H. Thaler, *The Winner's Curse: Paradoxes and Anomalies of Economic Life* (Free Press, New York, 1991).
4. M. G. Haselton, D. Nettle, P. W. Andrews, in *The Handbook of Evolutionary Psychology* (Wiley, 2005), pp. 724–746.
5. R. Thaler, *J. Econ. Behav. Organ.* **1**, 39 (1980).
6. B. M. Sweis *et al.*, *Science* **361**, 178 (2018).
7. H. R. Arkes, P. Ayton, *Psychol. Bull.* **125**, 591 (1999).
8. F. Hoeffler, *Why humans care about sunk costs while animals don't. An evolutionary explanation*, paper presented at the Royal Economic Society Annual Conference, Nottingham, UK, 20 April 2006.
9. P. Magalhães, K. Geoffrey White, *J. Exp. Anal. Behav.* **105**, 339 (2016).
10. M. Smith, J. Watzek, S. F. Brosnan, *Int. J. Comp. Psychol.* **31**, uclapsych_ijcp_37777 (2018).
11. J. Watzek, S. F. Brosnan, *Cognition* **178**, 109 (2018).
12. R. H. Thaler, C. R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Yale Univ. Press, ed. 1, 2008).

10.1126/science.aau3144

EVOLUTION

Many roads to convergence

Plant genomes highlight complex mechanisms behind evolutionary convergence

By **László G. Nagy**

Many plants form specialized symbiotic root structures, called nodules, that harbor beneficial associations with nitrogen-fixing bacteria in the genera *Rhizobium* or *Frankia* (see the photo). How this nitrogen-fixing root nodule (NFN) symbiosis arose repeatedly during plant evolution is an age-old mystery: It shows signatures of convergence (the repeated emergence of similarity during evolution) yet builds on similar gene sets in phylogenetically distant plants. On page 144 of this issue, Griesmann *et al.* (*1*) sequenced the genomes of 10 plant species to reveal the genetic correlates of the origin and loss of NFN symbiosis. Their work reveals intricate gain and loss patterns of symbiosis-associated genes, calling for new models to explain convergent evolution.

NFN symbiosis has an immense impact on plant growth and global nitrogen cycling but occurs only in four related plant orders, the Fabales, Fagales, Cucurbitales, and Rosales, collectively known as the NFN clade. Only a few phylogenetically diverse members of the NFN clade form symbiosis, making NFN symbiosis a phylogenetically patchy character (*2*, *3*). Despite its phylogenetic patchiness, nodule formation builds on surprisingly similar gene sets (*4*), with some key symbiosis-associated genes, such as *NODULE INCEPTION* (*NIN*) and *RHIZOBIUM-DIRECTED POLAR GROWTH* (*RPG*), being widely conserved across plants. Besides *NIN* and *RPG*, another 290 genes were recently found to be upregulated in nodules of *Medicago truncatula* and *Parasponia andersonii*, two plant species that independently evolved the ability to form symbiosis. This is surprising, given that these species diverged ~100 million years ago and belong to two separate orders, the Fabales and Rosales, respectively (*4*). How can this level of genetic similarity be reconciled with the phylogenetic patchi-

Synthetic and Systems Biology Unit, Biological Research Center, Hungarian Academy of Sciences, 6726 Szeged, Hungary. Email: lnagy@fungenomelab.com

ness of nodulation? Griesmann *et al.* outlined three possible evolutionary scenarios to describe the extant distribution of NFN symbiosis across the plant tree of life: It could be described by up to 16 independent evolutionary origins, a single origin with multiple losses, or an intermediate scenario combining homologous and convergently evolved components of the underlying genetics. The intermediate scenario invokes a predisposition event, the gain of a genetic element (gene, regulatory interaction, or other), that made descendant lineages more likely to evolve NFN symbiosis (*5*).

Griesmann *et al.* sequenced the genomes of 10 carefully selected plant species to cover several clades that independently gained the ability to form NFN symbiosis. The authors embarked on a sizable challenge to identify genomic evidence for a postulated predisposition event, either for independent gains or for independent losses. Searching multiple genomes for genetic changes underlying the evolution of a trait is like finding a needle in a haystack, although incorporating the gain and loss patterns of the trait across the phylogeny can help narrow the search space down (*6*). Indeed, the authors found several genes, such as *NIN* and *RPG*, that were lost in clades that lost the ability to form NFN symbiosis. Such losses have important implications for genetic engineering of NFN symbiotic plants. In nature, the symbiosis can be lost in nitrogen-rich habitats, where the cost of forming symbiosis outweighs its benefits, making symbiosis-associated genes dispensable. In agricultural settings, however, long-term fertilizer application might lead to similar losses of the ability to nodulate, which could be exploited with the development of engineered NFN symbiotic crop plants in regions where fertilizer availability is limited.

Another 52 gene families showed parallel expansions in multiple symbiotic subclades of the NFN clade and contained known nodulation-associated genes of *Medicago*, although most gene family expansions were specific to individual subclades. These observations would make NFN symbiosis a classic example of convergent evolution: independent acquisitions of similar functions via different evolutionary routes, with dissimilar genetic backgrounds. However, many symbiosis-associated genes predate the origin of the NFN clade and are universally up-regulated during nodule formation in independent clades (*4*), suggesting a more complex scenario involving homology that dates back to early, nonnodulating precursors.

Convergently evolved traits involving a similar genetic background are hard to explain. Although the multiple-origins model is often the phylogenetically most parsimonious for such traits, the likelihood of parallel changes affecting tens to hundreds of the same genes is so low that multiple independent origins seem implausible. More complex models, involving deep homologies (*7*), latent homologies (*8*), predisposition events, or parallel co-option of genes have been proposed to explain the evolution of phylogenetically patchy traits, like NFN symbiosis. Although these have been used to denote various, sometimes overlapping processes, they refer to at least two distinct evolutionary mechanisms. Traits are deeply homologous if homology is not evident at the level of the entire trait but historical continuity of shared function exists for at least some of the genes involved in its development. Arthropod and vertebrate eyes are textbook examples of deep homologies: They develop by largely nonhomologous mechanisms but build on homologous genes, such as opsins, developmental transcription factors, and photoreceptor cells that can be traced back to primordial photosensory organs of a cnidarian or bilaterian ancestor (*9*). In this case, both function and genetics are homologous, but evidence for this is blurred by divergence and/or lineage-specific refinements, overall leading to apparent convergence.

Conversely, latent homologies are developmental potentials (for example, gene regulatory circuits) that can be deployed (through co-option or exaptation) for new functions by simple genetic changes. Because a few mutations (for example, in regulatory sequences) are sufficient for their expression in a new context, latent homologies reduce the mutational target size for evolution and can thus promote convergence. Mechanistically, predisposition events can comprise the evolution of latent homologies that, as hypothesized for the NFN clade (*5*), lead to a higher likelihood for convergence to occur.

Teasing apart the possible mechanisms behind convergently evolved traits remains a substantial challenge even in the era of genomics. It nevertheless appears that case studies and models are emerging to explain the pervasive occurrence of convergence across the tree of life. NFN symbiosis represents one particularly intricate case of phylogenetically patchy traits. The phylogenomic comparisons by Griesmann *et al.* provide important clues about the genetics of the repeated emergence of NFN symbiosis and the genes that are repeatedly lost in secondarily nonsymbiotic clades, although some ambiguity remains as to the convergent origins. With the increasing accessibility of whole-genome sequences, such comparisons will likely play an increasing role in dissecting the genetic background of complex multigenic traits, such as nodulation, and decoding long-standing evolutionary questions, including the many roads that can lead to evolved similarity in disparate branches of the tree of life. ■

A colored scanning electron micrograph of a bean plant root nodule shows nitrogen-fixing *Rhizobium* bacteria (green) embedded in plant tissue, forming a symbiotic relationship (magnification ×50).

PHOTO: STEVE GSCHMEISSNER/SCIENCE SOURCE

REFERENCES

1. M. Griesmann *et al.*, *Science* **361**, eaat1743 (2018).
2. M. J. Telford, G. E. Budd, H. Philippe, *Curr. Biol.* **25**, R876 (2015).
3. L. G. Nagy, G. M. Kovács, K. Krizsán, *Biol. Rev.* 10.1111/brv.12418 (2018).
4. R. van Velzen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4700 (2018).
5. D. E. Soltis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2647 (1995).
6. L. G. Nagy *et al.*, *Mol. Biol. Evol.* **34**, 35 (2017).
7. N. Shubin, C. Tabin, S. Carroll, *Nature* **457**, 818 (2009).
8. L. G. Nagy *et al.*, *Nat. Commun.* **5**, 4471 (2014).
9. N. Shubin, C. Tabin, S. Carroll, *Nature* **388**, 639 (1997).

10.1126/science.aau2409

CHEMISTRY

Femtosecond structural photobiology

Time-resolved crystallography reveals how bacteriorhodopsin uses light to drive chemistry

By **Keith Moffat**

Absorption of a photon in the visible region of the solar spectrum is the fundamental step in the photochemical reactions that drive biological processes such as imaging, photosynthesis, and sensory perception. All known photoreceptors are proteins that contain a chromophore: a small organic molecule that absorbs in the visible region of the spectrum (*1*). The bacterial integral membrane protein bacteriorhodopsin (bR) uses retinal as its chromophore (see the figure), capturing the energy in a photon to drive retinal isomerization and pump protons across the membrane. bR is a readily accessible analog to the visual rhodopsins in, for example, the human eye. On page 145 of this issue, Nogly *et al.* (*2*) used time-resolved x-ray crystallography (*3*, *4*) to determine intermediate structures of bR at near-atomic resolution on the femtosecond to picosecond time scales at which isomerization occurs.

The photochemistry of retinal in bR differs substantially from that of free retinal. Interactions between the retinal and its protein surroundings tightly constrain the chromophore and confer specificity in trans-to-cis isomerization about the C13=C14 bond, high quantum yield, and efficient proton pumping. These constraints must be very selective. Isomerization requires carefully choreographed, ultrafast atomic motions that have evolved to direct the reaction along the desired path and inhibit all undesired reactions, such as isomerization about a different bond or rapid reversion to the ground state.

How is this molecular choreography achieved? Extensive earlier studies (*5*) have explored this question by using femtosecond optical spectroscopies; quantum mechanical, molecular mechanics, and molecular dynamics calculations; and structural probes with low spatial and temporal resolution. Spectroscopic data provide a simple linear model for the reaction path. Explicit structural information on the femtosecond time scale is needed to develop a more detailed picture.

Nogly *et al.* now provide these data in the face of major experimental and analytical difficulties. The structural changes that occur during isomerization are small; few atoms move, and those that do move do not move very far. The time-dependent changes in x-ray scattering that constitute the raw data are therefore also small, yet their accurate measurement is essential. The authors' visible laser pump–hard x-ray probe experiments are analogous to making a movie one frame at a time. A light pulse from a visible laser—the pump—initiates the reaction in the bR molecules in a tiny crystal. After a controlled time delay, the crystal is illuminated by an extremely intense pulse from an x-ray laser—the probe. By obtaining probe data over all crystal orientations, the authors determine the structure of the bR molecules at that time delay. By varying the time delay, the "light" structures are determined frame by frame—the reaction path. If the visible laser pulse is omitted, no reaction occurs, and the "dark" structure is obtained.

Synchrotron x-ray sources at which such experiments were pioneered are restricted to a time resolution of ~100 ps (*6*). Nogly *et al.* now access isomerization on the femtosecond to picosecond time scale by using the extremely intense, femtosecond, hard x-ray pulses from the Linac Coherent Light Source at Stanford, the world's first hard x-ray free–electron laser (XFEL) (*7*, *8*). However, XFELs are based on noise amplification and generate x-ray scattering data that are prone to systematic errors. Each tiny bR crystal is destroyed by the extremely intense XFEL pulse immediately after generating its diffraction pattern.

Nogly *et al.* present difference electron density maps that minimize systematic error by directly comparing the x-ray scattering from the light structure (at a known, variable time delay after the pump pulse) and dark structure (before the pump pulse). Such difference maps are highly sensitive to small structural differences (*9*). To obtain atomic structures, the authors converted the difference maps into time-dependent, extrapolated density maps and further averaged these maps into time bins centered on the anticipated time of maximum concentration of the I, J, and K spectroscopic states. They then refined a single atomic structure in each extrapolated density map.

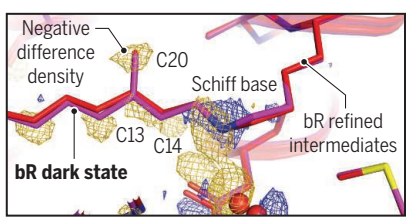
Two effects may complicate the accuracy of refinement. First, in any dynamic experiment, an individual time delay—here, a time bin—typically contains a mixture of structures (*3*, *4*). Second, absorption of a photon by retinal generates a vibrationally excited molecule far from equilibrium, with unusual bond lengths and angles. Reassuringly, the single refined structures are chemically plausible and consistent with both calculations (*2*) and the previously proposed aborted bicycle pedal model for isomerization (*10*, *11*).

The difference maps are of interest both for the atomic motions they reveal and for those they do not reveal. Certain protein atoms comprising the retinal pocket in the dark do not move; they impose constraints

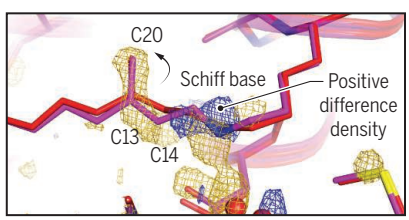
Tracking femtosecond retinal changes

Difference electron density maps after the pump laser pulse show the retinal structure in bacteriorhodospin (bR) before to isomerization (top) and during isomerization (bottom).

Time bin one (49 to 40 6fs)
Small structural changes in the initial electronic excited state occur after a photon is abosorbed but before isomerization has occurred.



Time bin two (457 to 646 fs)
In the J state, isomerization is half complete, with a torsion angle of 90° about the C13=C14 bond.



GRAPHIC: C. BICKEL/*SCIENCE*

Department of Biochemistry and Molecular Biology and Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60615, USA. Email: moffat@cars.uchicago.edu

on retinal motion throughout isomerization (see the figure). Other atoms and groups move in concert, restricting isomerization to the C13=C14 bond. A counter-ion cluster consisting of Asp^{85}, Asp^{212}, and a key buried water molecule (Wat^{402}) stabilizes the positive charge on the Schiff base in the dark state. Even before isomerization in the electronically excited state, this cluster is disrupted, and specific hydrogen bonds are weakened. The positive charge on the Schiff base is redistributed along the polyene chain. The alternation of single and double bonds along the polyene chain is perturbed and at least partial single-bond character conferred on the C13 to C14 bond (a double bond in the dark state) to facilitate its twisting as isomerization begins.

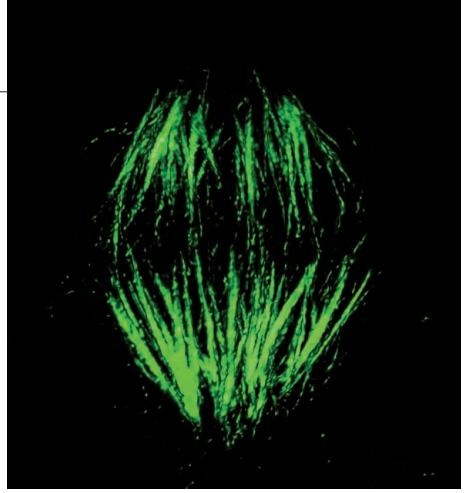
In bR, isomerization appears to be half complete (with a torsion angle of 90° about the C13=C14 bond) at ~500 fs. Isomerization has been studied with femtosecond time-resolved crystallography in two other proteins: photoactive yellow protein, whose small chromophore contains only a single double bond about which isomerization is also half complete at ~500 fs (*12*), and a photoswitchable fluorescent protein, in which a larger chromophore must execute molecular contortions to accomplish isomerization in ~1 ps (*13*). It is likely that electronic and structural similarities exist in the earliest steps of the elementary chemical process of isomerization in different chromophores.

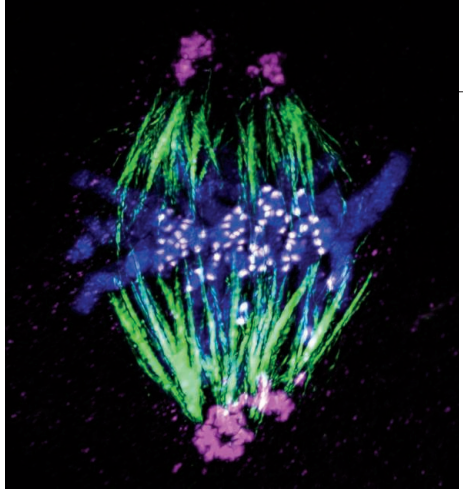
Challenging questions remain. Is the nature of the cis-to-trans isomerization pathway in the physiologically important visual rhodopsins simply the reverse of the trans-to-cis pathway seen here in bR? Are the structures and charge distributions of the retinal pockets in the trans and cis species in bR and visual rhodopsins closely similar? Nogly *et al.* show that the techniques are firmly in place to address these questions. ■

REFERENCES

1. A. Möglich, X. Yang, R. A. Ayers, K. Moffat, *Annu. Rev. Plant Biol.* **61**, 21 (2010).
2. P. Nogly *et al.*, *Science* **361**, eaat0094 (2018).
3. K. Moffat, *Chem. Rev.* **101**, 1569 (2001).
4. K. Moffat, in *X-ray Free Electron Lasers*, U. Bergmann, V. K. Yachandra, J. Yano, Eds. (Royal Society of Chemistry, 2017), vol. 18, p. 107.
5. C. Wickstrand, R. Dods, A. Royant, R. Neutze, *Biochim. Biophys. Acta* **1850**, 536 (2015).
6. V. Srajer *et al.*, *Science* **274**, 1726 (1996).
7. P. Emma *et al.*, *Nat. Photonics* **4**, 641 (2010).
8. J. Spence, *IUCr J.* **4**, 322 (2017).
9. R. Henderson, K. Moffat, *Acta Cryst.* **B27**, 1414 (1971).
10. A. Warshel, Z. T. Chu, *J. Phys. Chem. B.* **105**, 9857 (2001).
11. P. Altoè, A. Cembran, M. Olivucci, M. Garavelli, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20172 (2010).
12. K. Pande *et al.*, *Science* **352**, 725 (2016).
13. N. Coquelle *et al.*, *Nat. Chem.* **10**, 31 (2018).

10.1126/science.aau3200





EMBRYOGENESIS

Double trouble at the beginning of life

Dual-spindle assembly in early embryos can compromise mammalian development

By **Agata P. Zielinska** *and* **Melina Schuh**

Every human life begins with the fertilization of an egg (*1*). Once the egg and the sperm have fused, the parental chromosomes need to be united. To this end, the egg and sperm chromosomes are first packaged into two separate membrane-enclosed nuclei. These nuclei then slowly move toward each other and break down in the center of the fertilized egg, called the zygote. Only then are the maternal and paternal chromosomes united—but not quite. Surprisingly, the parental chromosomes do not mix immediately but instead occupy distinct territories in the zygote throughout the first cellular division (*2*, *3*). How the autonomy of parental genomes is retained after fertilization has remained unclear. On page 189 of this issue, Reichmann *et al.* (*4*) used elegant microscopy methods to illuminate this special moment, when the parental chromosomes first meet in live mouse zygotes, and follow how the chromosomes become distributed as the zygote divides. Their findings reveal an unexpected mechanism that keeps the parental genomes apart during the first division of the embryo: The male and female chromosomes each assemble their own chromosome separation machineries. This increases the probability that chromosomes are separated into multiple, unequal groups, which may compromise embryo development and give rise to spontaneous miscarriage.

Max Planck Institute for Biophysical Chemistry, Göttingen 37077, Germany. Email: melina.schuh@mpibpc.mpg.de

Reichmann *et al.* labeled the maternal and paternal chromosomes in different colors by taking advantage of distinct DNA sequences (*5*) in the parental chromosomes, which came from different mouse strains (*6*). To follow in detail how the chromosomes are united, zygotes have to be imaged at very high spatial and temporal resolution. However, embryos are light sensitive, which has hindered a detailed analysis in the past. The authors overcame this limitation by using an innovative light-sheet microscope, which illuminates the embryo selectively in the region of interest but not in adjacent regions, as is the case with standard microscopy approaches (*7*). This reduces the amount of light that the embryo is exposed to. Moreover, this method is fast and allowed the authors to reconstruct the entire volume that is occupied by the chromosomes with unprecedented spatial and temporal resolution.

The authors tracked the maternal and paternal chromosomes as they first met. In addition, they imaged microtubules, the proteinaceous fibers that form the spindle apparatus that captures, aligns, and distributes the chromosomes equally between the two daughter cells of the dividing zygote. Surprisingly, Reichmann *et al.* found that the maternal and paternal chromosome masses each assemble a separate spindle structure that autonomously initiates chromosome alignment (see the images). Later, the two spindles merge into a single spindle. However, the maternal and paternal chromosomes remain in separate regions of the merged spindle and do not mix.

PHOTOS: REICHMANN *ET AL.* (*4*)

Immunofluorescence staining of a mouse zygote (left image) shows parallel alignment of two mitotic spindles (green). This is overlaid (right image) with microtubule organizing centers (magenta), sites of microtubule binding to chromosomes (gray), and chromosomes (blue).

Whether the spatial separation of parental chromosomes has any advantages for the developing mammalian embryo is unclear. However, the fact that zygotes have a dual spindle creates a previously unforeseen source of potential error. The final task the zygote has to accomplish before it divides is to align the two spindle axes in parallel to each other, so that the two spindles can merge into a compact dual structure. If the poles of the spindles fail to align and merge, the genetic material of the zygote could be pulled into three or four directions instead of two (see the figure). Reichmann *et al.* demonstrate that spindle misalignment leads to the formation of multinucleated embryos that have more than one nucleus per cell. It is easy to envision that such an undesired partitioning of the DNA would have a negative impact on the fidelity of subsequent cell divisions and hence might compromise embryo development. Indeed, experience from in vitro fertilization (IVF) clinics shows that early human embryos cultured in vitro before implantation frequently have multiple nuclei in their cells (*8*) and then fail to develop further (*9*).

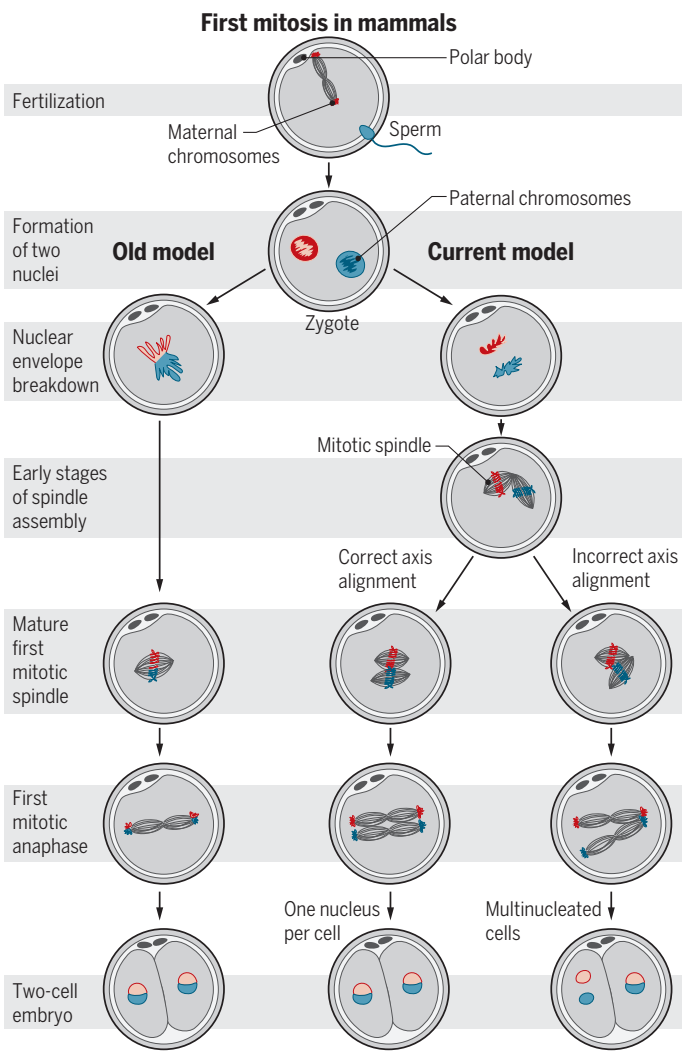
To evaluate the implications of these findings for human infertility, it is important to investigate whether cells with multiple nuclei in human embryos are also caused by defects during dual-spindle merging, as reported for mouse embryos in this study. In humans, the sperm delivers a centrosome into the zygote (*10*, *11*). Centrosomes are microtubule-organizing centers that help to generate a bipolar spindle in somatic cells (*12*). However, in mice, the sperm does not carry a centrosome, and the zygotic spindle needs to assemble without the help of centrosomes (*1*). Reichmann *et al.* demonstrate that chromosomes play a key role in driving spindle assembly in the mouse zygote. The chromosome-based spindle assembly mechanism seems to create a permissive environment for the formation of two independent spindle structures.

The presence of centrosomes in human zygotes could facilitate the assembly of a single bipolar spindle. However, it is still unclear whether spindle assembly in human zygotes is primarily centrosome driven, or whether they also assemble two distinct spindles from chromosome surfaces. Interestingly, spindle assembly in unfertilized human eggs is largely chromosome driven (*13*), which raises the possibility that also in human zygotes, even in the presence of centrosomes, chromosome-driven spindle assembly may occur.

In many countries, the law considers that new human life begins when parental chromosomes are united upon fertilization (*14*). This work is a reminder that the definition of the unification of parental genomes as the beginning of life is not as clear cut as was previously assumed. How we define the "beginning of life" not only has ethical implications but also practical repercussions for fertility treatments. This is because in some countries, such as Germany (*14*), parental nuclei are allowed to merge and hence life is allowed to "begin" in IVF clinics only in a very limited number of in vitro–cultured zygotes, all of which have to be transferred to the mother. This policy increases the rate of multiparous pregnancies, which are associated with severe risks to the health of the mother and her children. Additionally, because identification of the most promising fertilized eggs has to legally occur before the parental genomes have merged (so that embryos with merged genomes, defined as being human life, are not discarded), embryologists have to select embryos for implantation at the zygotic stage. Because our understanding of early human embryogenesis is still poor, it is difficult to select at this early stage the embryos that have the highest chance of dividing correctly, implanting, and giving rise to a healthy pregnancy. Considering that there is now comprehensive evidence that the parental DNA does not mix in live zygotes, this study highlights that a dialogue is needed to reshape the legal definition of when life begins, so that it is supported by up-to-date scientific knowledge. By revisiting this definition, we could increase the chances of a healthy pregnancy among couples who otherwise struggle to conceive. ■

Together, but still apart

Upon fertilization, parental chromosomes need to be united. Unexpectedly, in mice, male and female chromosomes are kept apart on separate mitotic spindles during the first division of the zygote. This increases the probability of forming multinucleated cells.



GRAPHIC: N. DESAI/*SCIENCE*

REFERENCES

1. D. Clift, M. Schuh, *Nat. Rev. Mol. Cell Biol.* **14**, 549 (2013).
2. W. Mayer *et al.*, *J. Cell Biol.* **148**, 629 (2000).
3. C. van de Werken *et al.*, *Nat. Commun.* **5**, 5868 (2014).
4. J. Reichmann *et al.*, *Science* **361**, 189 (2018).
5. S. Henikoff, K. Ahmad, H. S. Malik, *Science* **293**, 1098 (2001).
6. Y. Miyanari, C. Ziegler-Birling, M. E. Torres-Padilla, *Nat. Struct. Mol. Biol.* **20**, 1321 (2013).
7. P. Strnad *et al.*, *Nat. Methods* **13**, 139 (2016).
8. H. Balakier, K. Cadesky, *Hum. Reprod.* **12**, 800 (1997).
9. A. Egashira *et al.*, *J. Reprod. Dev.* **61**, 595 (2015).
10. G. Manandhar, C. Simerly, G. Schatten, *Hum. Reprod.* **15**, 256 (2000).
11. E. L. Fishman *et al.*, *Nat. Commun.* **9**, 2210 (2018).
12. M. Bettencourt-Dias, D. M. Glover, *Nat. Rev. Mol. Cell Biol.* **8**, 451 (2007).
13. Z. Holubcova *et al.*, *Science* **348**, 1143 (2015).
14. M. Kreyenfeld, D. Konietzka, *Childlessness in Europe: Contexts, Causes, and Consequences* (Springer, 2017).

10.1126/science.aau3216

INFECTIOUS DISEASES

Eradication genomics—lessons for parasite control

Genomic surveillance could help achieve targets for the elimination of tropical diseases

By **James A. Cotton, Matthew Berriman, Love Dalén, Ian Barnes**

Large-scale programs are seeking to control or eliminate infectious diseases with the greatest impact on global health. Many of these efforts target the neglected tropical diseases (NTDs) that disproportionately affect the lives of the poor. Often the aim is to eradicate the causative pathogens. The idea—inspired by the success of smallpox eradication in the 1960s and 1970s—is that a large-scale, but time-limited, effort could eliminate a disease for all successive generations, resulting in an enormous payoff both financially and in improved health. Here, we discuss the value of genomic approaches to support disease eradication efforts, particularly by analogy with how conservation genomics is supporting efforts to prevent extinctions.

Although there are fungal, viral, and bacterial NTDs, most are caused by infections with parasites (particularly protozoa and helminths), including the protozoa that cause sleeping sickness, leishmaniasis, and Chagas disease, or worms that cause diseases such as schistosomiasis, hookworm disease, and river blindness. Only two diseases—human smallpox and the cattle disease rinderpest—have been successfully eradicated; both of these campaigns relied on the availability of highly effective and long-lasting vaccines that are not available for any of the parasitic NTDs. Control methods vary between the different diseases and can include controlling parasite carriers (vectors), better sanitation, and improvements to health systems to allow faster detection and treatment. However, the main approach has been large-scale treatment of people at risk of infection, whether they actually have the disease or not. The size of these interventions is astounding: More than 1 billion people were treated for NTDs in 2016, with drugs donated by manufacturers. Nonetheless, this represents only 63% of the treatment needed to fully cover the populations at highest risk (*1*).

Key to the success of NTD control is measuring changes in pathogen populations, to ensure that interventions are on course to achieve elimination goals and to direct additional resources to areas of high or stubborn prevalence. However, monitoring pathogen populations by conventional techniques is challenging in many of the environments where these diseases flourish, particularly as only limited diagnostics are available (*2*). Monitoring will likely become particularly difficult as prevalence is reduced (*3*). The power of molecular data to understand pathogen populations is already widely appreciated: Genetics can help confirm suspected cases, discriminate between strains, identify new infectious agents, understand the evolutionary history of populations, and shed light on transmission. For example, in the polio eradication campaign, a global network sequenced part of every virus isolated from around 200,000 stool specimens per year to identify transmission links and track the disappearance of particular genotypes (*4*). Genetic data identified polio outbreaks caused by oral vaccine–derived revertant strains, highlighting a need to gradually withdraw serotypes from the vaccine (*5*).

> ***"Genomic surveillance could thus be a powerful tool for tracking... parasite populations."***

Whereas the most comprehensive genetic data are for prokaryotes and viruses, most NTDs are caused by eukaryotic parasites that at least occasionally reproduce sexually. Most viruses and bacteria reproduce clonally, and exchange genetic material unpredictably, so molecular epidemiology can be restricted to tracking the abundance of different lineages through space and time. Recombination in sexual eukaryotes makes each region of the genome an independent genetic marker, so genomic data for a small number of individuals can reveal a great deal about the population those individuals come from. Genomic surveillance could thus be a powerful tool for tracking changes in the demography or abundance of parasite populations.

In many ways, the study of eradication genomics is a mirror image of conservation genomics. Although the latter aims to understand populations to preserve biodiversity and ultimately reverse species declines, eradication genomics studies the genetics of populations that are deliberately being reduced by human activity, to enhance our ability to depress the populations further and ultimately remove them completely. For some time, population genetic data have been used in conservation biology to understand changes in endangered populations (*6*), but we expect that analysis of eukaryotic pathogen populations could result in some important, more subtle lessons from conservation genomics. For example, that isolated subpopulations are harder to conserve implies that eradication will be easier if pathogen populations can be isolated from one another. In small populations, stochastic changes in allele frequencies due to random death or survival of individuals can swamp the natural selection of beneficial genetic variation; the increased impact of this genetic drift can make it much harder for small populations to adapt to changing environments. Furthermore, inbreeding within small isolated populations is expected to be deleterious (*7*). Looking for genomic signatures such as reduced genetic diversity and longer sequences of homozygosity due to inbreeding may therefore identify populations close to elimination (*8*) (see the figure).

One difficulty is that levels of genome-wide diversity in species can be strongly affected by life history and ancient demographic events, as well as recent changes in population size. We may thus need a good picture of pathogen populations before control measures have an impact. Therefore, collecting and conserving samples for genetic analysis should be an urgent consideration for control programs. Guinea worm is the parasite closest to eradication; cases have declined from several million per year in the 1980s to 30 in 2017 (*9*). However, Guinea worm infections were reported in Chad in 2011 after a decade with no reported cases. Coincident with this reemergence were infections in dogs, although Guinea worm was not previously considered zoonotic (*10*). We assume that Guinea worm evolved during the 10-year gap, but the lack of suitable samples predating 2011 makes it impossible to exclude the possibility that worms migrated from elsewhere during this time. Understanding the dog infections is key to the eradication program: Although there were only 25 human cases reported worldwide in 2016, more than 1000 dog infections were identified in Chad that year. The need to retrospectively obtain samples is clear. There are no historical Guinea worm samples stored

Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK. Email: james.cotton@sanger.ac.uk

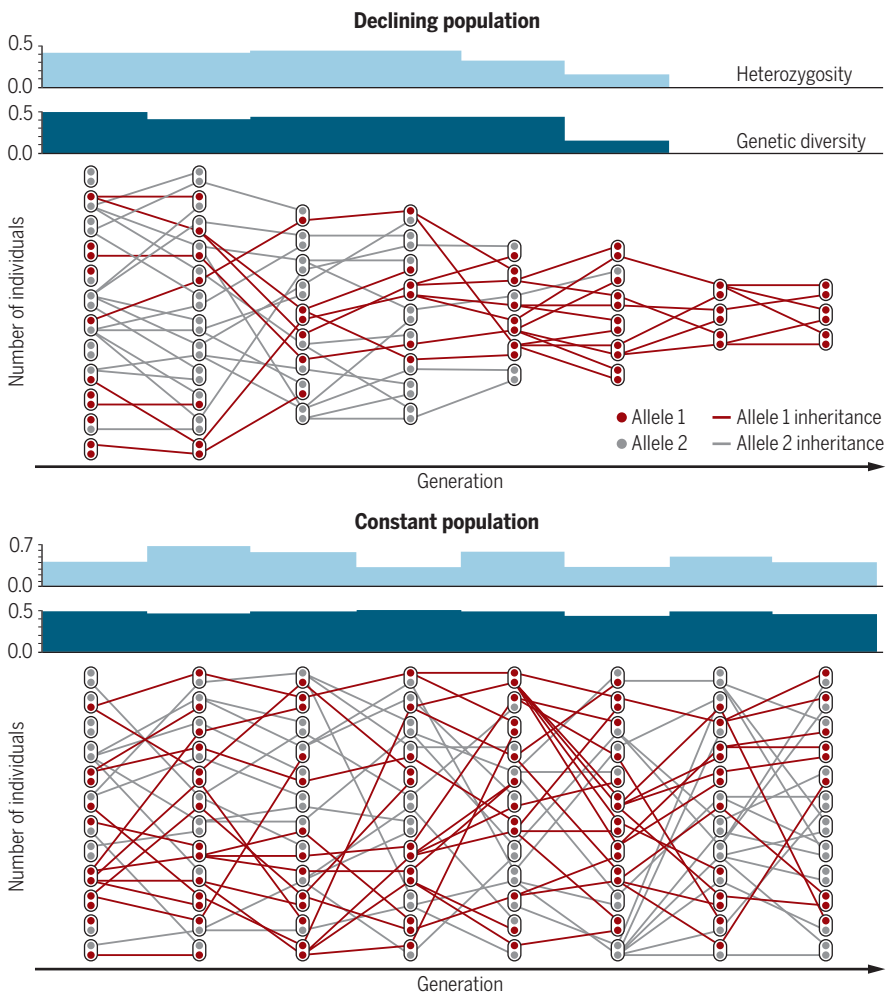
in a way that is likely to keep DNA intact, but there may be ways to circumvent the need for well-preserved DNA. Recent developments in paleogenomic methods, including improvements in the efficiency of building genomic libraries and the use of enzymatic treatments to reduce the impact of DNA damage, allow access to the highly fragmented and chemically modified genomic information recovered from archival specimens, and even from environmental sources such as archaeological specimens or sediments (*11, 12*). A precise picture of the historical population structure and size of these parasites will allow us to parameterize models of inbreeding and gene flow, improving the accuracy of inferences about the effect of control measures on current populations (*13*).

For many species, researchers have started exploring the genetics of populations using DNA "fingerprints" obtained from the differential distribution of microsatellites—repeat sequences in DNA—and other molecular markers. However, reductions in cost make whole-genome sequencing the most appropriate tool for the future, as genome-wide data are more powerful, and more comparable between populations and pathogens, and sequencing avoids the need to develop marker panels. At least for most key NTD pathogens, basic genomic resources such as annotated reference genomes are already available. Genome-wide data also enable the impact of control measures on pathogen populations to be understood beyond changes in distribution or numbers, as genomic data are sensitive to changes at functional loci and could reveal evolutionary changes of epidemiological relevance, such as the emergence of drug resistance. The emergence or sudden spread of new pathogen genotypes, signatures of strong recent selection on parts of the genome, or signs of a strong bottleneck in the absence of reduced prevalence are indications that a population is evolving in response to drug treatment or other control measures. This approach is being used to track artemisinin resistance in human malaria parasites in Asia, where population genomics approaches have played a key role in identifying the major locus responsible for reduced efficacy of the drug and understanding the genetic architecture of this trait (*14*).

Clearly, the highest priorities for control programs focus on delivering effective control measures and maintaining funding and political support. However, the rewards for collecting genomic data, and establishing pretreatment baselines through the analysis of archival samples, will be increasingly evident as programs mature. There will be many challenges in establishing genomic surveillance programs: the need to be comprehensive and coordinated, obtaining access to both prospective and archival samples, understanding the epidemiology and biology of the pathogens, to name a few. In some cases, theory will need to be developed to explain the population genetics of NTD pathogens, many of which have genetic systems distinct from those of obligately sexual, diploid animals. Translation of biomedical science to clinical practice has necessitated "bench to bedside" collaborations, and similarly, eradication genomics will require collaboration between technical experts in molecular biology and paleogenomics, as well as groups with expertise in the biology, control, and epidemiology of particular NTDs. The speed and scope of changes in NTD pathogen populations happening today is new territory for science. The ambitious goals set for NTD control will require pathogen populations to shrink at an unprecedented rate. However, success is not guaranteed (*15*) and will require constant vigilance. ■

Loss of diversity in a declining population

As control efforts reduce the size of a population, the chance that an individual is descended from closely related parents increases, leading to a decline in genetic diversity and heterozygosity.



Diagrams show inheritance of two alleles in simulated populations of 12 diploid individuals that declines or stays constant over eight generations.

GRAPHIC: ADAPTED BY A. KITTERMAN/*SCIENCE*

REFERENCES

1. Reaching a Billion: Fifth Progress Report on the London Declaration on NTDs (Uniting to Combat Neglected Tropical Diseases, 2017).
2. P. J. Hotez *et al.*, *PLOS Negl. Trop. Dis.* **10**, e0003895 (2016).
3. P. Klepac *et al.*, *Philos. Trans. R. Soc. London B Biol. Sci.* **368**, 20120137 (2013).
4. O. M. Diop *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **66**, 538 (2017).
5. N. C. Grassly, *Philos. Trans. R. Soc. London B Biol. Sci.* **368**, 20120140 (2013).
6. F. W. Allendorf *et al.*, *Nat. Rev. Genet.* **11**, 697 (2010).
7. B. Charlesworth *et al.*, *Genet. Res.* **74**, 329 (1999).
8. F. Abascal *et al.*, *Genome Biol.* **17**, 251 (2016).
9. Guinea worm wrap-up #253 (Centers for Disease Control and Prevention, 2018); www.cartercenter.org/resources/pdfs/news/health_publications/guinea_worm/wrap-up/253.pdf.
10. M. L. Eberhard *et al.*, *Am. J. Trop. Med. Hyg.* **90**, 61 (2014).
11. M.-T. Gansauge *et al.*, *Nat. Protoc.* **8**, 737 (2013).
12. N. Rohland *et al.*, *Philos. Trans. R. Soc. London B Biol. Sci.* **370**, 20130624 (2015).
13. D. Díez-del-Molino *et al. Trends Ecol. Evol.* **33**, 176 (2018).
14. O. Miotto *et al.*, *Nat. Genet.* **47**, 226 (2015).
15. C. J. M. Whitty, *Clin. Med.* **14**, 419 (2014).

10.1126/science.aar6609

FERROELECTRICITY

Perovskite ferroelectrics go metal free

Metal-free perovskites exhibit ferroelectric properties rivaling those of $BaTiO_3$

By **Wei Li**[1,2] *and* **Li-Jun Ji**[2]

Ferroelectric materials, which have spontaneous electric polarization that can be switched with an external electric field, have applications as capacitors, sensors, and data-storage devices (*1*). Since the discovery of high-performance $BaTiO_3$ and $LiNbO_3$, ferroelectric perovskite oxides have dominated industrial applications (*2*). However, these ceramics are expensive to produce as high-quality thin films, and they often contain heavy metals such as Pb^{2+}, which raises health and environmental concerns. Replacing heavy metals with organic components in the perovskite lattice was considered an alternative (*3*), but hybrid organic–inorganic perovskites have not shown comparable ferroelectricity and stability (*4*). On page 151 of this issue, Ye *et al.* (*5*) report metal-free perovskites that can compete with ferroelectric oxides (see the figure).

Perovskites have a general formula ABX_3, where A and B are large and small metal cations, respectively, and X is a bitopic anion (for example, O^{2-}) that coordinates to B. The structural topology can be maintained if the A- and X-site ions are replaced with organic molecules within certain size ranges (*6*). The obvious starting point would be substituting the large A-site metal cation, and the successful structural characterizations of $MAPbX_3$ (where MA is methylammonium, and X is Cl, Br, or I) opened up such possibilities (*7*). Many hybrid organic–inorganic perovskites have been synthesized by replacing the single atomic X site with bitopic organic linkers such as formate and dicynamide groups (*8*).

Can this strategy extend to the B site? The perovskite framework is constructed via coordination bonding in oxides and hybrid systems. Formation of coordination bonds requires that a ligand donate lone-pair electrons to the orbitals of transition metals. Likewise, for forming hydrogen (H) bonds, electronegative atoms with lone-pair electrons are needed as acceptors. In this sense, these two different bonding interactions exhibit some geometrical similarity. As the B site requires spherical symmetry to bond with six neighboring X sites for forming an octahedron, replacing it with a molecular cation that requires an octahedral H-bonding mode to bind with appropriate X-site acceptors could result in metal-free perovskite compounds. Indeed, the first metal-free perovskite, (piperazinium)$(NH_4Cl_3)\cdot H_2O$, was synthesized by assembling the NH_4^+ with chlorides through H-bond interactions (*9*).

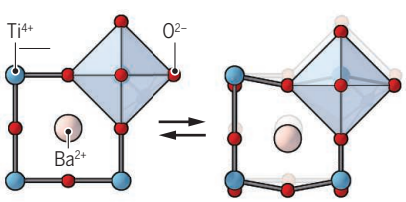
Guided by these molecular design principles, Ye *et al.* have designed a new class of ferroelectric metal–free perovskites. The inclusion of noncentrosymmetric or chiral organic amine cations imposed the polar symmetry that rendered the desired ferroelectric ordering in the perovskite lattice. The prototypical compound MDABCO–NH_4I_3 (MDABCO, *bis*-*N*-methyl-*N*′-diazabicyclo[2.2.2]octonium) exhibits a high Curie temperature (T_c of 448 K) and substantial spontaneous polarization (P_s of 22 μC/cm²), which compare well with the properties of commercial $BaTiO_3$. Such notable T_c and P_s values of MDABCO–NH_4I_3 should enable wide potential applications in environments where conventional oxide materials dominate. In addition, their multiple polarization directions enable facile rotation through electrical-field manipulation, which should make the poling process easier (*10*).

The substantial organic content of the metal-free perovskites makes these compounds soft, so they respond to stress differently compared with oxides and hybrid perovskites, in which stronger ionic and coordination bonding forces exist. This facilitates facile growth of high-quality thin films that are critical for managing electric dipoles through strain engineering. These new metal-free perovskite ferroelectrics are simple to synthesize, low cost, and lightweight.

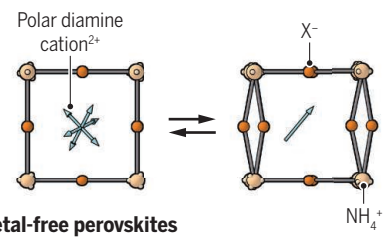
The marriage between symmetry and H bonding in these metal-free compounds extends the fields of perovskites and ferroelectrics to a new realm. However, some exciting aspects need further exploration. Bonding interactions between NH_4^+ and halides extend beyond H bonding, and their electronic nature needs to be further unveiled. Some of the transitions in these metal-free perovskites are strongly first-order and ferroelastic, which leads to substantial shear strains and distinctive twin-wall dynamics (*11*). Ferroelastic twin walls show exotic physical properties beyond the bulk material (*12*), so knowledge of the twin-wall nature in these metal-free perovskites would enrich their potential for high-density memory devices (*13*). Finally, the existence of chirality in these molecular perovskites complicates the alignment of electric dipoles in their crystal lattices, so understanding its coupling with ferroelectric ordering could lead to unexpected properties at the bulk and nanoscale levels (*14*). ■

Ferroelectric origins in perovskites

Ferroelectricity in perovskites is induced by discrete structural changes that occur at phase transitions.



Perovskite oxides
Ferroelectricity primarily arises from the displacement of the Ti^{4+} ion from the center of the TiO_6 octahedron in $BaTiO_3$.



Metal-free perovskites
Ferroelectricity in the materials of Ye *et al.* primarily stems from the order-disorder of the polar diamine cation.

[1]*Center of Rare Earth and Inorganic Functional Materials, National Institute for Advanced Materials, Nankai University, Tianjin 300350, China.* [2]*School of Physics, Huazhong University of Science and Technology, Wuhan 430074, China. Email: wl276@nankai.edu.cn*

REFERENCES AND NOTES

1. J. F. Scott, *Science* **315**, 954 (2007).
2. Y.-M. You *et al.*, *Science* **357**, 306 (2017).
3. B. Saparov, D. B. Mitzi, *Chem. Rev.* **116**, 4558 (2016).
4. W. Li *et al.*, *Nat. Rev. Mater.* **2**, 16099 (2017).
5. H.-Y. Ye *et al.*, *Science* **361**, 151 (2018).
6. G. Kieslich *et al.*, *Chem. Sci.* **5**, 4712 (2014).
7. D. Weber, *Z. Naturforsch. B* **33**, 1443 (1978).
8. W.-J. Xu *et al.*, *CrystEngComm* **18**, 7915 (2016).
9. C. A. Bremner *et al.*, *J. Am. Chem. Soc.* **124**, 10960 (2002).
10. J. Harada *et al.*, *Nat. Chem.* **8**, 946 (2016).
11. E. K. H. Salje, *Annu. Rev. Mater. Res.* **42**, 265 (2012).
12. S. Cherifi-Hertel *et al.*, *Nat. Commun.* **8**, 15768 (2017).
13. J. F. Scott, C. A. Paz de Araujo, *Science* **246**, 1400 (1989).
14. S. Seki *et al.*, *Science* **336**, 198 (2012).

ACKNOWLEDGMENTS

We acknowledge funding from the National Natural Science Foundation of China (no. 21571072).

10.1126/science.aat5729

GRAPHIC: C. BICKEL/*SCIENCE*

RETROSPECTIVE

Jens Christian Skou (1918–2018)

A pioneer in the biochemistry of membrane proteins

By **Poul Nissen**

Jens Christian Skou, a pioneer of biomembrane research and molecular physiology, died on 28 May, just a few months shy of his 100th birthday. The Danish physiologist discovered the enzyme known as the sodium-potassium adenosine triphosphatase (Na^+,K^+-ATPase), or sodium-potassium pump, which exports sodium ions from and imports potassium ions into cells. Almost all modern biochemistry, cell biology, and physiology textbooks showcase the sodium-potassium pump, which accounts for about 20 to 30% of the energy used by cells through hydrolysis of ATP in human and animal bodies and up to 70 to 80% in the brain. In recognition of this first discovery of an ion-transporting ATPase, Skou shared the 1997 Nobel Prize in Chemistry with chemists Paul D. Boyer and John E. Walker for their work on the mechanism of ATP synthesis.

Skou grew up in the coastal town of Lemvig at Limfjorden in Northern Jutland, Denmark. He earned his medical degree from the University of Copenhagen in 1944, and before any consideration of a research career had even struck him, he practiced medicine at a small hospital in the town of Hjørring, also in Northern Jutland. In 1947, to help his career in medicine along by earning a doctorate degree, Skou moved to the then very young Aarhus University to do research in its small department of physiology. His work focused on the mechanism of anesthetics. This brought him to lipid monolayers as models of the nerve cell membrane, and then to a research visit to the marine biology research station in Woods Hole, Massachusetts, in 1953. The trip was a transformative moment of his career. Unlike his small Aarhus department, Woods Hole was a large and dedicated research environment with intense scholar interactions and a huge library of scientific journals and books. There, Skou came across the curious observations of an ATPase activity in the nerves of giant squid. He took these studies back home to Aarhus and applied them to a more manageable source of nerves from local crabs, which he found displayed the same activity. Leaving aside his medical career plans, he instead devoted himself to research. Skou and a slowly growing research community characterized the enzyme, guided by the excellent controls of tractable substrates and specific inhibitors, and throughout his career, this system would define his work. He remained at Aarhus University and continued to do research even after his retirement in 1988.

In 1957, Skou published his first paper on the Na^+,K^+-ATPase enzyme. In it, he speculated that the enzyme might be the long-sought sodium-potassium pump that maintains steep electrochemical gradients for Na^+ and K^+ across the cell membrane

and enables the majority of transmembrane transport and signaling processes, including the action potentials of firing neurons. He proposed that the pump should span the cell membrane, an idea that at the time was heavily disputed. Despite insight just gained from the myoglobin structure that made the hypothesis seem unlikely, Skou persisted, and later in the 1960s and early 1970s, when better models of biomembranes were presented and the first structure of a membrane protein was visualized, his theory was shown to be true.

Aarhus University came of age, and in the mid-1960s when physiologist Peter Leth Jørgensen joined the department, a highly productive research community emerged with a critical spark of creativity and drive. Skou and Jørgensen's internal competition and fierce discussions were well-known; neither of them would prioritize compromises and "hygge"—a Danish social concept of relaxed tolerance, comfort, and contentment—over scientific rigor. Whereas Skou was the classical biochemist, Jørgensen became a molecular biologist. The two groups complemented each other in both style and results.

I first met Skou in 1991 when I was a master's student in Aarhus University's department of chemistry. In Jens Nyborg's lab, we were working on the EF-Tu protein and transfer RNA by crystallography. Skou came to request the use of our French press homogenizer to disrupt cell membranes and isolate Na^+,K^+-ATPase from shark rectal gland. I used the device often, so people directed him to me. Crystallization of the sodium-potassium pump seemed an unreal possibility at that time, but the encounter led to our first discussions about how to proceed.

It wasn't until 2007 that I talked again at length with Skou. Our group, which included biophysicist Jesper Vuust Møller, had determined the Ca^{2+}-ATPase crystal structure and then (with physiologist Bente Vilsen, also at Aarhus) the first crystal structure of Na^+,K^+-ATPase. Skou had invited himself to see it. I was still a newcomer to the Na^+,K^+-ATPase community, and I was quite nervous about hosting this giant of the field. Fortunately, Skou set me at ease by drawing me into a chat about sea trout fishing.

I learned that Skou was a man of the sea. He grew up by the shores of Limfjorden, and his research ideas had materialized at Woods Hole. His favorite sources for Na^+,K^+-ATPase studies came from crab and shark, and local fishermen were his most trusted suppliers. A devoted fly fisher of sea trout, he spent time on the shores of the beautiful peninsula of Mols across the Bay of Aarhus, where he had a perfectly located summer cottage.

People would often describe Skou as stubborn, but he was a modest man and a clear supporter of the social equality that emerged in Denmark in the 20th century. His wife, Ellen Margrethe, was active in politics and a respected voice on the public health care system. He was also a strong proponent of work-family balance and would always finish work at the lab in the afternoon to join his family. His Nobel Prize gave him the authority to advocate giving researchers the freedom he had been given in his early career to pursue creative ideas and follow his curiosity. He wished that today's younger generations could have the same leeway to explore that had allowed him to develop his original discoveries. Jens Christian Skou personified the creativity and courage of a great scientist, and he was a quiet and modest role model to us all. He will live on in our research, discussions, and memories. ■

ACKNOWLEDGMENTS

I am grateful to M. A. Skydsgaard and many colleagues for sharing their insights and memories.

10.1126/science.aau5275

Danish Research Institute of Translational Neuroscience (DANDRITE), Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark. Email: pn@mbg.au.dk

PHOTO: SOEREN KJELDGAARD

POLICY FORUM

CITIZEN SCIENCE

Citizen science, public policy

New research models may benefit from policy modifications

By **Christi J. Guerrini, Mary A. Majumder, Meaganne J. Lewellyn, Amy L. McGuire**

Citizen science initiatives that support collaborations between researchers and the public are flourishing. As a result of this enhanced role of the public, citizen science demonstrates more diversity and flexibility than traditional science and can encompass efforts that have no institutional affiliation, are funded entirely by participants, or continuously or suddenly change their scientific aims. But these structural differences have regulatory implications that could undermine the integrity, safety, or participatory goals of particular citizen science projects. Thus far, citizen science appears to be addressing regulatory gaps and mismatches through voluntary actions of thoughtful and well-intentioned practitioners. But as citizen science continues to surge in popularity and increasingly engage divergent interests, vulnerable populations, and sensitive data, it is important to consider the long-term effectiveness of these private actions and whether public policies should be adjusted to complement or improve on them. Here, we focus on three policy domains that are relevant to most citizen science projects: intellectual property (IP), scientific integrity, and participant protections.

Although the definitional bounds of citizen science are debated, there is general consensus that citizen science encompasses scientific endeavors in which individuals without specific scientific training participate as volunteers in one or more activities relevant to the research process other than (or in addition to) allowing personal data or specimens to be collected from them. These activities might take place at any point during the research process and include participation in study design, data collection and analysis, and dissemination of results (*1*). They might even encompass the entirety of the research process where, as in coordinated self-experimentation, the role of professional scientists is minimal.

Recognizing the potential for citizen science to advance scientific knowledge and promote public support of scientific activities, professional associations have emerged in the United States, Europe, and Australia to support citizen science efforts, including consideration of policy interactions. Meanwhile, the U.S. government recently passed legislation that supports agency use of citizen science and crowdsourcing to conduct projects that advance their missions (*2*).

INTELLECTUAL PROPERTY

Citizen scientists are usually not paid for their contributions, although they might benefit from participation in other ways. When citizen scientist volunteers are not compensated, their work is not subject to most employment-related laws and practices that govern their scientific collaborators. Some employment-related laws and practices have implications for ownership of IP, including copyrights and patents. Although domestic IP laws generally do not reach beyond national borders, international treaties have harmonized basic IP protections. Thus, application of many U.S. protections to citizen science contexts is generalizable to other countries, though some important differences persist.

U.S. copyright protection extends to authors of every original work fixed in a tangible medium. An exception exists for a "work made for hire" that is prepared by an employee within the scope of employment or by agreement as a specially ordered contribution to a collective work (*3*). Because citizen scientists are volunteers, any copyrightable works that they develop in the course of a project—for example, photographs, writings, and creative selections or arrangements of scientific data—likely do not constitute works for hire absent valid agreements otherwise. Rather, the volunteer who created the works is the copyright "author" and owner. Other countries do not recognize an automatic transfer of ownership to employers or commissioning parties (*3*).

Where projects do not require copyright assignment as a condition of volunteering, a citizen scientist who retains ownership of his or her works under governing domestic law can refuse to grant permission to publish them, which could disrupt the scientific process or prevent the dissemination of findings. However, projects can and often do avoid potential disputes using Creative Commons licenses, which allow creators to retain their copyrights while permitting others to copy and use their works.

A similar tool is not widely available for patents. In most countries, the legal inventor of a patentable discovery is one who contributes to its conception, and in the United States, only inventors may apply for patents. Employers typically require their scientist employees to assign any rights to future inventions as a condition of employment, including decisions to apply for patents and license inventions. By contrast, patent assignments do not appear to be a typical condition of volunteering in citizen science projects. Yet, it is possible that some citizen scientists will make contributions that support claims of sole or co-inventorship under domestic patent laws (*4*). For example, Sharon Terry, a self-described citizen scientist who helped discover the gene responsible for her children's rare disease, pseudoxanthoma elasticum, is a co-inventor on U.S. patents related to that discovery (*5*). Challenges may arise if the citizen science inventor who has retained IP rights exercises them to exclude projects from using the patented inventions or disagrees with co-inventors on whether to pursue patents or license inventions to others.

Contracts can be used to clarify rights and establish expectations related to patent and other IP rights. Projects might take a page from employment practices and require participants to assign future patent rights to project leaders as a condition of participation. However, such a practice seems incompatible with models that view citizen scientists as respected partners in the research process. The online citizen science game FoldIt, for example, does not require patent assignments. Instead, FoldIt's IP policy provides that "players who contributed to the discovery will be considered co-inventors for any discovery produced through play" and issues of ownership will be handled at a later date by the University of Washington, where the game was developed and is managed (*6*). It remains to be seen how well this approach will work in practice (*7*).

Alternatively, contracts can incorporate advance commitments by citizen scientists and professional scientists to cede stewardship of patents to nonprofit organizations that support research. Ms. Terry's advocacy organization, for example, is the assignee of the patented inventions that she helped discover (*5*). However, it is unclear whether this model would be readily accepted in contexts where citizen scientists do not have funding or other leverage to negotiate IP control.

Citizen scientists and professional scientists might forego patenting discoveries resulting from their collaborations or grant

Baylor College of Medicine, Center for Medical Ethics and Health Policy, 1 Baylor Plaza, Houston, TX 77030, USA. Email: guerrini@bcm.edu

each other nonexclusive, royalty-free cross-licenses to any discoveries. Although the development of standardized agreements for each of these scenarios seems achievable, support for such agreements may be limited where interest in commercialization is high (*7*). One-way material transfer agreements, which are a common vehicle for transferring research materials between institutions, might be adapted to promote sharing by and with citizen scientists. But in some cases, they may add unnecessary complexity to what might otherwise be straightforward transactions, and they are notoriously difficult to monitor and enforce.

In sum, although contractual approaches have limits, the ability to tailor contracts to circumstances makes them more practical than legislative approaches that would change the legal rules to produce different default outcomes—assuming that consensus on optimal outcomes even could be reached. As a policy matter, resources are probably better spent encouraging transparency and negotiation of IP terms at the outset of citizen science collaborations.

Citizen scientists collect samples in a U.S. national park.

PHOTO: NPS

RESEARCH INTEGRITY

Concerns about the quality of data contributed by citizen scientists and the soundness of their collection, reporting, and analytical techniques have long been raised (*8, 9*). Although professional scientists are not immune from quality transgressions, methodological rigor is central to their training and professional advancement, and formal mechanisms exist for holding professional scientists accountable for the quality of their work. Volunteers, on the other hand, may not experience similar external pressures to ensure research integrity (*10*) and may prioritize other aspects of their participation. Also, there may be fewer opportunities for professional scientists to address knowledge gaps and otherwise act as a check on quality if their involvement in projects is minimal.

Conflicts of interest can also undermine research integrity. Individuals who engage in citizen science can have biases stemming from their alliances with private, nonprofit, and political organizations, as well as their involvement in lawsuits (*9, 10*). They may also have biases based on their perceptions of how they or their community might be harmed or benefited by particular findings. Further, and especially with respect to research on politically charged topics, it is not beyond the realm of possibility that citizen scientists might be recruited to provide bad data to "improve" results or even sabotage research, although careful monitoring will help ferret out such attacks.

Of course, research conducted exclusively by professional scientists can also be undermined by personal biases and conflicts of interest, but institutional rules, funding stipulations, regulatory procedures, and professional norms work to compel their identification and disclosure whether the research is conducted in for-profit, nonprofit, or governmental settings. Citizen science projects that are not embedded within institutions, are self-funded, or are otherwise outside of regulatory control may operate without these traditional safeguards.

Citizen science projects have adopted strategies to promote research integrity, and many guides, tools, and templates are available to support projects from the planning stages through evaluation (*11, 12*). Studies indicate that citizen scientists can produce reliable data on par with those produced by professionals (*12*). But it is not yet clear how well the best practices that have been developed address the many flavors of citizen science or are being followed by its practitioners. These questions should be studied, but in the meantime, there may be opportunities to promote integrity through policy.

For example, recently enacted U.S. legislation aimed at promoting citizen science and crowdsourcing projects by federal agencies provides that all data collected through such efforts should be made publicly available where appropriate and to the extent possible (*2*). Such data accessibility creates opportunities to investigate questionable or poor-quality data and assess fitness for use through independent examination (*10*). The law also requires agencies to "make all practicable efforts" to ensure that participants adhere to federal research misconduct policies, which include sanctions for fabricating and falsifying data (*2*). For studies conducted with federal funds, the penalties for research misconduct—the most severe being a funding ban—will not apply to most citizen scientists because they do not seek such funding. However, their interests will be affected if professional collaborators lose the federal support that makes possible their work.

These regulations will not reach citizen science projects that are not conducted with U.S. federal employees or not federally funded. Journals that publish citizen science research help fill this gap through the peer review process and by placing pressure on authors to make their study data publicly available. Although these policies do not affect citizen science projects that are uninterested in traditional publication, some projects may have commitments to openness that match or even exceed journal requirements. For example, DIYgenomics.org publishes protocols and data on project wikis.

A related problem is presented when citizen scientists have conflicts of interest that are not disclosed. Many national regulations and journals have adopted requirements of conflicts reporting, but they may not reach citizen scientists who do not qualify as investigators or authors, although they facilitate research in meaningful ways. To promote accountability, such requirements should extend to the disclosure of relevant conflicts held by citizen scientists. Problems will arise, however, when disclosure is incompatible with volunteer terms that allow anonymous participation. We support project discretion to offer anonymity, especially where sensitive information is involved or citizen scientists could be subjected to intimidation or harm if identified. In such cases, consideration should be given to permitting disclosure at the aggregate level (*10*).

PARTICIPANT PROTECTIONS

As uncompensated volunteers, citizen scientists are unable to rely on traditional labor laws to prevent and redress harm because they do not qualify as protected workers. The Common Rule, which applies to all research involving human subjects that is conducted or supported by the U.S. government, is directed at protecting research participants. For studies governed by the Common Rule and not within an exemption, institutional review boards (IRBs) must ensure that any risks to subjects are minimized and reasonable in relation to anticipated benefits. Studies that take place in other countries may still be subject to the Common Rule or similar national requirements, so consideration of the Common Rule's provisions may be broadly instructive.

Importantly, Common Rule protections are directed toward "subjects" of research from whom identifiable data or biospecimens are collected. They do not explicitly authorize IRBs to consider risks or benefits to citizen scientists who facilitate research in other ways. Depending on the setting and study design, examples of specific risks to citizen scientists could include expectations of overwork, requirements to assume financial burdens, or vulnerability to harassment by others. Examples of specific benefits, on the other hand, could include training in scientific techniques, access to tools and data for personal use, or opportunities for coauthorship. Where citizen scientists are both subjects and facilitators of research, IRBs may interpret their jurisdiction to include risks and benefits associated with both roles. However, in some contexts, citizen scientists are only research facilitators and not also subjects, or their roles may be ambiguous (*13*). Federal protections also do not explicitly contemplate risks and benefits to communities, although community-based research review processes, including community IRBs and advisory committees, have emerged to complement traditional IRB review to ensure that involved communities are engaged in and directly benefit from proposed research and that study designs are culturally appropriate.

Many U.S.-led citizen science initiatives are not covered by the Common Rule because they are not federally conducted or supported (*14*). A small number may nevertheless be covered by research subject protections adopted by the U.S. Food and Drug Administration (FDA), which apply to clinical investigations that support applications for products regulated by FDA, regardless of funding source. Like the Common Rule, however, FDA research subject protections do not explicitly account for risks and benefits associated with supporting scientific initiatives in ways other than serving as research subjects.

For those citizen science initiatives outside the scope of these federal regulations and not conducted in the few states that extend these kinds of protections to all research regardless of funding source, there may be no legal requirement to evaluate or disclose potential risks to citizen scientists. Individuals owe a duty of ordinary care to avoid foreseeable injury to others, including general volunteers, but application of this common law doctrine is highly context-specific and may be unavailable to volunteers for any number of reasons. Moreover, some projects may require waivers of liability as a condition of volunteering, which are not allowed in the context of federally regulated research.

> ***"...citizen scientists are both subjects and facilitators of research..."***

Some unregulated projects, like Sage Bionetworks' mPower study of Parkinson's disease, which collects data through participants' mobile phones, voluntarily close this gap by securing independent IRB review of their protocols, although this can be cost-prohibitive for many projects. Others have proposed ethics evaluations by "citizen ethicists" who critique experiments and post their opinions online for potential participants to review, as well as downloadable ethics toolkits geared toward citizen scientists (*15*). Meanwhile, do-it-yourself biology projects may voluntarily adhere to a community code of ethics, which includes directives to pursue peaceful purposes and adopt safe practices. Until recently, the community also supported an online forum through which professional biosafety experts answered questions posed by citizen scientists. However, some projects may fail to engage in any kind of risk assessment or management.

Of the three policy domains discussed, participant protections require the most immediate attention given that they implicate direct physical harms, yet collective action focused on assessing those harms relative to other risks and benefits specific to citizen science has thus far been limited. How citizen science projects are making these risk-benefit calculations, and according to what processes, should be empirically investigated. Then, whether these data justify extending current legal protections to unregulated citizen science, or creating new policy frameworks altogether, merits study by policy-makers in partnership with the citizen science community.

If it is determined that additional regulation is warranted, it is unlikely to occur at the federal level given that a proposal to extend the Common Rule to additional kinds of research was recently considered and rejected. At a minimum, however, guidance should be developed by citizen science practitioners and participants in collaboration with ethicists to aid unregulated projects that cannot afford independent IRB review in conducting risk-benefit assessments. Such guidance also may be useful to traditional IRBs evaluating regulated projects in which the role of citizen scientists is multi-faceted or ambiguous. In addition, modes of risk assessment and management that are complementary or alternative to those required by the Common Rule and FDA regulations, including but not limited to ethics reviews conducted by non-traditional IRBs and citizen ethicists, should be evaluated to understand the citizen science contexts in which they could be useful.

As citizen science becomes more prevalent, it is increasingly important that policy opportunities to support participants and practitioners are identified and responsibly pursued. In the end, some policy adjustments may be in the best interests of society, but to be successful, they must appreciate the distinct ethos of citizen science and be guided by its diverse stakeholders. ■

REFERENCES AND NOTES

1. J. L. Shirk *et al.*, *Ecol. Soc.* **17**, 29 (2012).
2. Crowdsourcing and Citizen Science Act, Public Law 114-329 (6 January 2017), codified at 15 U.S.C. § 3724.
3. S. J. Marsnik, R. M. Lorentz, "Who owns employee works? Pitfalls in a globally distributed work environment," in *Managing the Legal Nexus Between Intellectual Property and Employees: Domestic and Global Contexts*, L. J. Oswald, M. A. Pagnattaro, Eds. (Elgar, 2015), chap. 8.
4. T. Scassa, H. Chung, "Managing intellectual property rights in citizen science" (Commons Lab, Woodrow Wilson International Center for Scholars, 2015).
5. S. F. Terry *et al.*, *Nat. Rev. Genet.* **8**, 157 (2007).
6. FoldIt, Terms and services of consent v. 3 (8 March 2013); https://fold.it/portal/legal.
7. L. S. Vertinsky, *Fl. State Univ. Law Rev.* **41**, 1067 (2014).
8. A. C. J. W. Janssens, P. Kraft, *PLOS Med.* **9**, e1001328 (2012).
9. Rise of the citizen scientist. *Nature* **524**, 265 (2015).
10. D. B. Resnik, K. C. Elliott, A. K. Miller, *Environ. Sci. Policy* **54**, 475 (2015).
11. A. Wiggins *et al.*, "Data management guide for public participation in scientific research" (DataONE PPSR Working Group, 2013); www.birds.cornell.edu/citscitoolkit/toolkit/steps/accept/DataONE-PPSR-DataManagementGuide.pdf.
12. M. Kosmala, A. Wiggins, A. Swanson, B. Simmons, *Front. Ecol. Environ.* **14**, 551 (2016).
13. R. Gellman, "Crowdsourcing, citizen science, and the law: legal issues affecting federal agencies" (Commons Lab, Woodrow Wilson International Center for Scholars, 2015).
14. M. A. Rothstein, J. T. Wilbanks, K. B. Brothers, *J. Law Med. Ethics* **43**, 425 (2015).
15. A. D. Marcus, "The ethics of experimenting on yourself," *Wall Street J.* (24 October 2014); www.wsj.com/articles/the-ethics-of-experimenting-on-yourself-1414170041.

ACKNOWLEDGMENTS

Development of this manuscript was funded by the U.S. National Human Genome Research Institute grant KO1- HG009355. The authors declare no financial conflicts of interest and thank C. Lee for research assistance.

10.1126/science.aar8379

ECOLOGY

To be a bee

A charming account celebrates the insects' idiosyncrasies and the people passionate about protecting them

By **Rachael Winfree**

The sweat bee *Augochloropsis sumptuosa* is one of the roughly 20,000 known bee species.

PHOTO: USGS BEE INVENTORY AND MONITORING LAB/FLICKR

We have bees to thank for some of the better features of our world, from the fruits and vegetables we eat to the world's flowering plants, which radiated alongside them during the Cretaceous. Nearly 90% of plant species require pollinators (an added benefit for anyone who suffers from windborne pollen allergies). Humans even seem to share aesthetic preferences with bees, insofar as the shapes, colors, and scents of the flowers we enjoy originally evolved to appeal to these and other pollinators.

Given the necessity of bees, the signs that they are in trouble are concerning. Although maladies of the domesticated honey bee (*Apis mellifera*), such as colony collapse disorder, have captured the most public attention, declines in the world's 20,000 wild bee species have the potential to be even more ecologically catastrophic. In *Buzz*, Thor Hanson deftly presents the predicament of bees today, in part by recounting conversations with scientists and practitioners who are leading the effort to understand and conserve bees. In only 304 pages, he successfully covers a wide range of topics, from bee evolution and natural history to the economic value of pollination.

Buzz
The Nature and Necessity of Bees
Thor Hanson
Basic Books, 2018, 304 pp.

Like all good naturalists, Hanson has a gift for seeing the world from his subject's point of view. At one point, he predicts the nest site preferences of a bumblebee queen (a boot, it turns out). At another, he empathizes with bees that won't pollinate alfalfa because the flowers' anthers bop them on the head.

In addition to his interest in entomology, Hanson also has an interest in etymology. Who knew, for example, that "dumbledore" is a historical name for the bumblebee? The book's bibliography reflects the diversity of topics covered, ranging from Virgil and Tolstoy to the scientific literature and CNN.

In one of the more surprising chapters (even to an ecologist who studies bees), Hanson presents evidence for the key role played by honey bees in human evolution. He begins with the honeyguide, an African bird species that feeds on bee hives but needs another, larger species to conduct the initial raid on the hive. The honey badger was long thought to be the honeyguide's original mutualist partner. Bringing some basic biology to bear, Hanson points out that honey badgers don't like to climb trees and in addition are nearsighted, hard of hearing, and nocturnal. By contrast, the Hadza people of Tanzania eagerly seek out honeyguides and harvest honey from wild honey bee hives—so much so that honey constitutes 15% of the calories in the Hadza diet. The availability of honey may even have furthered the evolution of our unusually large brains, the book reveals.

Hanson is an upbeat and often humorous guide, with a DIY attitude, the latter of which sometimes lands him in odd situations (such as sitting in a McDonald's, using tweezers to separate the bee-pollinated from the non–bee-pollinated components of a Big Mac). In another chapter, after reading that branched, pollen-collecting hairs are a distinctive trait that separates bees from their evolutionary forebears the wasps, he tests the hypothesis that minute particles adhere better to bees than wasps by weighing a bee and a wasp before and after dipping each in flour.

Hanson also has a sharp eye for the idiosyncratic—which he doesn't have to look far to find among bees or among bee enthusiasts. In the first category, he describes bees that nest inside active termite mounds and others that construct "elaborate vases" made of flower petals. Still others carry pollen on their abdomens, making each bee "look like she's wearing a tiny, brightly colored apron" (as compared with most other bees, "whose pollen loads look more like stockings pulled up high on their back legs.") In the second, are a retired insurance salesman who gives bees away as presents and an alfalfa farmer who slows his car to a crawl while driving so as to avoid killing even a single bee.

The only criticism I have of *Buzz*, and it is not a major one, is that the book underplays how little we actually know about bees. The chapter covering the decline and possible extinction of certain North American bumblebee (*Bombus*) species quite rightly focuses on the extent and causes of these declines. It doesn't mention, though, that *Bombus* is among the best-studied of the world's 443 bee genera. The conservation status of most bee species is simply unknown. In the United States, for example, there is no federal list of rare or declining bee species, and ad hoc assessment of which species are in trouble has been left to researchers and the Xerces Society for Invertebrate Conservation.

Overall, however, if you have time to read one book on what is happening with modern bees, you couldn't do better than *Buzz*. ■

The reviewer is at the Department of Ecology, Evolution, and Natural Resources, Rutgers University, New Brunswick, NJ 08901, USA, and is on the board of directors of the Xerces Society for Invertebrate Conservation, Portland, OR 97232, USA. Email: rwinfree.rutgers@gmail.com

10.1126/science.aat9250

FILM

Learning from different disciplines

Conversations spark connections as scientists search for inspiration in other fields

By **Brian Uzzi**

If you knocked on Heaven's door, and God greeted you, what question would you ask? What is the nature of human consciousness, and how can it be expanded? Where does the Universe begin and end? What is time, and why isn't it constant? What causes deviance? And further still, if you received an answer, could you decipher it? Or would the answer only be a clue leading to the next clue?

The Most Unknown
Ian Cheney, director
Vice/Motherboard, 2018, 88 minutes. Available on Netflix, August 2018.

The Most Unknown, a documentary film in the Simons Foundation Science Sandbox series, takes the viewer on a fantastical journey of nine scientists as they intrepidly knock on Heaven's door. Each asks a profound question in a different way. But all are making a difference in how science explains the Universe, the world we live in, and the worlds within us.

Microbiologist Jennifer Macalady collects samples from a water source in the Frassasi caves in Italy.

Although the filmmaker, Ian Cheney, is not a scientist, he makes a point that is now well documented scientifically: Scientists from different disciplines who immerse themselves in one another's work and exchange ideas solve the hardest problems (*1*). Midway through the film, for example, California Institute of Technology geobiologist Victoria Orphan describes how tiny microbes, too small to see with the naked eye, can "eat" greenhouse gases and therefore are a potential source of big solutions for climate change. Orphan then travels to Boulder, Colorado, where she meets physicist Jun Ye, who is obsessed with measuring time with a clock so precise that it only loses one second in accuracy every million years. As the two scientists converse, they make a potential connection: The secret to the success of greenhouse-eating microbes may be their particular experience with time. At the bottom of the ocean, strong gravitation waves bend and slow down time, reveals Ye. The change isn't a lot, but it may be a clue to understanding the microbes' specific dietary requirements.

From Boulder, Ye travels to Sussex, United Kingdom, where the world's most accurate clock maker comes face to face with neuroscientist Anil Seth, who studies why humans can't keep time without reference points of activity. In Seth's laboratory, subjects in magnetic resonance imaging machines are asked to guess the runtime of video clips. They find that the human perception of time is linked to the activity level in the video. When persons in a video are highly active or objects move quickly, subjects underestimate the length of the video and vice versa for low-activity videos. One potential implication of his work is that although our life spans are increasing, social media and frenzied news cycles may lead to the perception that life is shorter and less fulfilling.

Science is replete with examples of mysteries that were unlocked through the accumulation of sometimes small and often separate advances that came from varied perspectives. The process of discovery is uncertain. What looks like a breakthrough may prove to be a dead end, whereas accidents may provide a result—a piece of a larger puzzle—that suddenly makes all the other separate pieces make sense collectively.

On an uninhabited island off the coast of Puerto Rico, Seth meets Yale psychologist Laurie Santos, who studies deviance—theft, to be exact. Santos is specifically interested in the conditions under which monkeys in the wild risk taking a five-finger discount of succulent treats. In a moment of levity, an accident messes up the experiment to be filmed, but the presence of the documentary crew reveals something unexpected about monkey behavior: Even if the researcher is looking away, attempted theft is less likely to occur when there are potential third-party witnesses. A serendipitous mistake gives Santos her next clue.

Woven into the film's electrifying tapestry of ideas, persons, and places is the story of the hard work of science. Great discovery, we see, comes part and parcel with emotional frustration and disappointment, unusual hours that buck circadian rhythms, tight spaces, cold water, and dark places.

In *The Most Unknown*, Cheney connects apparently disparate journeys of discovery, illustrating where interdisciplinary teamwork can fill in a scientist's blind spots. When seen through this lens, science can be viewed as a big import-export business of ideas. Conventional, well-understood ideas in one area, when brought into another scientific domain, are suddenly seen in a new light.

Newton's observations that "standing on the shoulders of giants" is key to scientific discovery may only have been half right. As documented in *The Most Unknown*, scientists also have much to learn from their contemporaries. ■

The reviewer is at the Kellogg School of Management, the McCormick School of Engineering, and the Northwestern Institute on Complex Systems (NICO), Northwestern University, Evanston, IL 60208, USA. Email: uzzi@kellogg.northwestern.edu

REFERENCE

1. B. Uzzi, S. Mukherjee, M. Stringer, B. Jones, *Science* **342**, 468–472 (2013).

10.1126/science.aau1692

PHOTO: ABRAMORAMA

Brazil may remove restrictions on the trade of introduced species such as this American bullfrog.

PHOTO: ISTOCK.COM/CMCNEILL17

Edited by **Jennifer Sills**

Brazil naturalizes non-native species

Brazil's national policies are putting the country's megadiversity at risk (*1*–*3*). After passing a 2016 law that will put constraints on biodiversity research (*3*), the Ministry of Agriculture, Livestock, and Supply has taken another controversial action: A new ordinance proposes that introduced aquatic species in Brazil should be considered "native" (*4*), including invasive species in the Neotropics (*1*, *5*). Classifying introduced aquatic species as if they are indigenous to Brazil could potentially cause even more introductions and lead to the loss of ecosystem services and functions, as well as traditional knowledge about native species (*6*). Moreover, Brazil shares some large river basins (such as Paraguai, Paraná, and Amazon) with other countries. Therefore, Brazil will become a major source of non-native species for other countries in South America. The rate of introductions in Brazil will likely outpace the research investigating their negative effects (*1*, *3*).

This is not the first time that political decisions have tried to categorize non-native species as native. In 2009, the Brazilian Congress proposed a law that intended to "naturalize by decree" several non-native fishes to foster aquaculture development (*5*). The most recent ordinance is based on a 2016 law that considers introduced species with established populations as part of the Brazilian genetic heritage. Non-native species such as the Malaysian giant prawn (*Macrobrachium rosenbergii*), African catfish (*Clarias gariepinus*), and American bullfrog (*Lithobates catesbeianus*) have established populations in some localities (*7*), and the approval of this new ordinance will permit their free trade and rearing across Brazil.

This retrogression conflicts with several Aichi Biodiversity Targets, especially the one related to the prevention, control, or eradication of non-native species (*8*). Brazil harbors the most diverse aquatic biota in the world (*9*), and it is imperative that local authorities take appropriate measures that value and preserve native biodiversity. Basic research and knowledge produced by scientists (*10*) should play a vital role in these decisions.

Marcelo Fulgêncio Guedes Brito,[1]* André Lincoln Barroso Magalhães,[2] Dilermando Pereira Lima-Junior,[3] Fernando Mayer Pelicice,[4] Valter M. Azevedo-Santos,[5] Diego Azevedo Zoccal Garcia,[6] Almir Manoel Cunico,[7] Jean Ricardo Simões Vitule[8]

[1]Programa de Pós-Graduação em Ecologia e Conservação, Universidade Federal de Sergipe, São Cristóvão, SE, 49100-000, Brazil. [2]Programa de Pós-Graduação em Tecnologias para o Desenvolvimento Sustentável, Universidade Federal de São João Del Rei, Ouro Branco, MG, 36420-000, Brazil. [3]Laboratório de Ecologia e Conservação de Ecossistemas Aquáticos, Universidade Federal do Mato Grosso, Pontal do Araguaia, MT, 78698-000, Brazil. [4]Núcleo de Estudos Ambientais, Universidade Federal do Tocantins, Porto Nacional, TO, 77500-000, Brazil. [5]Universidade Estadual Paulista "Júlio de Mesquita Filho," Botucatu, SP, 18618-970, Brazil. [6]Laboratório de Ecologia de Peixes e Invasões Biológicas, Universidade Estadual de Londrina, Londrina, PR, 86057-970, Brazil. [7]Programa de Pós-Graduação em Aquicultura e Desenvolvimento Sustentável, Universidade Federal do Paraná, Palotina, PR, 85950-000, Brazil. [8]Laboratório de Ecologia e Conservação, Setor de Tecnologia, Departamento de Engenharia Ambiental, Universidade Federal do Paraná, Curitiba, PR, 81531-970, Brazil.
*Corresponding author.
Email: marcelictio@gmail.com

REFERENCES

1. J. R. S. Vitule *et al.*, *Nature* **513**, 315 (2014).
2. J. R. S. Vitule *et al.*, *Science* **347**, 1427 (2015).
3. F. A. Bockmann *et al.*, *Science* **360**, 865 (2018).
4. Portaria Nº 103 (4 May 2018); www.recursosgeneticos.org/Recursos/Arquivos/editor/portariasmc103.pdf [in Portuguese].
5. F. M. Pelicice *et al.*, *Conserv. Lett.* **7**, 55 (2014).
6. K. L. Speziale *et al.*, *Biol. Inv.* **14**, 1609 (2012).
7. A. O. Latini, D. C. Resende, V. B. Pombo, L. Coradin, *Espécies exóticas invasoras de águas continentais no Brasil* (Ministério do Meio Ambiente, 2016) [in Portuguese].
8. D. P. Lima-Junior *et al.*, *Ambio* **47**, 427 (2018).
9. A. A. Padial *et al.*, *Biodivers. Conserv.* **26**, 243 (2017).
10. V. M. Azevedo-Santos *et al.*, *Biodivers. Conserv.* **26**, 1752 (2017).

10.1126/science.aau3368

Free satellite data key to conservation

Biodiversity is in crisis, with extinction rates orders of magnitude higher than background levels (*1*). Underfunded conservationists need to target their limited resources effectively. Over the past decade, satellite remote sensing has revolutionized our ability to monitor biodiversity globally, and is now used routinely, especially by nongovernmental organizations, to detect changes, set priorities, and target conservation action. The U.S. Geological Survey (USGS) unlocked high-resolution Landsat data in 2008 (*2*), making data available online (*3*), and the Copernicus program from the European Commission subsequently made their data available as well (*4*). These resources have been instrumental to biodiversity research. Assessments of environmental changes such as deforestation are now readily available. The current spatial and spectral resolution of Landsat

data make them appropriate to many conservation applications, and although they are not always ideal, pragmatic researchers with limited resources use them regularly. Conservationists have already called for these data to remain free (*5*). Consequently, the news that USGS may charge for data (*6*) is deeply troubling.

USGS has recently convened an advisory committee to determine whether users would be prepared to pay for increased spectral and spatial resolution images (*7*). Requiring users to pay would put these images beyond the reach of conservationists. It would halt time-series analyses that have been useful in monitoring the effects of climate change, land-cover change, and ocean surfaces, likely hindering the achievement of the Sustainable Development Goals (*8*). We urge the USGS to reconsider their position and continue to provide data from the Landsat program freely to all users.

G. M. Buchanan,[1]* A. E. Beresford,[1] M. Hebblewhite,[2] F. J. Escobedo,[3] H. M. De Klerk,[4] P. F. Donald,[5] P. Escribano,[6] L.P. Koh,[7] J. Martínez-López,[8] N. Pettorelli,[9] A. K. Skidmore,[10] Z. Szantoi,[4] K. Tabor,[7] M. Wegmann,[11] S. Wich[12]

[1]RSPB Centre for Conservation Science, Royal Society for the Protection of Birds, Edinburgh, EH12 9DH, UK. [2]Wildlife Biology Program, Department of Ecosystem and Conservation Sciences, W.A. Franke College of Forestry and Conservation, University of Montana, Missoula, MT 59812, USA. [3]Faculty of Natural Sciences and Mathematics, Universidad del Rosario, Bogotá, DC, 11122, Colombia. [4]Department of Geography and Environmental Studies, Stellenbosch University, Stellenbosch 7602, South Africa. [5]BirdLife International, David Attenborough Building, Pembroke Street, Cambridge, CB2 3QZ, UK. [6]CAESCG, University of Almería, Cañada de San Urbano s/n 04120 Almería, Spain. [7]Betty & Gordon Moore Center for Science, Conservation International, Arlington, VA 22202, USA. [8]BC3-Basque Centre for Climate Change, Scientific Campus of the University of the Basque Country, 48940, Leioa, Spain. [9]Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY, UK. [10]University of Twente, Faculty of Geo-Information Science and Earth Observation, 7500 AE Enschede, Netherlands. [11]Institute of Geography and Geology, 97074 Würzburg, Germany. [12]School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, L33AF, UK.
*Corresponding author. Email: graeme.buchanan@rspb.org.uk

REFERENCES

1. J. DeVos *et al.*, *Cons. Biol.* **29**, 452 (2015).
2. C. Woodcock *et al.*, *Science* **320**, 1011 (2008).
3. USGS, Earth Explorer (https://earthexplorer.usgs.gov/).
4. European Commission, Copernicus (http://copernicus.eu/).
5. W. Turner *et al.*, *Biol. Conserv.* **182**, 173 (2015).
6. G. Popkin, *Nature* **556**, 417 (2018).
7. USGS, Landsat Advisory Group undertakes a Landsat Cost Recovery Study (2018); www.usgs.gov/center-news/landsat-advisory-group-undertakes-a-landsat-cost-recovery-study.
8. Sustainable Development Goals (https://sustainabledevelopment.un.org/?menu=1300).

10.1126/science. aau2650

Funding agencies can prevent harassment

Harassment and lack of physical safety in fieldwork and laboratories exists across a range of disciplines (*1*, *2*). Editorials and #MeToo stories have recently highlighted that research is often conducted under "macho" conditions in which harassment, bullying, and unsafe work environments are common (*3*, *4*). In response, codes of conduct for researcher safety are on the rise (*3*, *5*). However, national research funds, private funding organizations, and monitoring agencies rarely require that the recipients of their grants implement codes of conduct or safety standards (*2*). Opportunities for cultural change should rest not only with individual scientists, teams, and professional societies. Funding agencies should share the responsibility.

The cost of ensuring researcher safety should be part of the overall budget, and predefined safety standards should prevent situations in which harassment could occur (*2*). For example, when companies or institutions need scientists to do contracted monitoring work, bidding prices often determine whom they select. Unless funding agencies require safety standards, such bidding prices will always favor low-cost solutions that neglect safety. As another example, when principal investigators (PIs) write applications, they should budget for training and counseling to prevent and address harassment. Such measures would be more widespread if funding agencies acknowledged them.

Funding agencies have the power to participate in changing the culture by requiring codes of conduct for acceptable behavior from their grant recipients. Forcing researchers and companies to incorporate safety standards into grant proposals and assignment bids will increase awareness about harassment and stressful working environments. Only through full support from the broad spectrum of players involved in science will it be possible to create an inclusive and responsible culture that ensures safe workspaces.

Lars L. Iversen[1,2]* and Mette Bendixen[3]

[1]School of Life Sciences, Arizona State University, Tempe, AZ 85281, USA. [2]Center for Macroecology, Evolution and Climate, National Museum of Natural Sciences, University of Copenhagen, Denmark. [3]Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO 80309, USA.
*Corresponding author. Email: lliversen@bio.ku.dk

REFERENCES

1. K. B. Clancy *et al.*, *PLOS ONE* **9**, e102172 (2014).
2. M. A. Rinkus *et al.*, *Soc. Nat. Res.* 10.1080/08941920.2018.1471177 (2018).
3. *Nat. Ecol. Evol.* **1**, 1787 (2017).
4. R. E. Bell, L. S. Koenig, *Science* **358**, 1223 (2017).
5. E. Marín-Spiotta, *Nature* **557**, 141 (2018).

10.1126/science.aau3979

TECHNICAL COMMENT ABSTRACTS

Comment on "Designing river flows to improve food security futures in the Lower Mekong Basin"

John G. Williams

Sabo *et al.* (Research Articles, 8 December 2017, p. 1270) use sophisticated analyses of flow and fishery data from the Lower Mekong Basin to design a "good" hydrograph that, if implemented by planned hydropower dams, would increase the catch by a factor of 3.7. However, the hydrograph is not implementable, and, if it were, it would devastate the fishery. Further, the analyses are questionable.
Full text: dx.doi.org/10.1126/science.aat1225

Comment on "Designing river flows to improve food security futures in the Lower Mekong Basin"

Ashley S. Halls and Peter B. Moyle

The designer flow regime proposed by Sabo *et al.* (Research Articles, 8 December 2017, p. 1270) to support fisheries in the Lower Mekong Basin fails to account for important ecological, political, and economic dimensions. In doing so, they indicate that dam impacts can be easily mitigated. Such an action would serve to increase risks to food and livelihood futures in the basin.
Full text: dx.doi.org/10.1126/science.aat1989

Response to Comments on "Designing river flows to improve food security futures in the Lower Mekong Basin"

G. W. Holtgrieve, M. E. Arias, A. Ruhi, V. Elliott, So Nam, Peng Bun Ngor, T. A. Räsänen, J. L. Sabo

Sabo *et al.* presented an empirically derived algorithm defining the socioecological response of the Tonle Sap Dai fishery in the Cambodian Mekong to basin-scale variation in hydrologic flow regime. Williams suggests that the analysis leading to the algorithm is flawed because of the large distance between the gauge used to measure water levels (hydrology) and the site of harvest for the fishery. Halls and Moyle argue that Sabo *et al.*'s findings are well known, and contend that the algorithm is not a comprehensive assessment of sustainability. We argue that Williams' critique stems from a misunderstanding about our analysis; further clarification of the analysis is provided. We regret not citing more of the work indicated by Halls and Moyle, yet we note that our empirical analysis provides additional new insights into Mekong flow-fishery relationships.
Full text: dx.doi.org/10.1126/science.aat1477

TECHNICAL COMMENT

SUSTAINABILITY

Comment on "Designing river flows to improve food security futures in the Lower Mekong Basin"

John G. Williams

Sabo *et al*. (Research Articles, 8 December 2017, p. 1270) use sophisticated analyses of flow and fishery data from the Lower Mekong Basin to design a "good" hydrograph that, if implemented by planned hydropower dams, would increase the catch by a factor of 3.7. However, the hydrograph is not implementable, and, if it were, it would devastate the fishery. Further, the analyses are questionable.

In a methodological tour de force, Sabo *et al.* (*1*) relate components of the flow regime of the Mekong River, quantified as stage at Stung Treng, to catch in the important Dai fishery on the Tonle Sap River, Cambodia. They use these empirical relationships to design a "good" flow regime proposed as an "ecological objective function" for the operation of controversial dams planned for the Mekong River. They project that this flow regime would increase yield of the Dai fishery by a factor of 3.7. However, simple flow routing considerations show that the "good" regime could not be implemented, and if it were, it would devastate the Dai fishery.

Sabo *et al.* also depict the historical flow regime and a "bad" regime as plots of stage over time (Fig. 1). The plot of the "good" regime is a roughly rectangular wave, with long low-flow periods connected to shorter high-flow periods by transitions ~15 days long. The transitions are more gradual in the "bad" regime.

Stung Treng is several hundred kilometers upstream from the Tonle Sap–Mekong confluence. The Tonle Sap connects the Mekong to a large natural lake and adjacent floodplains that support large populations of many fishes (*2*). Water flows up the Tonle Sap to the Great Lake during high flows in the Mekong, and back down the Tonle Sap to the Mekong during low flows. Flow in the Mekong is a major driver of stage in the Great Lake, but the Great Lake receives ~40% of its water from tributaries and rainfall (*3*), and the stage records differ substantially (Fig. 2).

Sabo *et al.* state that "…we used estimated hydrologic drivers of the historical bag net, or 'Dai,' fishery on the Tonle Sap River—the largest commercial fishery in the Mekong—to design better fisheries futures by comparing designed flows to current and pre-dam (natural-flow) regimes," and they hypothesize that "Flow variation (high and low) may drive production by controlling redox conditions in floodplain soils, …" This implies that they intend their "good" flow regime for the Tonle Sap–Great Lake system, and perhaps for other reaches of the lower Mekong with large floodplains.

Implementing the "good" flow regime would entail managing releases from the most downstream dam, well upstream from the Tonle Sap–Mekong confluence. During the high-flow period, releases would be large and the floodplain would be inundated. At the end of the high-flow period, releases from the dam would be cut to create the desired decrease in stage. Downstream from the dam, water would drain from the floodplain to the channel as soon as releases were cut, reducing the rate of decrease in stage (and discharge) in the channel farther downstream. The opposite would happen, more strongly because of downstream overland flow, at the end of the low-flow season, prolonging the transition from dry to wet conditions. The situation is too complex for simple analysis, but based on hydrodynamic modeling, about 35 km^3 of water flowed onto the left bank of the Mekong upstream and

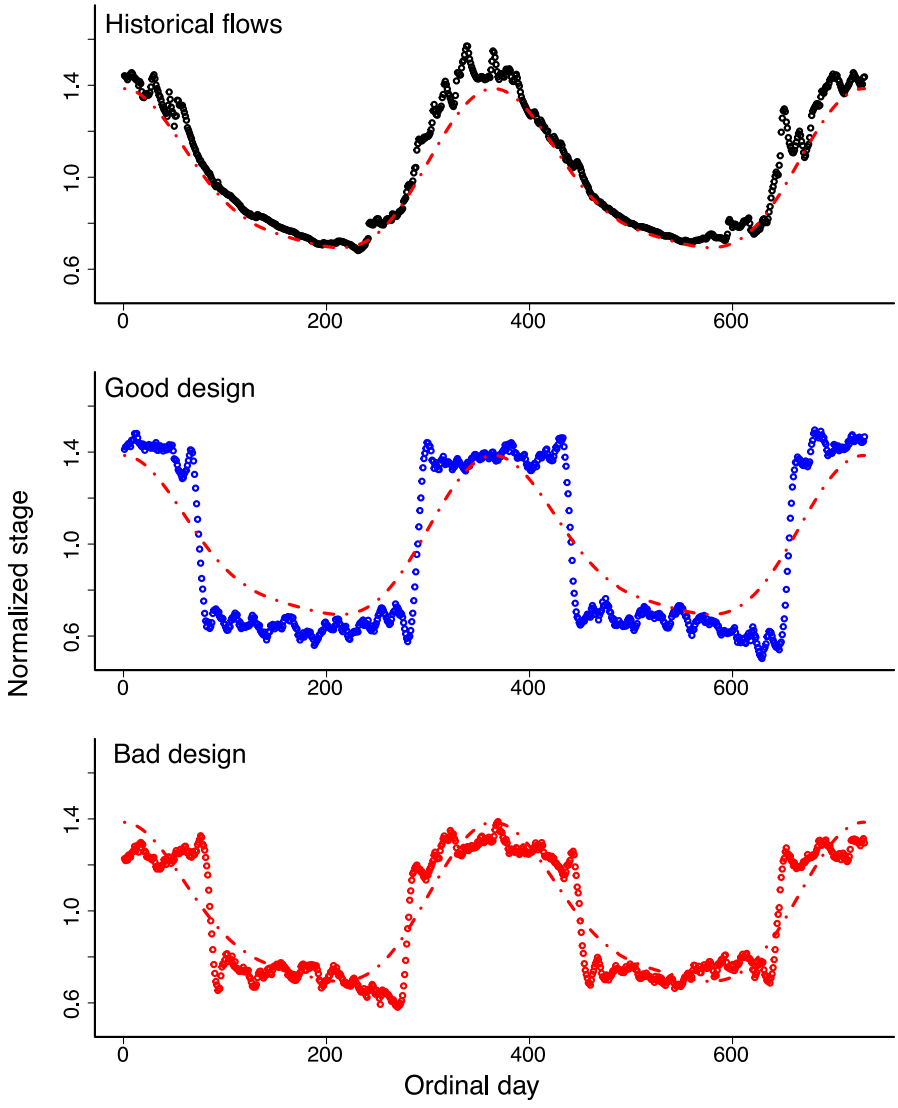


Fig. 1. Illustration of designed flows, including a "good" design and a "bad" design, compared to the reconstructed historical flow regime. [Copied from Sabo *et al*. (*1*)]

Retired consultant. Email: jgwill@frontiernet.net

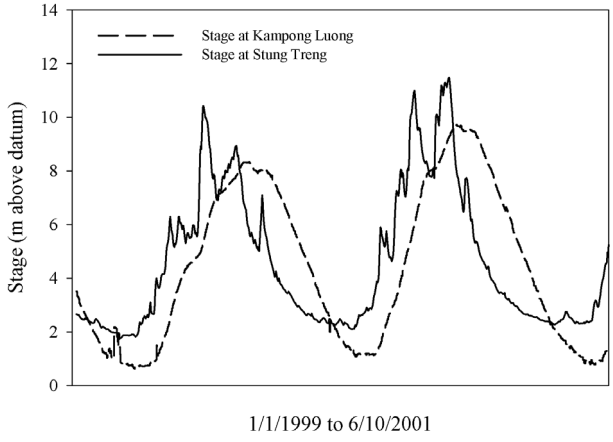


Fig. 2. Stage of the Mekong River at Stung Treng (solid line) and the Great Lake at Kampong Luong (dashed line) for two flood cycles. The flood pulse in the Great Lake lags that at Stung Treng, and high-frequency variation is suppressed. [Data from Sameng Preap, Mekong River Commission]

on the opposite side from the Tonle Sap confluence in 2002 (*4*), so flows between the channel and floodplain are substantial. Thus, the proposed rectangular wave would degrade to something more like the "bad" flow regime as it moved downstream.

Regarding the Tonle Sap, the average annual outflow from the Great Lake exceeds 75 km^3 (*3*). Draining this volume in 15 days would require a constant discharge of ~58,000 m^3 s^{-1}, which is almost 6 times the high flow reported in (*2*) and greater than current peak flows in the Mekong. This discharge could not plausibly be carried by the existing Tonle Sap channel.

The Dai fishery uses what are essentially anchored trawls, with mouths facing upstream, that depend on current to function. The nets target fishes migrating out of the Great Lake to the Mekong River with the falling limb of the hydrograph, as the Great Lake drains over 5+ months (*2*). If somehow the "good" flow regime could be implemented, shortening the falling limb of the hydrograph to ~15 days would truncate the period when the nets might be used by a factor of about 11. Until the channel enlarged in response to the higher flows, the flow probably would be too swift to use the fine mesh nets needed to catch the small fishes that make up the bulk of the Dai fishery catch (*2*), even if it were possible to anchor the nets.

Sabo *et al.* "hypothesized that high fisheries yields are driven by measurable attributes of hydrologic variability, and that these relationships can be used to design and implement future flow regimes that improve fisheries yield through control of impending hydropower operations." Indeed, the size of the flood pulse in the Great Lake is a good predictor of the Dai fishery catch (*2*). Sabo *et al.* found associations between several other attributes of variability in stage at Stung Treng and the catch, but whether any are "hydrologic drivers" of catch remains an open question because the flow-catch records are short (17 years); the analysis was not specified in advance (*5*); other important factors such as habitat cover, sedimentation, and net primary production (*6*) were ignored; and it seems improbable that one drastic change in the flow regime would be good for the fishery whereas a modestly different drastic change would be bad (*7*). Intuitively, variation in stage in the Great Lake seems more likely to influence catch in nets targeting fishes migrating out of the lake than does variation in stage in the Mekong River hundreds of kilometers upstream.

Despite the sophistication of the methods used to develop empirical relationships, predictions based on them depend on the assumption of causality and the assumption that all else will remain more or less equal, which clearly will not be true if the proposed dams are built and the hydrograph is radically altered. Complex computer-based analyses can be marvelously powerful tools, but they should be used carefully, and it is always necessary to think clearly about the results.

REFERENCES

1. J. L. Sabo *et al.*, *Science* **358**, eaao1053 (2017).
2. A. S. Halls, B. R. Paxton, N. Hall, N. Peng Bun, L. Lieng, N. Pengby, N. So, "The Stationary Trawl (Dai) Fishery of the Tonle Sap-Great Lake, Cambodia," MRC Technical Paper No. 32 (Mekong River Commission, Phnom Penh, Cambodia, 2013).
3. M. Kummu, J. Sarkkula, *Ambio* **37**, 185–192 (2008).
4. H. Fujii *et al.*, *Int. J. River Basin Manage.* **1**, 253–266 (2003).
5. A. Gelman, E. Loken, *The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem Even When There Is No "Fishing Expedition" or "p-Hacking" and the Research Hypothesis Was Posited Ahead of Time* (Department of Statistics, Columbia University, 2013); www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf.
6. M. E. Arias *et al.*, *Ecol. Modell.* **272**, 252–263 (2014).
7. J. P. Ioannidis, *PLOS Med.* **2**, e124 (2005).

25 January 2018; accepted 25 May 2018
10.1126/science.aat1225

TECHNICAL COMMENT

SUSTAINABILITY

Comment on "Designing river flows to improve food security futures in the Lower Mekong Basin"

Ashley S. Halls[1] and Peter B. Moyle[2]*

The designer flow regime proposed by Sabo *et al*. (Research Articles, 8 December 2017, p. 1270) to support fisheries in the Lower Mekong Basin fails to account for important ecological, political, and economic dimensions. In doing so, they indicate that dam impacts can be easily mitigated. Such an action would serve to increase risks to food and livelihood futures in the basin.

Sabo *et al*. (*1*) propose that dams in the Lower Mekong Basin (LMB) could produce an artificial ("designer") hydrograph to mitigate impacts on downstream fisheries that feed millions of people. Modeling studies that predict responses of tropical river fisheries to hydrological variation are not new, even in the LMB (*2*–*4*). A study similar to (*1*) was recently undertaken by the Mekong River Commission (MRC) (*5*) using a shorter segment of the same CPUE (catch per unit of effort) time series for the Cambodian Dai fishery. Hydrology was measured in the Tonle Sap River where the fishery takes place, not several hundred kilometers away at Stung Treng as in (*1*).

This MRC study showed that a flood index, termed "FPExt" (flood pulse extent) (*1*), is the primary driver of fisheries yields; fish growth increases exponentially with the flood index, likely reflecting improved feeding opportunities (*6*). The response was modeled "allowing predictions to be made of how…fishery yield is likely to vary under different flooding conditions, whether natural, or modified [by] climate change and/or water management projects [dams]" (*5*).

Sabo *et al*. also found this flood index to be a significant driver of CPUE, together with a composite index of dry-season anomalies. However, the authors appear to have disregarded the MRC findings (*5*): "…previous work suggests two hypotheses to explain how dams diminish production of inland fisheries: (i) by reducing connectivity and dispersal, and (ii) by reducing primary productivity through the entrapment of sediment supply and associated nutrient delivery from headwaters to downstream nursery habitats" (*1*).

Sabo *et al*. explain that their positive relationship between CPUE and FPExt "…may in part be related to increased catchability of fish at high water levels." However, catchability declines in large floods because fish density decreases. Therefore, fishermen in tropical rivers typically fish during the receding flood and dry seasons rather than peak flood (*7*).

The model of Sabo *et al*. also does not take into account other serious impacts of dams besides modifying downstream flows, such as creating barriers to fish migrations. Passage through dams increases mortality and diminishes reproductive success in fish populations. Dams trap nutrient-bearing sediments important for maintaining primary production and fish habitat, thereby diminishing water quality (*8*, *9*). The barrier effects of dams alone could end important fish migrations in the LMB, irrespective of the availability of designer flows (*10*).

The "Good" design flow of (*1*) features large, almost rectangular, flood pulses, punctuated by long dry seasons, and very rapid transitions—approximately 15 days, versus 4 to 5 months under natural conditions. Power production is the primary function of the Mekong dams, which are designed and financed according to predefined operating procedures. Existing (and planned) dams generate downstream flows that are the inverse of the "Good" design flow (i.e., diminished flood pulse and elevated dry-season flow). The "Good" flow would compromise the economic viability of dams in the LMB (*11*). We estimate that Cambodia alone would not have necessary storage (approximately 28 km^3) to generate the "Good" design flow regime (Table 1). Creating the "Good" design flow would require coordination of dam operations by all six countries in the Mekong Basin, only three of which have a major stake in the Lower Mekong River fisheries (Table 1).

Even if the flow could somehow be generated, the designer flow regime is an extrapolation of an empirical model of a pattern that has never been observed in the LMB. The rapid changes in flow between dry and wet seasons threaten fish populations upon which the Dai and other fisheries in the LMB depend (*12*) by removing migration cues, hampering upstream spawning migrations, and sweeping downstream-drifting eggs and larvae past suitable nursery habitats. Stranding and diminished reproductive success under rapidly changing water levels will further diminish fish catch (*13*). Other flow-dependent sectors of the economy, such as flood recession agriculture (rice farming) and navigation, would also be affected by these new flow regimes.

[1]Aquae Sulis Research Ltd., Midway House, Turleigh, Wiltshire BA15 2LR, UK. [2]Department of Wildlife, Fish, and Conservation Biology, Center for Watershed Sciences, University of California, Davis, CA 95616, USA.
***Corresponding author. Email: pbmoyle@ucdavis.edu**

REFERENCES

1. J. L. Sabo *et al*., *Science* **358**, eaao1053 (2017).
2. A. S. Halls, R. L. Welcomme, *River Res. Appl.* **20**, 985–1000 (2004).
3. A. S. Halls, G. P. Kirkwood, A. I. Payne, *Ecohydrol. Hydrobiol.* **1**, 323–339 (2001).
4. R. L. Welcomme, D. Hagborg, *Environ. Biol. Fishes* **2**, 7–24 (1977).
5. A. S. Halls *et al*., "The Stationary Trawl (Dai) Fishery of the Tonle Sap-Great Lake, Cambodia," MRC Technical Paper No. 32 (Mekong River Commission, Vientiane, Lao PDR, 2013).
6. R. L. Welcomme, *River Fisheries* (FAO Fisheries Technical Paper No. 262, Food and Agriculture Organisation of the United Nations, 1985).
7. D. D. Hoggarth *et al*., *Management Guidelines for Asian Floodplain River Fisheries* (FAO Fisheries Technical Paper No. 384, Food and Agriculture Organisation of the United Nations, 1999).
8. G. Marmulla, *Dams, Fish and Fisheries* (FAO Fisheries Technical Paper No. 419, Food and Agriculture Organisation of the United Nations, 2001).
9. G. M. Kondolf *et al*., *Sci. Total Environ.* **625**, 114–134 (2018).

Table 1. Estimates of live storage in existing and planned dams in the LMB and estimated annual fish catch in each LMB country. Estimates are from (*14*) and (*15*), respectively. The necessary live storage to generate the "Good" design flow proposed by Sabo *et al*. (*1*) is approximately 28 km^3.

Country	Live storage (km^3)	Proportion of total live storage	Total annual catch (tonnes)	Proportion of total annual catch
Lao PDR	57.5	68%	208,503	8%
Cambodia	18.9	22%	481,537	20%
Thailand	5.4	6%	911,485	37%
Vietnam	3.2	4%	852,823	35%
Total	85.0	100%	2,454,348	100%

10. A. S. Halls, M. Kshatriya, "Modelling the Cumulative Barrier and Passage Effects of Mainstream Hydropower Dams on Migratory Fish Populations in the Lower Mekong Basin" (Mekong River Commission, Vientiane, Lao PDR, 2009).
11. MRC, "Assessment of Basin-wide Development Scenarios: Main Report. Basin Development Plan Programme, Phase 2" (Mekong River Commission, Vientiane, Lao PDR, 2011); www.mrcmekong.org/assets/Publications/basin-reports/BDP-Assessment-of-Basin-wide-Dev-Scenarios-2011.pdf.
12. A. F. Poulsen *et al*., "Fish Migrations of the Lower Mekong River Basin: Implications for Development, Planning and Environmental Management" (Mekong River Commission, Phnom Penh, Cambodia, 2006).
13. R. L. Welcomme, A. S. Halls, *Ecohydrol. Hydrobiol.* **1**, 313–321 (2001).
14. C. Yermoli, "MRC Hydropwer Database. Hydropower Sector Review for the Joint Planning Process. Basin Development Plan Programme, Phase 2" (Mekong River Commission, Vientiane, Lao PDR, 2009); www.mrcmekong.org/assets/Other-Documents/BDP/BDP2-Regional-Hydropower-Sector-Review-5-Mar-09.pdf.
15. K. G. Hortle, "Consumption and Yield of Fish and Other Aquatic Animals from the Lower Mekong Basin" (Mekong River Commission, Vientiane, Lao PDR, 2007).

2 February 2018; accepted 25 May 2018
10.1126/science.aat1989

TECHNICAL RESPONSE

SUSTAINABILITY

Response to Comments on "Designing river flows to improve food security futures in the Lower Mekong Basin"

G. W. Holtgrieve[1]*, M. E. Arias[2], A. Ruhi[3], V. Elliott[4], So Nam[5], Peng Bun Ngor[6], T. A. Räsänen[7], J. L. Sabo[8,9,10]

Sabo *et al*. presented an empirically derived algorithm defining the socioecological response of the Tonle Sap Dai fishery in the Cambodian Mekong to basin-scale variation in hydrologic flow regime. Williams suggests that the analysis leading to the algorithm is flawed because of the large distance between the gauge used to measure water levels (hydrology) and the site of harvest for the fishery. Halls and Moyle argue that Sabo *et al*.'s findings are well known and contend that the algorithm is not a comprehensive assessment of sustainability. We argue that Williams' critique stems from a misunderstanding about our analysis; further clarification of the analysis is provided. We regret not citing more of the work indicated by Halls and Moyle, yet we note that our empirical analysis provides additional new insights into Mekong flow-fishery relationships.

Sabo *et al*. (*1*) presented an empirical analysis that mathematically defines the socioecological response of an ecosystem (fish catch) to an increasingly human-controlled physical ecosystem driver worldwide (hydrology). This analysis yielded three critical findings: (i) The timing of river discharge, as much as the magnitude, controls fish catches in the Tonle Sap River Dai fishery; (ii) the two aspects of the hydrology most closely related to fish catch—water height during the flood and duration of the low-flow period—have been changing with existing dam development in the direction opposite to what promotes fish catches; and (iii) a "designed" hydrograph has the potential to maintain or even improve fish catches without the constraint of maintaining a completely natural hydrologic regime. Together, these results provide a new perspective on managing flow regimes in heavily developed rivers and thus provide directions for future research, while simultaneously presenting a potentially viable approach for mitigating some of the impacts of dam development on capture fisheries in the Mekong River system.

In his Comment, Williams (*2*) contends that this empirical analysis is flawed because of the distance between the river gauging station used in the analysis and the location of the fishery, and that the "good design" flow scenario presented in Sabo *et al*. is thus unable to be implemented. Both questions are addressed below, which we believe arise from Williams' misunderstanding of the analysis presented in Sabo *et al*. In a second Comment, Halls and Moyle (*3*) note that previous research by Halls and others has identified flood magnitude and duration as important to Dai fishery catches, which were not cited by Sabo *et al*. They also question the use of the Stung Treng gauging station on the Mekong River for our analysis as opposed to the more proximate Prek Kdam gauging station on the Tonle Sap River. Lastly, Halls and Moyle raise concerns about implementation of a designed-flow regime and messaging to regional policy-makers that designed-flow regimes are a complete solution to environmental impacts arising from dam development in the Mekong. We appreciate the contributions of Williams (*2*) and Halls and Moyle (*3*) and the opportunity to discuss these issues.

It is indeed unfortunate that Sabo *et al*. did not cite the full complement of research showing the relationship between fishery catches and flood-pulse magnitude and duration, as we agree that there is substantial literature in support of this result. Beyond this, however, Sabo *et al*. also discerned that the magnitude and duration of the low-flow period is equally as important to fisheries as the high-flow period. This more novel and unexpected result is nonetheless consistent with other research and general theories suggesting that the production of organic matter and nutrients in floodplains during the dry season supports the exceptionally high productivity of flood-pulse ecosystems observed worldwide (*4*–*6*). Governing mechanisms and the magnitude of nutrient and organic transfer from floodplains to fishes remain active areas of research, but current information suggests that their importance to fishery productivity is substantial [e.g., (*7*–*9*)].

The multivariate autoregressive modeling approach used by Sabo *et al*. reveals a statistical relationship between multiple aspects of hydrologic variance at the Stung Treng gauging station—namely, water height at maximum flood and duration of the low-flow period—and annual fish catches at the Dai fishery on the Tonle Sap River. Stung Treng was chosen because it had the longest period of record, which was important for quantifying pre-dam hydrologic conditions. The gauging station in the Tonle Sap River at Prek Kdam is the closest to the fishery, but the water level record was too short to carry out the full analysis. The distance between the gauge station and the fishery is irrelevant, however, as long as the hydrology at the reference location and at the fishery location are consistently related to each other. The two records need not be identical in shape or magnitude in order to be valid for analysis and forecasting, as systematic differences in hydrology between the locations are inherently captured in the empirical analysis.

The data presented in figure 1 of Williams (*2*) are from the Tonle Sap Lake itself at Kampong Luong, yet despite the issues inherent in attempting to compare height of water in a lake basin to height in a flowing river for only 2 years, the data show a high degree of systematic coherence. The more appropriate comparison between the Mekong stage at Stung Treng and the Tonle Sap River stage at Prek Kdam is given in Fig. 1 here. We have also extended the comparison to a 27-year period to better represent interannual variability. This figure shows two important aspects of the system hydrology: (i) the strong coherence between the two records (Fig. 1A), and (ii) the nonlinear relationship on the falling limb of the hydrograph, with stage decreasing faster in the Mekong than at Prek Kdam (Fig. 1B). The former validates the use of Stung Treng as the reference station in our analysis. The latter indicates that because of landscape morphometry, declining river stage at Stung Treng will precede and be more rapid than declines in stage in the Tonle Sap River. Because this effect is inherently captured in our analysis, it renders the calculation by Williams moot, as the "good design" presented in Sabo *et al*. is for the hydrograph at Stung Treng, not at the Tonle Sap River, and it is wrong to assume that the two stations should have identical hydrographs. In short, Williams has made an incorrect assumption that our "good design" hydrograph is to be applied at any location in the basin; it is applicable at Stung Treng only, with the response of fish catch applicable at the Dai fishery only.

Any meaningful assessment of the implementation of designed-flow scenarios in the Mekong will involve a complex modeling exercise that

[1]School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98105, USA. [2]Department of Civil and Environmental Engineering, University of South Florida, Tampa, FL 33620, USA. [3]Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA. [4]Betty and Gordon Moore Center for Science, Conservation International, Arlington, VA 22202, USA. [5]Mekong River Commission Secretariat, Ban Sithane Neua, Sikhottabong District, Vientiane 01000, Lao PDR. [6]Fisheries Administration, No. 186, Preah Norodom Boulevard, Khan Chamcar Morn, P.O. Box 582, Phnom Penh, Cambodia. [7]Consultant, Rinkelitie 3A2, P.O. Box 02580, Siuntio, Finland. [8]Future H_2O, Knowledge Enterprise Development, Arizona State University, Tempe, AZ 85287, USA. [9]Julie Ann Wrigley Global Institute of Sustainability, Arizona State University, Tempe, AZ 85287, USA. [10]Faculty of Ecology, Evolution and Environmental Science, School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA.
*Corresponding author. Email: gholt@uw.edu

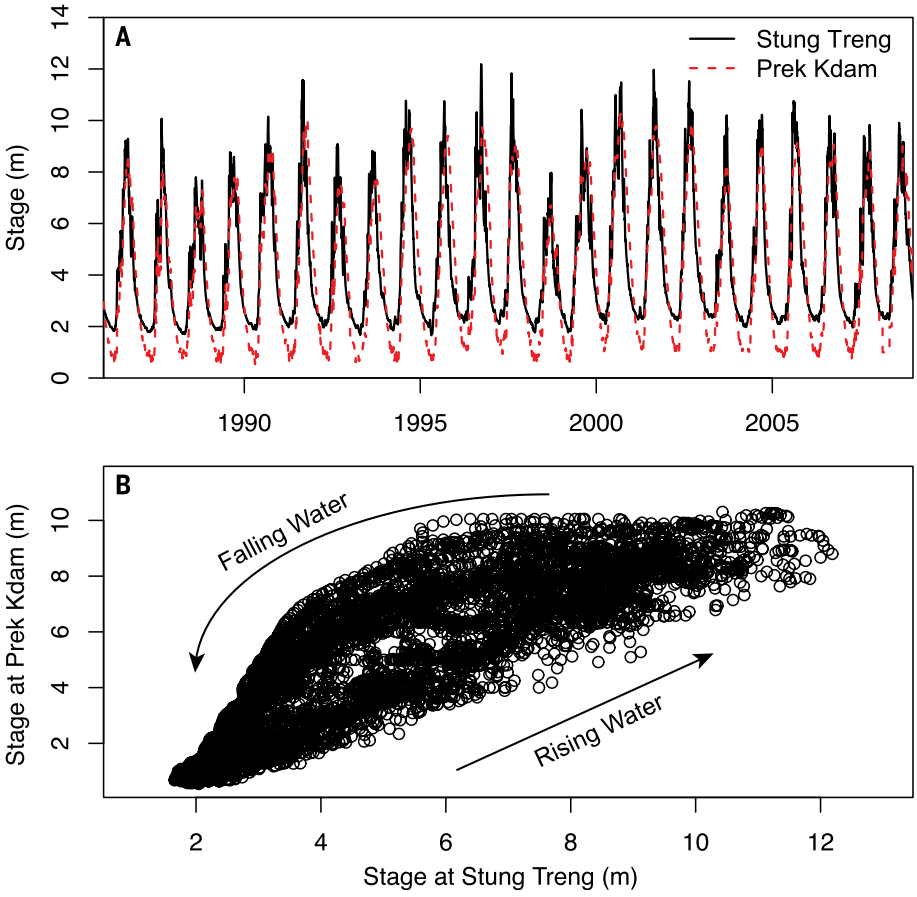


Fig. 1. Hydrograph coherence and nonlinearity in the Lower Mekong Basin. (**A**) River stage at Stung Treng on the Mekong River and Prek Kdam on the Tonle Sap River from 2 January 1986 to 31 December 2008. The correlation coefficient (Pearson *r*) of $\log_{10}$-transformed stage is 0.89. (**B**) The same data plotted against each other, showing a linear relationship in water height between the two stations on the rising limb of the hydrograph but a nonlinear relationship on the falling limb.

considers, at least, movement of water across the landscape and through time, the number and placement of dams throughout the watershed, operation rules for electricity production, coordination (or lack thereof) of dam operations, and future climate. Similarly, as Halls and Moyle (*3*) note, multiple other societal objectives beyond fisheries—such as rice production, flood mitigation, biodiversity protection, and electricity production—must also be taken into account when considering how river hydrology might be managed in the future. This was obviously beyond the scope of Sabo *et al.*, and continued future research is needed to provide a more comprehensive multisectoral analysis. The goal of such an analysis should be to present a broader suite of futures to local policy-makers to better highlight critical cross-sector interdependencies.

Finally, as rapid economic development proceeds in the Lower Mekong Basin, an altered flow regime is one of many ecosystem changes likely to have an impact on fish production and fisheries. Sabo *et al.*'s focus on flow regime does not imply that other potential impacts—for example, barriers to migration, reduced sediment supply, and eutrophication—are unimportant or need not be considered. The potential for unintended ecosystem consequences of a non-natural flow regime must also be further researched and fully considered in the decision-making process. Designing flow regimes is one area where there is scope for management action, however, and Sabo *et al.* provide a method to robustly quantify the relationship between flow regime and fisheries catches in a way that managers can use. With this, there is now means for formal trade-off evaluation useful to policy-makers and a potential path toward mitigating some of the impacts that dam development will have on lower Mekong fisheries. Exactly how designed flow regimes might be implemented in the real world will require substantial additional research and stakeholder engagement.

REFERENCES AND NOTES

1. J. L. Sabo *et al.*, *Science* **358**, eaao1053 (2017).
2. J. G. Williams, *Science* **361**, eaat1225 (2018).
3. A. S. Halls, P. B. Moyle, *Science* **361**, eaat1989 (2018).
4. P. B. Bayley, *Bioscience* **45**, 153–158 (1995).
5. J. W. Junk, P. B. Bayley, R. E. Sparks, *Can. Spec. Publ. Fish. Aquat. Sci.* **106**, 110–127 (1989).
6. R. L. Welcomme, *Inland Fisheries Ecology and Management* (Fishing News Books, 2001).
7. G. W. Holtgrieve *et al.*, *PLOS ONE* **8**, e71395 (2013).
8. D. Lamberts, J. Koponen, *Ambio* **37**, 178–184 (2008).
9. S. B. Correa, K. Winemiller, *Oecologia* **186**, 1069–1078 (2018).

ACKNOWLEDGMENTS

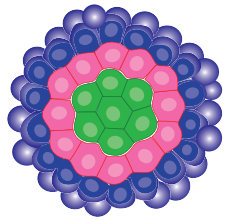
Supported by MacArthur Foundation award 100717 and NSF grant DEB-1541694 (G.W.H.); NSF grant CBET-1204478 (J.L.S.); NSF grant EAR-1740042 (J.L.S. and G.W.H.); the National Socio-Environmental Synthesis Center (SESYNC) with funding from NSF grant DBI-1052875 (A.R.); and MacArthur Foundation awards 100718 and 15-108455-000-CSD to Conservation International. All authors thank the Mekong River Commission and the Inland Fisheries Research and Development Institute (IFReDI) of Cambodia for access to data.

14 February 2018; accepted 25 May 2018
10.1126/science.aat1477

RESEARCH

Modeling tissues with synthetic biology

Toda et al., *p. 156*



IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

A farmer tends flooded deepwater rice fields in Bangladesh.

PLANT SCIENCE

How rice defeats the floodwaters

Deepwater rice varieties grow taller when flooded, in a growth response driven by the plant hormones gibberellin and ethylene. This keeps the leaves above the water. Kuroha *et al*. identified the genes underlying this phenotype, which encode a component of the gibberellin biosynthetic pathway and its regulatory ethylene-responsive transcription factor. This genetic relay drives growth of the plant stem internodes in response to flooding. Modern cultivated deepwater rice, which has been domesticated for adaptation to the monsoon season of Bangladesh, emerged from the genetic variation found in wild rice strains over a broader geographic region. —PJH

Science, this issue p. 181

PHYSICS

Simulating correlated electron systems

Correlated electron systems are generally difficult to simulate because of limited capabilities of computational resources. Harris *et al.* used a D-Wave chip based on a large array of superconducting elements to simulate the phases of a complex magnetic system. They tuned the amount of frustration within the lattice and varied the effective transverse magnetic field, which revealed phase transitions between a paramagnetic, an ordered antiferromagnetic, and a spin-glass phase. The results compare well to theory for this spin-glass problem, validating the approach for simulating problems in materials physics. —ISO

Science, this issue p. 162

GEODETIC MONITORING

Nuclear testing under the radar

North Korea conducted its sixth underground nuclear weapons test in September 2017. The seismic waves generated from the test allow for triangulation and explosive yield estimates. However, Wang *et al.* show that synthetic aperture radar (SAR) should be added to the arsenal of techniques used to detect and characterize nuclear tests. SAR tracks deformation from space, which resulted in a better constraint of source parameters by using deformation from the nuclear test and the subsequent collapse of Mount Mantap. The test occurred at a depth of about half a kilometer, with an explosive yield around 10 times that of the Hiroshima explosion. —BG

Science, this issue p. 166

GREENHOUSE GASES

A leaky endeavor

Considerable amounts of the greenhouse gas methane leak from the U.S. oil and natural gas supply chain. Alvarez *et al.* reassessed the magnitude of this leakage and found that in 2015, supply chain emissions were ~60% higher than the U.S. Environmental Protection Agency inventory estimate. They suggest that this discrepancy exists because current inventory methods miss emissions that occur during abnormal operating conditions. These data, and the methodology used to obtain them, could improve and verify international inventories of greenhouse gases and provide a better understanding of mitigation efforts outlined by the Paris Agreement. —HJS

Science, this issue p. 186

EVOLUTIONARY COGNITION

The impact of time wasted

The amount of time already spent on a task influences human choice about whether to continue. This dedicated

CREDITS: (FROM TOP) S. TODA *ET AL.*; RICHARD I'ANSON/GETTY IMAGES

Published by AAAS

time, known as the "sunk cost," reduces the likelihood of giving up the pursuit of a reward, even when there is no indication of likely success. Sweis *et al.* show that this sensitivity to time invested occurs similarly in mice, rats, and humans (see the Perspective by Brosnan). All three display a resistance to giving up their pursuit of a reward in a foraging context, but only after they have made the decision to pursue the reward. —SNV

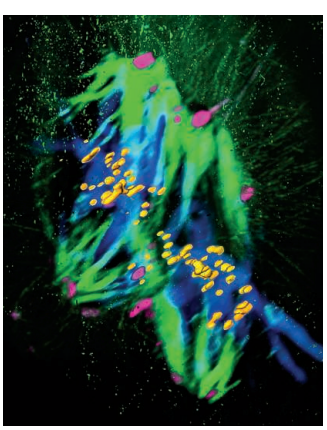
Science, this issue p. 178; see also p. 124

EARLY DEVELOPMENT

It takes two to tango

Fusion of egg and sperm combines the genetic material of both parents in one cell. In mammals, including humans, each parental genome is initially confined in a separate pronucleus. For the new organism to develop, the two genomes must be spatially coordinated so that the first embryonic division can create two cells that combine both genomes in one nucleus. Reichmann *et al.* found that at the beginning of the first division, two microtubule spindles organize the maternal and paternal chromosomes and subsequently align to segregate the parental genomes in parallel (see the Perspective by Zielinska and Schuh). Failure of spindle alignment led to two-celled embryos with more than one nucleus per cell. Dual-spindle assembly in the zygote thus offers a potential mechanistic explanation for division errors frequently observed in human embryos in the fertility clinic. —SMH

Science, this issue p. 189; see also p. 128



Dual spindles separate maternal and paternal chromosomes in parallel.

HEPATITIS C VIRUS

Hampering HCV transmission

No hepatitis C virus (HCV) vaccine is currently available. Furthermore, evidence from studies in nonhuman primates suggests that any future human HCV vaccine would be unlikely to induce complete immunity. Major *et al.* examined whether lowered HCV titers potentially resulting from an imperfect vaccine might still stem HCV transmission in people who inject drugs. They measured the HCV RNA from infected human plasma retained in contaminated needles and syringes. Their mathematical model combining these measurements with published data on HCV viral kinetics suggested that a partially effective vaccine could reduce the HCV transmission risk from sharing contaminated needles and syringes. —CAC

Sci. Transl. Med. **10**, eaao4496 (2018).

CELL BIOLOGY

An inflammatory way to activate AMPK

In its role as a sensor of energy status, the kinase AMPK is activated by phosphorylation mediated by the tumor suppressor LKB1. Antonia and Baldwin found that AMPK activation could also be mediated by a pathway involving TAK1, a kinase associated with inflammatory pathways, and its target, IKK. IKK phosphorylated AMPK independently of LKB1. Combining an IKK inhibitor with the cancer drug phenformin improved the drug's ability to kill LKB1-deficient cancer cells, highlighting a potential treatment for cancers lacking this tumor suppressor. —WW

Sci. Signal. **11**, eaan5850 (2018).

EMERGING THERAPIES

Gene editing gets a head start

The development of gene-editing technologies into therapies for human disease is an exciting prospect. A crucial question is whether there are advantages to correcting disease-causing mutations before rather than after birth, and whether this approach is even feasible. In a proof-of-concept study in mice, Ricciardi *et al.* accomplished successful in utero correction of a hemoglobin gene mutation that causes β-thalassemia, a serious blood disorder. They injected nanoparticles containing gene-editing machinery (triplex-forming peptide nucleic acids and single-stranded donor DNA) intravenously into midgestational mouse fetuses. After birth, the treated mice showed sustained amelioration of anemia and survived longer than untreated mice. —PAK

Nat. Commun. 10.1038/s41467-018-04894-2 (2018).

CANCER

Finding the precise drug

Treating cancer according to the molecular makeup of tumor cells is challenging. How can we find a target that will kill tumor cells and match it to an effective drug? Alvarez *et al.* analyzed the protein-signaling networks in human gastroenteropancreatic neuroendocrine tumors (GEP-NETs) to find master regulators of tumor survival. Subsequent screening of GEP-NET–derived cells with various drugs confirmed that their top hit, entinostat, inhibited GEP-NET growth in vivo. This approach has the potential to provide unbiased precision therapy for many cancers, especially rare cancers, such as GEP-NETs. —GKA

Nat. Genet. 10.1038/s41588-018-0138-4 (2018).

CREDITS: (FROM LEFT) ARTASIOVA/HOSSAIN/REICHMANN/ELLENBERG/EMBL; KTSDESIGN/SCIENCE SOURCE

MICROBIOTA

Cell-to-cell action in cholera

Gut bacteria are competitive. They jostle for niches and nutrients by means of a variety of mechanisms. Gram-negative bacteria devote a lot of resources to the comparatively large (up to 600 nm) type 6 secretion system (T6SS) used to stab neighboring bacterial and host cells. Contrary to expectations, Fast *et al.* discovered that commensal bacteria are not always protective against T6SS-toting pathogens. *Vibrio cholerae* T6SS contributes to host death in the fly model of cholera. *Acetobacter pasteurianus* is a common, normally harmless, commensal in the fly gut. If A. pasteurianus is removed, surprisingly, *V. cholera* is enfeebled, too. This effect seems to be mediated by signaling through the immune deficiency (IMD) pathway of the fly. The IMD is "primed" by the commensal organism. —CA

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.1802165115 (2018).

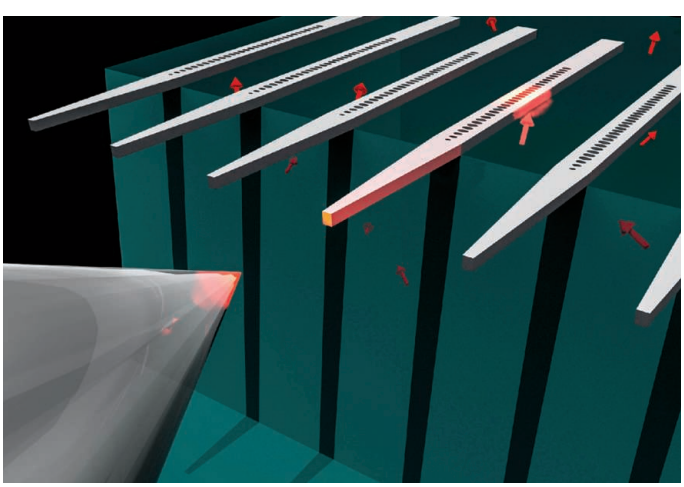
The pathogenicity of *Vibrio cholerae* (illustrated at left) is reduced by removal of a commensal bacterium.

QUANTUM OPTICS

Single photons for optic fibers

Being robust, fast, and able to encode information in a number of different ways, single photons are ideal carriers of quantum information. Exploiting the vast optic fiber network to route single photons between distant points, however, is hampered by the lack of single-photon sources operating at telecom wavelengths. Dibos *et al.* show that single erbium ions doped in a solid-state host material produce single photons at just the right wavelength for long-distance transmission. They use a silicon-based nanophotonic crystal in close proximity to a doped ion to enhance the extraction of single photons and to couple them in an optic fiber. The results demonstrate the basis of an architecture to develop long-distance quantum optical networks. —ISO

Phys. Rev. Lett. **120**, 243601 (2018).



Schematic illustration of silicon waveguides and a YSO crystal

ILLUSTRATION: A. M. DIBOS *ET AL.*, *PHYS. REV. LETT.* **120**, 243601 (2018)

CELL BIOLOGY

mTORC1 jams cell traffic

The cytoplasm of a cell is stuffed with large molecules and organelles. Such crowding may cause some molecules to aggregate, which could in turn affect their biological function. To monitor cytoplasmic crowding, Delarue *et al.* made a probe from bacterial proteins that form large scaffolds linked to fluorescent proteins. In yeast and human cells, the probes revealed that the protein kinase complex mTORC1 (mechanistic target of rapamycin complex 1) enhanced cytoplasmic viscosity. It did so by increasing production of ribosomes while inhibiting their degradation. Because ribosomes occupy about 20% of a cell's volume, mTORC1 activity tunes cytoplasmic crowding and thus influences cell physiology. —LBR

Cell 10.1016/j.cell.2018.05.042 (2018).

CARBON CAPTURE

Help wanted

Preventing global surface air temperatures from rising too high will require emissions to be reduced and CO_2 to be removed (captured) from the air. Many ways to do this have been proposed, although none to date are both technologically and economically feasible, so new and better strategies are needed. Rau *et al.* suggest that generating H_2 with a combination of saline water electrolysis and mineral weathering, powered by electricity not derived from fossil fuels, could greatly increase energy generation and CO_2 removal, at a lower cost than methods involving biomass energy plus carbon capture and storage. This approach also would allow carbon to be sequestered as long-lived ocean alkalinity rather than as concentrated CO_2. —HJS

Nat. Clim. Change **8**, 621 (2018).

THERMOELECTRICS

Lower symmetry for higher performance

Thermoelectric materials interchange heat and electricity. Lowering thermal conductivity while maintaining electrical conductivity is important for developing promising thermoelectric materials. Li *et al.* depart from the usual strategy of using high-symmetry materials by moving from cubic to rhombohedral GeTe, which ends up boosting the thermoelectric performance. The performance metric called the figure of merit is 2.4 at 600 K. The general strategy may be applicable to other materials, providing another pathway to improve performance. —BG

Joule **2**, 976 (2018)

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

PLANT SCIENCE

Genomic traces of symbiosis loss

A symbiosis between certain bacteria and their plant hosts delivers fixed nitrogen to the plants. Griesmann *et al*. sequenced several plant genomes to analyze why nitrogen-fixing symbiosis is irregularly scattered through the evolutionary tree (see the Perspective by Nagy). Various genomes carried traces of lost pathways that could have supported nitrogen-fixing symbiosis. It seems that this symbiosis, which relies on multiple pathways and complex interorganismal signaling, is susceptible to selection and prone to being lost over evolutionary time. —PJH

Science, this issue p. 144; see also p. 125

NEUTRINO ASTROPHYSICS

Neutrino emission from a flaring blazar

Neutrinos interact only very weakly with matter, but giant detectors have succeeded in detecting small numbers of astrophysical neutrinos. Aside from a diffuse background, only two individual sources have been identified: the Sun and a nearby supernova in 1987. A multiteam collaboration detected a high-energy neutrino event whose arrival direction was consistent with a known blazar—a type of quasar with a relativistic jet oriented directly along our line of sight. The blazar, TXS 0506+056, was found to be undergoing a gamma-ray flare, prompting an extensive multiwavelength campaign. Motivated by this discovery, the IceCube collaboration examined lower-energy neutrinos detected over the previous several years, finding an excess emission at the location of the blazar. Thus, blazars are a source of astrophysical neutrinos. —KTS

Science, this issue p. 147, p. 146

SYNTHETIC BIOLOGY

Engineering multilayered cellular structures

The ability to program the manufacture of biological structures may yield new biomaterials or synthetic tissues and organs. Toda *et al*. engineered mammalian "sender" and "receiver" cells with synthetic cell surface ligands and receptors that controlled gene regulatory circuits based on Notch signaling. Programming the cells to express cell adhesion molecules and other regulatory molecules enabled spontaneous formation of multilayered structures, like those that form during embryonic development. The three-layered structures even showed regeneration after injury. —LBR

Science, this issue p. 156

FERROELECTRICITY

Perovskites go organic

The perovskite structure accommodates many different combinations of elements, making it attractive for use in a wide variety of applications. Building perovskites out of only organic compounds is appealing because these materials tend to be flexible, fracture-resistant, and potentially easier to synthesize than their inorganic counterparts. Ye *et al*. describe a previously unknown family of all-organic perovskites, of which they synthesized 23 different family members (see the Perspective by Li and Ji). The compounds are attractive as ferroelectrics, including one compound with properties close to the well-known inorganic ferroelectric $BaTiO_3$. —BG

Science, this issue p. 151; see also p. 132

ORGANIC CHEMISTRY

A silver cleaver splits cyclic amines

Carbon-carbon single bonds are fairly unreactive when they are not strained in a tight ring. Roque *et al*. now report that a silver salt can cleave C–C bonds in unstrained cyclic amines such as pyrrolidines and piperidines. Paired with an electrophilic fluorine source in aqueous solution, the silver first oxidizes the α carbon adjacent to the nitrogen. Ring-opening fluorination of the β carbon then proceeds by an apparent radical mechanism. The reaction offers a versatile means of introducing fluorine to structural motifs common in pharmaceutical research. —JSY

Science, this issue p. 171

EARLY OCEAN

The rise of oxygen

To understand the evolution of the biosphere, we need to know how much oxygen was present in Earth's atmosphere during most of the past 2.5 billion years. However, there are few proxies sensitive enough to quantify O_2 at the low levels present until slightly less than 1 billion years ago. Lu *et al*. measured iodine/calcium ratios in marine carbonates, which are a proxy for dissolved oxygen concentrations in the upper ocean. They found that a major, but temporary, rise in atmospheric O_2 occurred at around 400 million years ago and that O_2 levels underwent a step change to near-modern values around 200 million years ago. —HJS

Science, this issue p. 174

INFECTIOUS DISEASES

Tracking parasite decline

Neglected tropical diseases, such as malaria, are often caused by parasites, the population dynamics of which are difficult to track. Control programs that are aimed at eradicating these diseases would benefit from an ability to monitor parasite populations to assess efficacy and tailor responses. In a Perspective, Cotton *et al.* discuss the use of genomics to track parasite populations and their evolution in the context of control programs. They also discuss how conservation genomics can inform us about the dynamics of extinction. —GKA

Science, this issue p. 130

HUMAN GENETICS

Methylation patterns in early embryos

The methylation state of an embryo can be influenced by its environment, although whether this affects a person's health later in life is unclear. A mother's diet can influence genomic regions that vary greatly in methylation state, known as metastable epialleles (MEs). Kessler *et al.* performed genome-wide screening on 687 candidate MEs in preimplantation embryos of North American Caucasians. Compared with ME patterns previously observed in a Gambian population and in Chinese embryos, regions of the embryonic genome that are sensitive to season of conception harbored many MEs and atypical methylation patterns. Thus, MEs are sensitive to genetic and environmental features. —PJB

Sci. Adv. 10.1126/sciadv.aat2624 (2018).

THYMUS

Metabolic signaling circuits in thymocytes

Cell differentiation is often accompanied by metabolic changes. Yang *et al.* report that generation of double-positive (DP) thymocytes, which have both CD4 and CD8 antigens on their surface, from double-negative (DN) thymocytes coincides with dynamic regulation of glycolytic and oxidative metabolism. Given the central role of mTORC1 (mechanistic target of rapamycin complex 1) signaling in regulating metabolic changes, the authors examined the role of the mTORC1 pathway

in thymocyte development. They conditionally deleted RAPTOR, a key component of the mTORC1 complex. Loss of RAPTOR impaired the DN-to-DP transition, but unexpectedly also perturbed the balance between αβ and γδ T cells. —AB

Sci. Immunol. **3**, eaas9818 (2018).

STRUCTURAL DYNAMICS

Look fast

Organisms from bacteria to humans sense and react to light. Proteins that contain the light-sensitive molecule retinal couple absorption of light to conformational changes that produce a signal or move ions across a membrane. Nogly *et al*. used an x-ray laser to probe the earliest structural changes to the retinal chromophore within microcrystals of the ion pump bacteriorhodopsin (see the Perspective by Moffat). The excited-state retinal wiggles but is held in place so that only one double bond of retinal is capable of isomerizing. A water molecule adjacent to the proton-pumping Schiff base responds to changes in charge distribution in the chromophore even before the movement of atoms begins. —MAF

Science, this issue p. 145; see also p. 127

RESEARCH ARTICLE SUMMARY

PLANT SCIENCE

Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis

Maximilian Griesmann, Yue Chang, Xin Liu, Yue Song, Georg Haberer, Matthew B. Crook, Benjamin Billault-Penneteau, Dominique Lauressergues, Jean Keller, Leandro Imanishi, Yuda Purwana Roswanjaya, Wouter Kohlen, Petar Pujic, Kai Battenberg, Nicole Alloisio, Yuhu Liang, Henk Hilhorst, Marco G. Salgado, Valerie Hocher, Hassen Gherbi, Sergio Svistoonoff, Jeff J. Doyle, Shixu He, Yan Xu, Shanyun Xu, Jing Qu, Qiang Gao, Xiaodong Fang, Yuan Fu, Philippe Normand, Alison M. Berry, Luis G. Wall, Jean-Michel Ané, Katharina Pawlowski, Xun Xu, Huanming Yang, Manuel Spannagl, Klaus F. X. Mayer, Gane Ka-Shu Wong, Martin Parniske*, Pierre-Marc Delaux*, Shifeng Cheng*

INTRODUCTION: Access to nutrients such as nitrogen is required for plant growth. Legumes and nine additional plant families benefit from the nitrogen-fixing root nodule (NFN) symbiosis, in which roots develop nodules that intracellularly host nitrogen-fixing bacteria. In this mutually beneficial symbiosis, the bacteria convert atmospheric nitrogen into ammonium and deliver it to the host plant. NFN symbiosis thus enables plant survival under nitrogen-limiting conditions in terrestrial ecosystems. In agriculture, this symbiosis reduces reliance on nitrogen fertilizer, thus reducing the costs, ecological impact, and fossil fuel consumption attendant on large-scale application of fertilizers.

RATIONALE: Molecular phylogenies show that NFN symbiosis is restricted to four angiosperm orders—Fabales, Fagales, Cucurbitales, and Rosales—that together form the monophyletic NFN clade. However, only 10 of the 28 plant families within this clade contain species engaged in the NFN symbiosis. Even within these 10 families, most genera do not form this symbiosis. The NFN symbiosis requires the coordinated function of more than 30 essential genes. Presence of this symbiosis in related families suggests that a genetic change in the ancestor of the NFN clade enabled evolution of NFN symbiosis in this clade. The scattered distribution of functional NFN symbiosis across the clade has led to the question of whether NFN symbiosis evolved multiple times independently in a convergent manner or was lost multiple times regardless of the number of times it arose. Fossil data have been unable to answer this question. Here we used molecular evidence to ask how the current pattern of plant species with NFN symbiosis evolved.

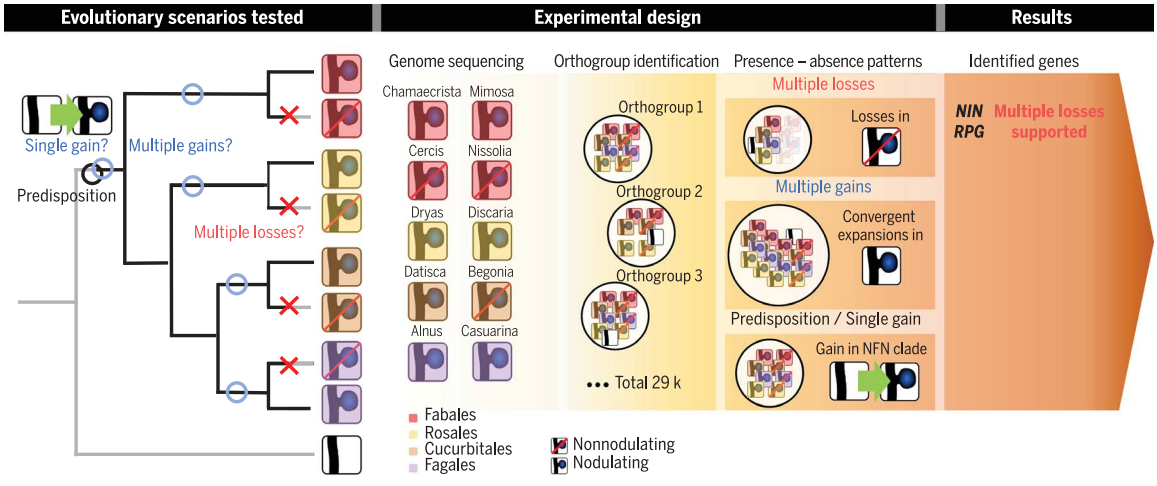
RESULTS: We sequenced the genomes of seven nodulating species belonging to the Fagales, Rosales, and Cucurbitales orders and the legume subfamily Caesalpinioideae. We complemented this dataset by sequencing three genomes of nonnodulating species from the Cucurbitales and from the legume subfamilies Cercidoideae and Papilionoideae. Using a genome-wide phylogenomic approach, we found that all legume genes with a characterized role in NFN symbiosis are conserved in nodulating species with one exception. We observed larger numbers of order-specific gene family expansions that, solely because of their phylogenetic distribution, may include genes contributing to multiple gains or subsequent refinements of the symbiosis. In parallel, we discovered signatures of multiple independent loss-of-function events for the gene encoding the indispensable NFN symbiosis regulator *NODULE INCEPTION* (*NIN*) in 10 of 13 genomes of nonnodulating species within the NFN clade. The pattern suggests at least eight independent losses of NFN symbiosis.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat1743

CONCLUSION: We found that multiple independent losses of NFN symbiosis occurred in the four orders of the NFN clade. These results suggest that NFN symbiosis has previously been more common than currently evident and that this symbiosis is subject to an underestimated adverse selection pressure. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: chengshf@genomics.cn (S.C.); pierre-marc.delaux@lrsv.ups-tlse.fr (P.-M.D.); parniske@lmu.de (M.P.)
Cite this article as M. Griesmann *et al.*, *Science* **361**, eaat1743 (2018). DOI: 10.1126/science.aat1743



Phylogenomics and evolution of NFN symbiosis. Genome sequencing of nodulating and nonnodulating species combined with 27 previously available genomes resulted in a dataset spanning the NFN clade and species outside the NFN clade as an outgroup. Orthogroups were identified and filtered following three phylogenetic patterns. This genome-wide analysis identified two genes involved in NFN symbiosis, *NIN* and *RHIZOBIUM-DIRECTED POLAR GROWTH* (*RPG*), that were lost in most nonnodulating species. The occurrence of multiple losses (red crosses) of NFN symbiosis suggests an adverse selection pressure.

RESEARCH ARTICLE

PLANT SCIENCE

Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis

Maximilian Griesmann[1,2], Yue Chang[3,4], Xin Liu[3,4], Yue Song[3,4], Georg Haberer[2], Matthew B. Crook[5], Benjamin Billault-Penneteau[1], Dominique Lauressergues[6], Jean Keller[6], Leandro Imanishi[7], Yuda Purwana Roswanjaya[8], Wouter Kohlen[8], Petar Pujic[9], Kai Battenberg[10], Nicole Alloisio[9], Yuhu Liang[3,4], Henk Hilhorst[11], Marco G. Salgado[12], Valerie Hocher[13], Hassen Gherbi[13], Sergio Svistoonoff[13], Jeff J. Doyle[14], Shixu He[3,4], Yan Xu[3,4], Shanyun Xu[3,4], Jing Qu[3,4], Qiang Gao[3,15], Xiaodong Fang[3,15], Yuan Fu[3,4], Philippe Normand[9], Alison M. Berry[10], Luis G. Wall[7], Jean-Michel Ané[16,17], Katharina Pawlowski[12], Xun Xu[3,4], Huanming Yang[3,18], Manuel Spannagl[2], Klaus F. X. Mayer[2,19], Gane Ka-Shu Wong[3,20,21], Martin Parniske[1]*, Pierre-Marc Delaux[6]*, Shifeng Cheng[3,4]*

The root nodule symbiosis of plants with nitrogen-fixing bacteria affects global nitrogen cycles and food production but is restricted to a subset of genera within a single clade of flowering plants. To explore the genetic basis for this scattered occurrence, we sequenced the genomes of 10 plant species covering the diversity of nodule morphotypes, bacterial symbionts, and infection strategies. In a genome-wide comparative analysis of a total of 37 plant species, we discovered signatures of multiple independent loss-of-function events in the indispensable symbiotic regulator *NODULE INCEPTION* in 10 of 13 genomes of nonnodulating species within this clade. The discovery that multiple independent losses shaped the present-day distribution of nitrogen-fixing root nodule symbiosis in plants reveals a phylogenetically wider distribution in evolutionary history and a so-far-underestimated selection pressure against this symbiosis.

Nitrogen is one of the main requirements for plant growth. Members of the legume family (Fabaceae, order Fabales) and of nine additional plant families benefit from symbiosis with nitrogen-fixing bacteria, either phylogenetically diverse rhizobia or species of the genus *Frankia*, which are hosted inside plant cells found within nodules—specialized host-derived lateral organs typically found on roots. In this mutualistic nitrogen-fixing root nodule (NFN) symbiosis, intracellular bacteria convert atmospheric nitrogen into ammonium by means of the enzyme nitrogenase (*1*). This "fixed nitrogen," delivered to the host plant, is an essential building block for amino acids, DNA, RNA, tetrapyrroles such as chlorophyll, and many other molecules. This symbiosis enables plant survival under nitrogen-limiting conditions. In agriculture, this independence from chemical nitrogen fertilizer reduces costs and fossil fuel consumption imposed by the Haber-Bosch process (*2*). Since the discovery of the NFN symbiosis, with rhizobia in 1888 and with *Frankia* in 1895 (*3*, *4*), it has been unclear why it is restricted to only a limited number of flowering plant species.

A major scientific step forward in our understanding was the reorganization of the phylogenetic tree of angiosperms in 1995, which revealed that plants forming the NFN symbiosis are restricted to the Fabales, Fagales, Cucurbitales, and Rosales orders, which together form the NFN clade (*5*, *6*). However, this reorganization also raised new questions, because only 10 of the 28 plant families within the NFN clade contain plants that form nodules (referred to here as "nodulating species") and these do not form a monophyletic group; moreover, within 9 of these 10 families, most genera do not form NFN symbiosis (*7*). In addition to this scattered distribution, a further unsolved mystery that surrounds the evolution of NFN symbiosis is its diversity at multiple levels: Legumes (Fabales) and the nonlegume *Parasponia* (Rosales) (*8*) form nodules with rhizobia, whereas species of the actinobacterial genus *Frankia* infect actinorhizal plants from eight plant families of the orders Fagales, Cucurbitales, and Rosales (*9*). A diversity of infection mechanisms has been described (*9*), and root nodule structures display wide variations (*5*, *7*). The most parsimonious hypothesis to explain the restricted and yet scattered distribution pattern of such diverse NFN symbioses predicted a genetic change in the ancestor of the NFN clade, a predisposition event, that enabled the subsequent independent evolution of NFN symbiosis specifically and exclusively in this clade (the multiple-origins hypothesis), along with a number of losses (*5*, *7*, *10*–*12*). Recent quantitative phylogenetic modeling studies supported scenarios with independent gains and switches between the nonnodulating and nodulating states during the evolution of the NFN clade (*11*, *13*). However, none of these analyses provide direct evidence or the molecular causes of the specific gains and losses that explain the distribution of NFN symbiosis in extant genera.

Exploring the genetic basis underlying the evolutionary dynamics of the NFN symbiosis in plants will improve our understanding of the diversity of symbiotic associations observed in extant taxa and the ecosystems they inhabit and potentially provide keys to engineer it in crops and to predict the stability of this trait over long evolutionary times. Here we used a genome-wide comparison including genomic and phylogenomic methods (*14*) to address the long-standing conundrum of the evolution of NFN symbiosis and identify the underlying genetic players (*15*, *16*).

Results

Genome sequencing in the NFN clade

Sequenced genomes of nodulating species were only available for a few agriculturally relevant legume species belonging to a single subfamily (Papilionoideae), all derived from a single predicted evolutionary origin of NFN symbiosis, with no representation either of taxa representing possible additional origins within legumes or of nonlegume nodulating species widely accepted as representing multiple additional origins (*17*–*19*).

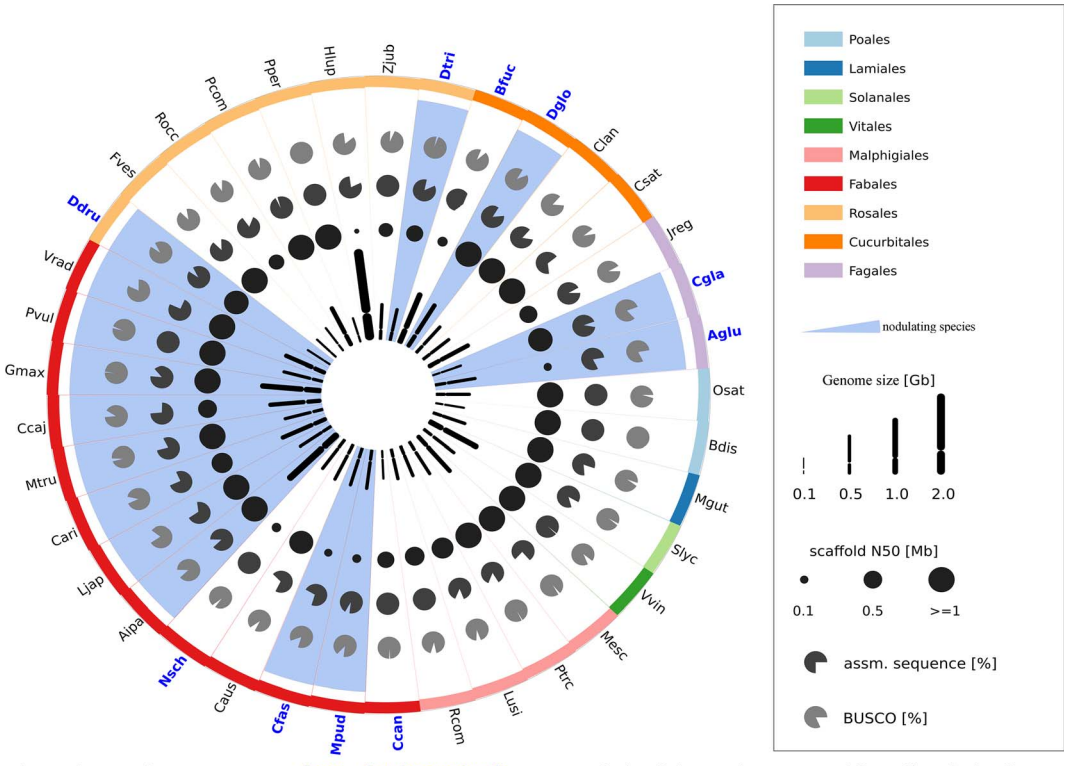
[1]Faculty of Biology, Genetics, LMU Munich, Großhaderner Strasse 2-4, 82152 Martinsried, Germany. [2]Plant Genome and Systems Biology (PGSB), Helmholtz Center Munich–German Research Center for Environmental Health, 85764 Neuherberg, Germany. [3]BGI-Shenzhen, Shenzhen 518083, China. [4]China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China. [5]Department of Microbiology, Weber State University, Ogden, UT 84408-2506, USA. [6]Laboratoire de Recherche en Sciences Végétales (LRSV), Université de Toulouse, CNRS, UPS, 24 chemin de Borde Rouge, Auzeville, BP42617, 31326 Castanet Tolosan, France. [7]LBMIBS, Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, CONICET, R. Saénz Peña 352, B1876BXD Bernal, Argentina. [8]Laboratory for Molecular Biology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, Netherlands. [9]Université Lyon 1, Université de Lyon, CNRS, Ecologie Microbienne, UMR 5557, Villeurbanne, 69622 Cedex, France. [10]Department of Plant Sciences, University of California, Davis, Davis, CA 95616, USA. [11]Laboratory for Plant Physiology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, Netherlands. [12]Department of Ecology, Environment and Plant Sciences, Stockholm University, 106 91 Stockholm, Sweden. [13]French National Research Institute for Sustainable Development (IRD), UMR LSTM (IRD–INRA–CIRAD–Université Montpellier–Supagro), Campus International de Baillarguet, TA A-82/J, 34398, Montpellier Cedex 5, France. [14]Section of Plant Breeding and Genetics, School of Integrated Plant Science, Cornell University, Ithaca, NY 14853, USA. [15]BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China. [16]Department of Agronomy, University of Wisconsin–Madison, Madison, WI 53706, USA. [17]Department of Bacteriology, University of Wisconsin–Madison, Madison, WI 53706, USA. [18]James D. Watson Institute of Genome Sciences, Hangzhou 310058, China. [19]TUM School of Life Sciences Weihenstephan, Technical University of Munich, 85354 Freising, Germany. [20]Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada. [21]Department of Medicine, University of Alberta, Edmonton, AB T6G 2E, Canada.
***Corresponding author. Email: chengshf@genomics.cn (S.C.); pierre-marc.delaux@lrsv.ups-tlse.fr (P.-M.D.); parniske@lmu.de (M.P.)**

Conversely, sequenced genomes of nonnodulating species were restricted to the Fagales, Rosales, and Cucurbitales and did not include nonnodulating legume taxa (table S1). To overcome this sampling bias, which restricted the phylogenomic analysis, we sequenced de novo the genomes of seven nodulating species belonging to the Cucurbitales, Fagales, and Rosales orders and the Caesalpinioideae subfamily of the Fabaceae, representing a possible second origin of NFN symbiosis in legumes. Three nonnodulating species from the NFN clade were also sequenced, notably *Nissolia schotii*, a papilionoid legume that has lost the ability to form the NFN symbiosis and which therefore provides insights on the genomic consequences of losing the symbiosis (Fig. 1 and fig. S1). For each species, 144 to 381 Gb of Illumina reads were obtained, covering the estimated genomes at least 189-fold and up to 1113-fold, with the resulting scaffold N50 length between 96 kb and 1.18 Mb and an average genome completeness of 96% (Fig. 1, figs. S1 and S2, and tables S2 to S29). Altogether, the genomes of species sequenced here represent six families and most known nodule anatomy types and root infection pathways; include hosts for the main classes of nodule-inducing symbiotic nitrogen-fixing α-proteobacteria, β-proteobacteria, and actinobacteria; and cover six to seven independent evolutionary origins of NFN symbiosis, according to the multiple-gains hypothesis (*5*, *7*, *10*, *11*). These 10 sequenced genomes, together with 18 other genomes from the NFN clade and 9 genomes from other flowering plants as an outgroup (Fig. 1), were compared to detect molecular traces supporting any of the three postulated events in the evolution of NFN symbiosis: (i) predisposition to evolve it, (ii) multiple independent gains, and (iii) multiple independent losses of NFN symbiosis.

The putative predisposition event did not involve NFN clade-specific gene gains

The predisposition event postulated by Soltis *et al.* in 1995 (*5*) may be based on the acquisition of one or several genes or sequence modifications specific to the NFN clade. This acquisition would be also consistent with a single-origin hypothesis, in which NFN symbiosis in all taxa is predicted as a homologous trait. Genes acquired during the predisposition event are expected to be specific to the NFN clade and present in all nodulating species. To search for genes following this evolutionary pattern, we identified gene families across all 37 plant genomes in our dataset using the Orthofinder pipeline (*20*). For each of the resulting 29,433 gene family clusters, we calculated a separate phylogeny and subsequently inferred orthologs for all genes of the reference species, *Medicago truncatula* (*16*). We selected groups for which orthologs were absent in the nine species outside of the NFN clade but were retained in nodulating species (fig. S3). To obtain a candidate set for manual validation from the total of 29,213 orthologous groups, we used an automated filtering approach with relaxed criteria, which allowed for the absence of orthologs in a small subset of nodulating species (*16*). This step was necessary to avoid the loss of putative candidates owing to missed gene models resulting in false negatives, as they often occur in automated gene-prediction pipelines. Our relaxed filter identified a total of 31 orthologous candidate groups (table S30). All of these candidate groups underwent an iterative manual curation, including a search for missed gene models and recalculation of phylogenies and orthologs (*16*).

Not a single candidate gene was identified that matched the evolutionary pattern expected for predisposition-related genes, suggesting that genes gained in the most recent common ancestor of the NFN clade have been conserved or lost irrespectively of the symbiotic state of the lineages. If the predisposition indeed occurred, this result indicates that it did not involve the



Osat	*Oryza sativa*	**Ccan**	***Cercis canadensis***	Ccaj	*Cajanus cajan*	Hlup	*Humulus lupulus*
Bdis	*Brachypodium distachyon*	**Mpud**	***Mimosa pudica***	Gmax	*Glycine max*	Zjub	*Ziziphus jujuba*
Mgut	*Mimulus guttatus*	**Cfas**	***Chamaecrista fasciculata***	Pvul	*Phaseolus vulgaris*	**Dtri**	***Discaria trinervis***
Slyc	*Solanum lycopersicum*	Caus	*Castanospermum australe*	Vrad	*Vigna radiata*	**Bfuc**	***Begonia fuchsoides***
Vvin	*Vitis vinifera*	**Nsch**	***Nissolia schottii***	**Ddru**	***Dryas drummondii***	**Dglo**	***Datisca glomerata***
Mesc	*Manihot esculenta*	Aipa	*Arachis ipaensis*	Fves	*Fragaria vesca*	Clan	*Citrullus lanatus*
Ptrc	*Populus trichocarpa*	Ljap	*Lotus japonicus*	Rocc	*Rubus occidentalis*	Jreg	*Juglans regia*
Lusi	*Linum usitatissimum*	Cari	*Cicer arietinum*	Pcom	*Pyrus communis*	Csat	*Cucumis sativus*
Rcom	*Ricinus communis*	Mtru	*Medicago truncatula*	Pper	*Prunus persica*	**Cgla**	***Casuarina glauca***
						Aglu	***Alnus glutinosa***

Fig. 1. Genome features of species used in this study. Genome statistics are shown as pictograms for species used in this study, with nodulating species highlighted by blue sectors. Species names are shown as four-letter abbreviations at the outer circle, with their taxonomic order color coded as shown in the top-right legend. Newly sequenced species are in bold blue letters. The next two circles show, as pie charts, the proportion of complete BUSCO genes [see (*68*)] detected in the genome assembly (light gray) and the percentage of assembled sequence relative to the estimated genome size (dark gray), respectively. Scaffold N50 values are depicted as bubble charts (black) capped to a maximal N50 of 1 Mb to reduce graphical biases by finished genomes assembled to pseudochromosomes. Note that even for assemblies in this study with low contiguity, the BUSCO results suggest that the gene space has been well covered (tables S4 and S5). The innermost circle represents the genome size by proportional chromosome pictograms.

acquisition of genes but rather the co-option of existing genes and their corresponding pathways.

Gene family dynamics is compatible with multiple gains

Multiple molecular mechanisms leading to the convergent evolution of a trait have been identified (*21*). Deep homology (*22*), the independent recruitment of a homologous gene set for the development of nonhomologous traits, has been proposed for the NFN symbiosis (*7*, *23*). Indeed, several genes initially identified for their symbiotic role in legumes were later found to also play a symbiotic role in Fagales (*24*–*27*), Rosales (*28*, *29*), and Cucurbitales (*30*). An alternative mechanism for the evolution of new traits is gene family expansion, as exemplified by the parallel diversification of the zinc-finger transcription factor family in the evolution of a dominant yeast form in fungi (*31*) or the acquisition of strigolactone perception in parasitic plants (*32*). We hypothesized that if NFN symbiosis evolved multiple times, independent expansions in the same gene family might have been involved. Given that our dataset covers six to seven of the predicted independent gains of NFN symbiosis, it allowed us to search for gene families whose evolutionary patterns are consistent and would support the multiple-gains hypothesis (*7*). We analyzed Orthofinder clusters for copy-number variation to identify gene family expansion events at each of these nodes (Fig. 2) (*16*). Multiple alternative models have been proposed for the independent gain of NFN symbiosis. In one scenario (*7*), a single gain before the radiation of the legume family has been suggested that correlates with the expansion of 33 clusters in our analysis (Fig. 2). To test whether any of these clusters were enriched with differentially expressed genes (DEGs) in nodule tissue versus root tissue, we derived for *M. truncatula* (Medicago) gene expression data for both conditions from the gene expression atlas (*33*) and tested for an enrichment of Medicago DEGs in each cluster (table S31). We only found one such cluster that belongs to the nitrate transporter family NRT1/PTR (*34*). However, this gene family appears to also be expanded at nodes not related to independent gains within and outside of the NFN clade (table S31). An alternative model proposes two independent gains within the legumes: the first at the most recent common ancestor of the papilionoids and the clade to which *Castanospermum australe* belongs (21-291 enriched-expanded clusters), or at the base of the papilionoids (4-32), and the second gain at the base of the caesalpinioid and mimosoid clade (7-92). This clade could alternatively comprise two independent gains for *Chamaecrista fasciculata* (39-710) and mimosoids (37-723). Larger numbers of expanded gene families were observed for the predicted events in the Rosales, Fagales, and Cucurbitales (Fig. 2 and table S31). Taken together, we found 52 gene family clusters that were enriched with differentially expressed genes in Medicago nodules and expanded multiple times at proposed independent gain nodes. However, similar to the NRT1/PTR family, all of the enriched clusters also expanded outside the NFN clade. Inside the NFN clade, these clusters expanded beside the hypothesized gain of NFN symbiosis nodes at many additional nodes (table S31). In our survey of all independent gene family expansions, we did not identify any cluster that displayed parallel expansions, one possibility among others that would indicate convergent recruitment for the independent gains of NFN symbiosis (*7*, *16*). If NFN symbiosis indeed evolved multiple times, our genome-wide analysis revealed hundreds of clade-specific candidate genes that, together with gene co-option, may have played a role in the putative independent evolutions of this trait.

Genomic evidence for multiple losses

Testing homologies of a trait shared by multiple taxa typically involves assessing the trait in these species and inferring origins of the trait once homology is accepted or rejected. However, information on the origin of a trait, and thus its homology, can also be obtained from taxa lacking the trait, by distinguishing primary absence (the taxon never had the trait) from secondary loss of the trait (*23*). It has been demonstrated that genes involved in a specific biological process are lost following the loss of that trait, a process known as gene coelimination (*35*–*37*). To test the multiple-losses hypothesis, we searched the ortholog groups calculated above for an evolutionary pattern that retained orthologs in all nodulating species but lost them in nonnodulating ones. To filter and confirm the list of candidate orthologous groups, we used the same two-step process combining an automated pipeline with relaxed criteria for nodulating species followed by a careful manual curation and evaluation step with stringent criteria (presence in all nodulating species required). The automated pipeline with relaxed criteria resulted in a list of 121 candidate groups (table S32). During our manual confirmation and refinement, we rejected 31 of these candidate groups because orthologs were absent from one or more nodulating species. Another 62 candidate groups were rejected because orthologs of more than 50% of nonnodulating species were present. A weak phylogenetic signal did not allow for inferring a reliable orthology for 27 candidate groups. A single gene, *NODULE INCEPTION* (*NIN*), was confirmed to be present in the genomes of all the nodulating species and in the genome of species outside the NFN clade, and absent from most nonnodulating ones in the NFN clade (Fig. 2 and fig. S4). Forward genetic screens in the legumes *Lotus japonicus* (*27*), *Pisum sativum* (*38*), and *M. truncatula* (*24*) identified *NIN* as indispensable for the two developmental aspects of the NFN symbiosis: initiation of root nodule development and the formation of the plant structure facilitating intracellular uptake of bacteria (*27*). Furthermore, RNA interference–based suppression of *NIN* expression in *Casuarina glauca* (Fagaceae, Fagales) impaired nodule formation (*25*), consistent with a conservation of the role of *NIN* in NFN symbiosis in actinorhizal host plants.

Confirmation of NIN absence by microsynteny

In addition to the presence of *NIN* in all the nodulating species in our dataset, synteny analysis revealed the conservation of the syntenic blocks surrounding the *NIN* locus across the NFN clade (Fig. 2). By contrast, 10 of 13 nonnodulating species of the Fabales, Fagales, Cucurbitales, and Rosales underwent partial (four species) or complete (six species) deletions of *NIN* from the conserved genomic block (Fig. 2). The legume family is divided into six subfamilies, two of which include nodulating species (*39*). According to previous estimates (*7*), *Cercis* (a member of Cercidoideae, which is sister to most or all other legumes) and *Castanospermum* (a member of a clade sister to the crown group of papilionoid legumes, in which nearly all members form the NFN symbiosis) both represent lineages in which NFN symbiosis never occurred. By contrast, *Nissolia* and its sister genus *Chaetocalyx* are nonnodulating genera nested within the nodulating crown group of papilionoids and thus have been predicted to represent secondary loss of NFN symbiosis (*11*, *40*). Our synteny approach allowed the discovery of the complete absence of *NIN* from the genome of *N. schottii* (Fig. 2). In addition, we confirmed the absence of *NIN* in three *Chaetocalyx* species (fig. S5). Because *NIN* was present in the most recent common ancestor of the NFN clade and conserved in nodulating species, the absence of this gene in the *Nissolia-Chaetocalyx* lineage represents a loss that correlates with, and is sufficient to explain, the loss of NFN symbiosis. *Cercis canadensis* harbored only a *NIN* pseudogene remnant in the genomic block, whereas the genome of *C. australe* completely lacked *NIN* (Fig. 2). These results demonstrate three independent losses of *NIN* in the legume family. Similarly, the synteny analyses confirmed a minimum of three independent losses in nonnodulating Rosales and two in the Cucurbitales (Fig. 2). Together, the diversity of *NIN* deletions in the nonnodulating species is indicative of at least eight independent evolutionary events that led to the loss of *NIN* function (Fig. 3). Following the multiple-gains hypothesis (*7*), NFN symbiosis was predicted to have evolved independently at least six times in the species space sampled here (Fig. 2). Loss of *NIN* provides an alternative model with at least eight independent losses of NFN symbiosis (Fig. 3). Thus, the current distribution of NFN symbiosis might be a combination of these two complementary and not mutually exclusive models.

Loss of RPG

In parallel with these multiple confirmed losses of *NIN*, the synteny analysis also confirmed the presence of *NIN* in three nonnodulators of the NFN clade (Fig. 2). We hypothesized that the loss of genes other than *NIN* may explain the nonnodulating state of these species and that such

genes may have been missed by the specific and stringent criteria for the detection of the presence-absence patterns. In addition to *NIN*, we determined the presence-absence pattern of 21 genes that were identified by forward and reverse genetics as critical for NFN symbiosis in legumes (Fig. 3 and table S33). Among these genes, 20 were conserved in nodulating species and most nonnodulating ones (figs. S6 to S25). By contrast, *RHIZOBIUM-DIRECTED POLAR GROWTH* (*RPG*) (*41*), which is present outside the NFN clade, is missing in *N. schottii* and 11 other nonnodulating species from the four orders of the NFN clade (Fig. 3). These losses were confirmed by microsynteny analyses for

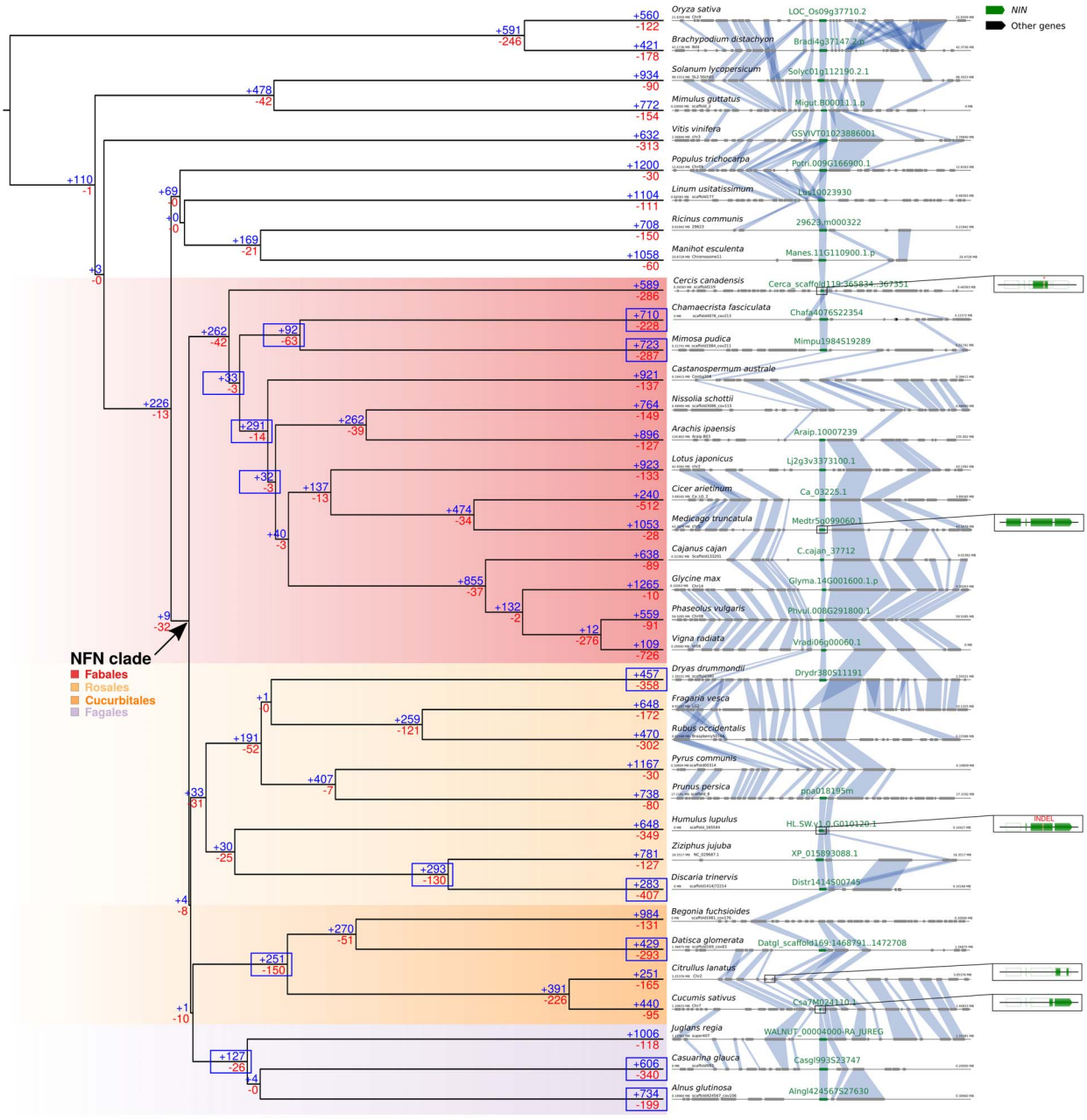


Fig. 2. Gene family expansions and contractions in the NFN clade. In the left panel, a species phylogeny of the used dataset is depicted. For each node, the numbers of gene families showing expansions (+, blue) and contractions (−, red) are given. Blue boxes point out nodes hypothesized to be positions of independent gain events of the NFN symbiosis, including all suggested alternative models in the given dataset (*7*). For example, among the Fabales, the dataset could comprise one, two, or three independent gains. A black arrow marks the base of the NFN clade. The right panel depicts syntenic relationships of the *NIN* region. *NIN* genes are colored in green. *NIN* gene IDs are shown above the gene symbol. The synteny analysis upholds orthologous relationships drawn from the phylogenetic analysis and supports the absence of *NIN* in several species by verifying the existence of contiguous *NIN* regions without *NIN* genes. Enlarged gene models are only shown for fragmented *NIN* genes in comparison to the full *M. truncatula NIN* gene. Blocks with no fill and green fill represent parts absent and present, respectively, when compared to Medicago *NIN*. Insertions and deletions (INDEL) and premature stop codons (*) are symbolized by a vertical red line.

all nonnodulating species (fig. S26). In the *M. truncatula rpg* mutant, infection threads are still present but their structure is abnormal (*41*), indicating that *RPG*, similar to *NIN*, is required for proper infection-thread progression. However, in contrast to *nin* mutants, nodules are formed on *rpg* mutant roots (*41*). Among the nodulating species, *RPG* is absent in the papilionoid *Arachis ipaensis* (Fig. 3). In the genus *Arachis*, infection threads were not observed; instead, rhizobia appear to infect nodules intercellularly (*42*). Polymorphism in *RPG* may represent an intermediate step on an evolutionary path toward the loss of this symbiosis in *Arachis*, a genus in which NFN symbiosis is described as a labile trait (*43*, *44*). Absence of *RPG* in *Arachis* also explains why the genome-wide comparative phylogenomic approach did not identify this gene, given that the pipeline required the candidate genes to be present in all nodulating species. *RPG* was one of the genes rejected for not fulfilling this criterion. Among nonnodulating species, *Juglans regia* (Fagales), *Ziziphus jujuba*, and *Prunus persica* (Rosales) have lost *RPG* but retained *NIN*, suggesting that additional mutations might be causative for the loss of NFN symbiosis in these species. Some of the candidate mutation targets, for example, lysine motif (LysM) receptors involved in the perception of symbiotic signals produced by nitrogen-fixing nodulating rhizobia, will be difficult to identify by phylogenomic analysis, because of the rapid evolution and expansionary dynamics of these gene families (*45*). In contrast to other symbiosis-relevant genes involved in infection, *NIN* and *RPG* are only known to have NFN symbiosis–specific functions, whereas the mutation of other genes may have more pleiotropic effects. For example, the key signaling components *SYMRK*, *CCaMK*, and *CYCLOPS* are involved in both NFN symbiosis and arbuscular mycorrhizal symbiosis, the most widespread symbiosis in land plants (*46*). Mutations in any of these three genes would also affect arbuscular mycorrhizal symbiosis. Illustrating this dual selection pressure, retention of these genes has been described in the genus *Lupinus*, which lost the arbuscular mycorrhizal symbiosis but retained NFN symbiosis (*37*, *47*, *48*). Given that both *NIN* and *RPG* are present in species outside the NFN clade (Fig. 3 and figs. S4 and S6), their consistent losses in nonnodulating species suggest the shift in constraints on sequence evolution specifically in the NFN clade, which might be mirrored by signature of relaxed or positive selection on both genes at the base of the NFN clade. We investigated the selective pressure acting on the NFN clade for these two genes using the PAML package (*49*). Results did not reveal a statistically significant positive or relaxed selection occurring in the NFN clade that would have reflected a putative neo functionalization for NFN symbiosis–specific functions (table S34).

Discussion

In recent decades, the favored model to explain the scattered occurrence of NFN symbiosis in flowering plants predicted a single predisposition event at the base of the NFN clade followed by up to 16 origins, even though the occurrence of multiple losses was never excluded (*7*, *11*).

Our genome-wide comparative analysis did not detect gene gains specific to the NFN clade and maintained in all nodulating species. Such genes would have been ideal candidates for either the predisposition event (in the multiple-gains hypothesis) or the evolution of NFN symbiosis itself in the hypothesis that NFN symbiosis evolved only once in the most recent common ancestor of the NFN clade (the single-gain hypothesis). This indicates that this step involved either fast-evolving genes that were not captured by our phylogenomics pipeline or subtler genetic changes. Evolutionary developmental genetics in plant, fungal, and animal systems have revealed that even more than gains of genes, new traits often arise from the rewiring of existing gene networks via gains or losses of cis regulatory elements leading to the co-option of ancestral genes (*50*). A similar mechanism may have acted in the most recent common ancestor of the NFN clade.

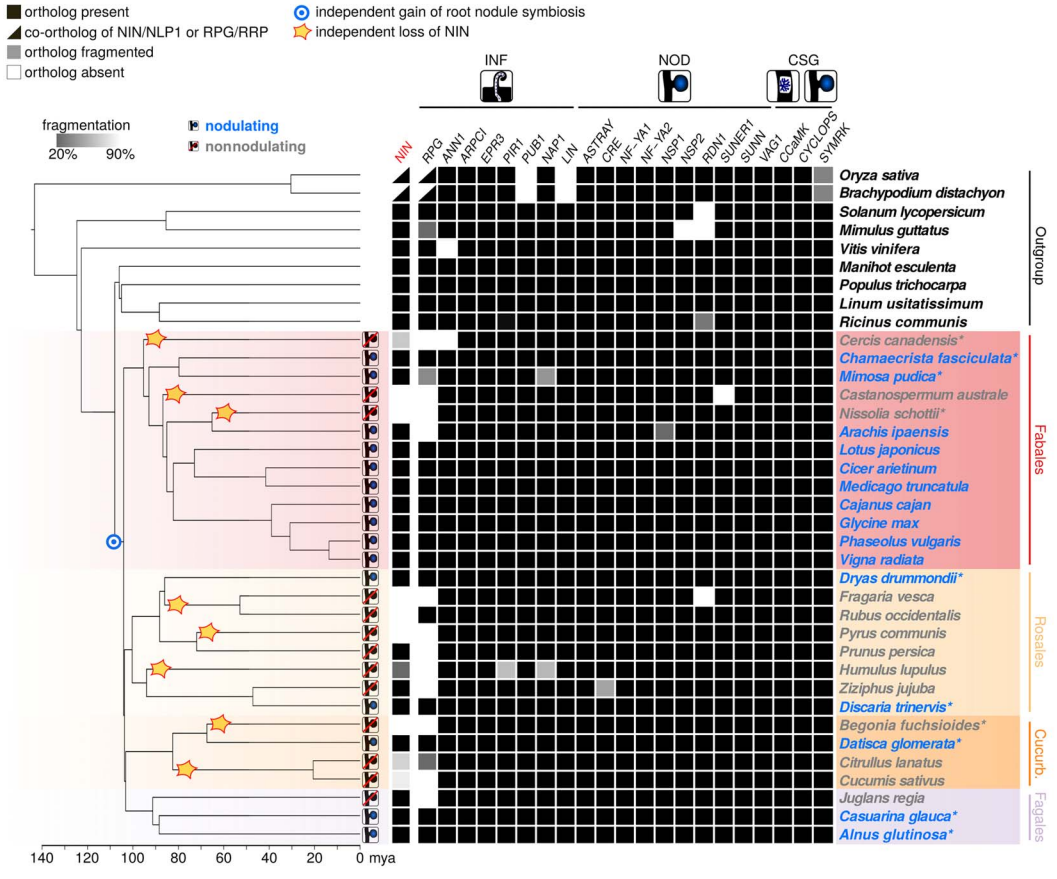


Fig. 3. Phylogenetic pattern of NFN symbiosis–related genes. The chronogram contains nodulating (blue) and nonnodulating (gray) species from all four orders of the NFN clade (blue circle with dot), to which NFN symbiosis is limited. Nine species outside the NFN clade are included as outgroups at the top. The absence or presence of entire or fragmented copies of 21 symbiosis genes are indicated by white, black, and gray boxes, respectively. Stars indicate independent losses of *NIN*. The independent loss or fragmentation of *NIN* correlates with the absence of nodules after the emergence of the NFN clade. *RPG* is lost or fragmented in even more nonnodulating species than *NIN* and also in the nodulating species *A. ipaensis* and *Mimosa pudica*. Asterisks indicate species sequenced for this study. INF, genes required for infection; NOD, genes involved in nodule organogenesis and regulation; CSG, genes required for both NFN symbiosis and arbuscular mycorrhiza symbiosis; mya, million years ago.

Co-option of different, or homologous, in the case of deep homology, genetic components may lead to the convergent evolutions of nonhomologous traits in multiple species (*50*). Deep homology has been invoked for the evolution of NFN symbiosis, during either the predisposition or the following putative multiple gains (*7*). Our results support this hypothesis, given that all the genes characterized for their involvement in NFN symbiosis in legumes were already present in the most recent common ancestor of the NFN clade and that we did not detect genes specific to the NFN clade that were conserved in all nodulating species (Figs. 2 and 3). For the putative multiple gains of NFN symbiosis, it cannot be excluded that gene gains were also involved in addition to co-option of ancestral pathways. We identified hundreds of such lineage-specific candidate genes (Fig. 2). However, considering that most of the predicted gains in our analysis are located in terminal taxa that are known to accumulate orphan genes and species-specific duplications in comparative approaches, it can be anticipated that only a subset of them participated in the evolution of NFN symbiosis (in the multiple-gains hypothesis) or in lineage-specific refinements of the trait (in the single-gain hypothesis).

Our results validate another hypothesis: multiple independent losses of NFN symbiosis in the four orders of the NFN clade. In a classical model of evolution, if the number of losses necessary to explain the distribution of a trait in a given clade outnumbers the predicted gains, multiple gains will be favored over multiples losses to explain the distribution of the trait. In legumes, up to six gains of NFN symbiosis were predicted (*7*), whereas the clear losses that we identified in *C. canadensis*, *C. australe*, and *N. schotii* now argue for a single origin before the radiation of the family (Fig. 3). Beyond legumes, the number of validated losses of NFN symbiosis is consistent with a single gain of this symbiosis in the most recent common ancestor of the NFN clade, even though it does not reject the possible occurrence of multiple gains. The recent identification of loss of NFN symbiosis in the Rosales *Trema orientalis* brings further support to this hypothesis (*51*). Besides NFN symbiosis, the single or multiple origin(s) of other traits, such as the evolution of complex multicellularity in fungi, are currently debated with the accumulating evidence of multiple losses demonstrated by the loss of associated essential genes (*52*). Thus, multiple gains and multiple losses are not mutually exclusive scenarios to explain the evolution of complex traits such as NFN symbiosis. This also suggests that reduction, similarly to the evolution of complexity, might be a major driver of the phenotypic diversity observed in extant organisms (*36*, *53*).

The fixation of loss-of-function alleles of *NIN* (either complete loss or pseudogenization) in nonnodulating species provides the genetic explanation for the loss of NFN symbiosis in 10 species representing eight nonnodulating lineages. Fixation of such alleles requires ecological conditions in which the cost of symbiotic nitrogen-fixation—involving infection, building nodules to host bacteria, and providing carbon to feed them—outweighs the benefit to the plant. In most terrestrial habitats, nitrogen is limiting (*54*), suggesting that the scale should be tipped toward the conservation of NFN symbiosis once this complex trait evolves. In nitrogen-rich habitats, NFN symbiosis is known to be inhibited in legumes (*55*). Long-term fertilizer application would make NFN symbiosis–specific genes superfluous, leading to their eventual mutational inactivation and loss. In addition to this abiotic constraint, NFN symbiosis may be undermined by "cheating" bacteria that gain entry into root nodules and are fed by the plant but do not deliver nitrogen (*56*–*58*). Cheaters may therefore imbalance the trade-off between the costs and benefits of the association, as already proposed on the basis of patterns of legume NFN symbiosis in Africa (*59*). This would result in the loss of NFN symbiosis being adaptive, thus providing an ecological explanation for the occurrence of this symbiosis in a few flowering plant species. In this context, the finding that *NIN* participates in shaping the root microbiome beyond NFN symbiosis makes it a target of adaptive selection against this symbiosis (*60*).

Engineering biological nitrogen fixation in crops remains a goal of plant synthetic biologists, with the aim of improving food production in developing countries in which the application of nitrogen fertilizer is limited by economic and infrastructural constraints. Our results supporting the occurrence of multiple independent losses indicates that the apparent selection against NFN symbiosis must be taken into account by projects whose aim is to improve legumes and, even more, when considering the engineering of nitrogen fixation in other crops.

Materials and methods

Plant material and sample preparation

The origin of the plant material used for DNA or RNA extraction in this study is summarized in data S2.

Methods for plant growth, DNA extractions, and RNA extractions are described in the supplementary materials online.

Genome sequencing

Figure S1 describes the overall strategy and results of the dataset production in this study.

Whole-genome sequencing for the 10 genomes was performed using Illumina sequencing technology (HISEq. 2000 and HISEq. 4000) at BGI-Shenzhen. Hierarchical library construction strategy was applied that typically included multiple paired-end libraries with insert sizes of 170, 250, 350, 500, and 800 bp and mate-pair libraries with insert sizes of 2, 5, 10, and 20 kb. Most of the paired-end and mate-pair libraries were prepared from large genomic fragments, typically of size 20 to 40 kb, or even larger. For some species, more small-insert-size PE libraries were constructed to complement the limited mate-pair libraries. The library construction for each species is summarized in table S36. Deep genome sequencing was performed for most of the species, with at least 110-fold coverage after a stringent data filtering and the highest sequencing depth reached 535-fold in cleaned data.

The overview statistics of data production are summarized in table S2 and fig. S1.

To minimize sequencing errors and reduce genome assembly artifacts, several quality control steps were taken to filter out low-quality sequencing reads:

Removal of N-rich reads: Reads that contained more than 10% of "N"s bases or polyA structure were removed.

Removal of low-quality reads: Reads in which 40% of the bases were low-quality (quality scores ≤7) were filtered out.

Filtering of reads with ≥10 nt aligned to the adapter sequences: Adaptor sequences were aligned to read1 and read2 using a dynamic programming approach; if the aligned fragments from read1 or read2 were reversed complementary to each other, the pair was also removed.

Filtering of small-insert-size reads with insert size (170 to 800 bp): The overlapping length between read1 and read2 is ≥10 bp; 10% mismatch was allowed.

Filtering of PCR duplicates: If the PE read1 and read2 were 100% identical, these reads were treated as duplicates and only one was retained.

Trimming of read ends: The low-quality bases from read ends (5′–5 bp, 3′–8 bp) were directly trimmed.

This filtering process was carried out using an in-house Perl program. After filtering, on average, 150-fold sequencing coverage was generated. For each species, clean reads were then passed to the genome assembler pipeline for de novo genome assembly.

The statistics of clean data are also summarized in table S2.

Genome assembly

To optimize the strategy for genome assembly, a genome survey is necessary to estimate the genome complexity. Some genomes are abundant in repetitive content and/or maintain a high heterozygosity rate through genome k-mer–analysis. A k-mer refers to a continuous sequence with k base pairs, typically extracted from the reads (thus, shorter than the read length, e.g., 17 bases per k-mer). If an "ideal" sequencing dataset is produced from a randomly whole-genome shotgun process without sequencing errors or coverage bias, the start positions of reads along the genome will follow Poisson distribution (*61*). Supposing that the read length is far shorter than the genome size, the k-mer can be regarded as randomly generated from the genome and their occurrence (sequencing depth) also is expected to be Poisson distributed (fig. S3A). Based on this assumption, the genome size can be estimated as (*62*):

$$\text{Genome size} = \frac{\text{k} - \text{mernumber}}{\text{Average sequencing depth}}$$

For a "normal" diploid genome, the k-mer frequency produced from adequate reads would

follow Poisson distribution. For genomes which are either repeat-rich or highly heterozygous, an additional peak either indicative of highly repetitive content (typically twofold depths of the main peak) or of high heterozygosity (*63*) (typically half depths of the main peak) is expected next to the "main peak" (indicative of the normal diploid genome) from the frequency distribution, or some even more complex scenarios either caused by the unexpected non-canonical genomic characteristics (e.g., degree of heterozygosity, complexity of the size and distribution of repetitive content, etc.) coupled with the use of different k-mer sizes.

K-mer statistics and distributions are presented in fig. S2 and tables S6 to S25.

During de novo genome assembly, we tried different k-mers (from 23- to 33-mer) to construct contigs, and the best k-mer (with the largest contig N50 length) was selected for the final run. Owing to differences in genome complexities between species, multiple genome assemblers were applied to achieve the optimal assembly result. As described in fig. S1, SOAPdenovo2 (version 2.04) (*64*) was the most frequently used assembler and Platanus (version 1.2.4) (*65*) for highly heterozygous genomes. After several rounds of assembly evaluations regarding contig contiguity and genome completeness, the best assemblies (largest contig N50 and highest BUSCO gene mapping rate) were selected for the downstream gap-closing step by Gapcloser (version 1.2) (*64*). For the assembly of *Discaria trinervis* genome, we used the Celera assembler CA8.3rc1. Because the Celera assembler (*66*) is sensitive to excessive coverage, library sizes were down-sampled to equal sized batches with a total coverage of approximately 50-fold. In a subsequent step, the entire sequence information was used to generate scaffolds and close gaps with SSpace (*67*) and Gapcloser, respectively.

Table S3 indicates the assembly strategy for each genome.

The statistics of genome assemblies are summarized in table S3, and the details for each species are presented in tables S6 to S25.

Genome assembly evaluation

The assembly evaluations for all of the genomes are provided in table S4.

Basically, mapping of the 1440 ultra-conserved core eukaryotic genes from the BUSCO (*68*) dataset resulted in >90% of the core eukaryote genes recovered for most of the genome assemblies. Taken together, these results indicate good genome assembly qualities for most of the newly sequenced species in this study, especially with respect to the genic regions (Fig. 1).

Genome annotation

A schematic workflow for genome annotation is given in fig. S1.

Repeat identification

Identification of transposable elements (TEs) was carried out by RepeatMasker (version 4-0-5) (*69*). A custom repeat library was constructed for each species by careful self-training. To construct the repeat custom library, we first collected the miniature inverted repeat transposable elements (MITEs) from many closely-related species and created a lineage-specific custom library by MITE-hunter (*70*) with default parameters. For the prediction of long terminal repeats (LTRs), we used LTRharvest (*71*) integrated in Genometools (version 1.5.8) (*72*), defining LTR in the length of 1.5 to 25 kb, with two terminal repeats ranging from 100 to 6000 bp with ≥99% similarity. Elements with intact PPT (poly purine tract) or PBS (primer binding site) were necessary to define LTR, which were identified by LTRdigest (*71*) using a eukaryotic tRNA library (http://gtrnadb.ucsc.edu/), whereas elements without appropriate PPT or PBS location were removed. To remove false positives such as local gene clusters and tandem local repeats, 50-bp flanking sequences on both sides of the LTRs of each candidate element were aligned using MUSCLE (*73*) with default parameters; if the identity was greater than or equal to 60%, the LTR element was considered as a false positive and removed. LTR elements nested with other inserted, but unrelated, components were also removed. Exemplars were built using a cutoff of 80% identity in 90% of element length from an all versus all BLASTn search. Terminal repeat retrotransposon in miniature (TRIM) libraries, with lengths of 70 to 500 kb, were built following a similar prediction strategy. Furthermore, the genomic sequence was masked to run RepeatModeler (version 1-0-8) (*69*) to extensively de novo predict repetitive sequences for each species. The MITE, LTR, and TRIM repetitive sequence libraries were integrated together to make a complete and nonredundant custom library. This custom repeat library was taken as the input for RepeatMasker to identify and classify transposable elements genome-wide for each species.

Gene annotation

Repeat elements were masked for each genome assembly before gene model prediction. Protein-coding genes were identified using the MAKER-P pipeline (version 2.31) (*74*) with two rounds of iterations. To obtain an optimal gene prediction, a series of trainings was performed. First, for genomes that have RNA samples sequenced, a set of transcripts was generated by a genome-guided approach using Trinity and then mapped back to the genome using PASA (version 2.0.2) (*75*). This process generated a set of complete gene models from each genome assembly and thus obtained real gene characteristics (size and number of exons and introns per gene, distribution of genes, features of splicing sites, etc.) by Augustus (*76*). Genemark-ES (version 4.21) (*77*) was self-trained with default parameters. SNAP (*78*) was trained using RNA- or protein-based gene models from the first iteration of MAKER-P pipeline. For RNA-seq aided gene annotation, RNA clean reads were assembled into inchworms using Trinity (*79*). For some species, transcriptome and ESTs data were obtained from NCBI or 1KP database if available (https://sites.google.com/a/ualberta.ca/onekp/). An optimal core protein set was collected from several closely related species for homolog-based gene prediction for each species, for example, gene models from the model plants like *Arabidopsis thaliana* and *Oryza sativa*, as well as from some well-annotated legumes like Medicago and *Glycine max*. Default parameters were used to run MAKER-P with all integrated annotation sources and to produce the final set of gene models for each species. The number of gene models for each species is summarized in table S26, and detailed statistics are summarized in table S28. BUSCO evaluation suggests complete and reliable gene annotation for all newly sequenced genomes (tables S4 and S5).

HMMER-based engineer InterproScan (version 5.11) (*80*) was used to predict gene function from several functional databases. The motifs and domains of genes were determined by searching against protein databases. An integrated gene functional annotation is summarized in table S29 for all species.

Transcriptome sequencing

RNA samples were sequenced for seven species to assist gene prediction in this study. The overview of total RNA samples with various tissues is summarized in table S37. For each RNA sample, a pair-end library with insert size of ~200 bp was constructed following the manufacturer protocol. Libraries were barcoded and pooled together as input to the Illumina Hiseq 4000 platform for sequencing. All of the RNA samples were sequenced in-depth, with an average of 6-Gb sequences per sample, to ensure a complete coverage for each transcriptome.

Genome-wide comparative phylogenomic analysis

Clustering of gene families

Clustering of gene families is based on predicted proteomes of the gene annotation of 37 species (table S1). To generate a set of nonredundant representative sequences, we removed multiple isoforms of a gene applying a cd-hit clustering using an identity threshold of 99.5% (*81*). Subsequently, homologs were identified with an all versus all blastp (v.2.2.30+) search of the 37 species preoteomes. For each query-subject-pair, we summed up the aligned sequence of all its HSPs (blast high scoring pair), ignoring overlaps, and compared it with the sequence length of both subject and query. We removed all query-subject-pairs from the blast tables for which the alignment coverage was less than 40% of either the total query or subject sequence length. According to Yang and Smith (*82*), this hit fraction filter step with the used cutoff of 40% substantially improves phylogenetic trees and orthology inference in the subsequent steps. On the basis of these modified blast tables, we clustered the remaining homologs to gene families with OrthoFinder (inflation parameter of 1.3) (*20*). The resulting 29,433 gene families were used as a starting point for the genome-wide phylogeny-based ortholog presence-absence analysis and candidate

confirmation. We also calculated gene family clusters without applying the 40% hit fraction filter and used these 23,869 gene family clusters for the analysis of gene copy-number variation. Figure S3 provides an overview both for the individual steps and the complete pipeline.

Genome-wide phylogeny-based ortholog presence-absence analysis

For each OrthoFinder gene family cluster, a separate phylogeny was calculated with FastME (*83*). As suggested by Yang and Smith (*82*), unreliable or wrongly resolved super-long branches were removed from each family tree in the following way: For each tree, average length and standard deviation of terminal branches were calculated, and branches longer than the mean of the average terminal branch length plus threefold standard deviation were removed. A python script was written to root pruned trees with farthest-oldest outgroup method [implemented in the Python package ETE 3 (*84*)], and ortholog-paralog relationships were inferred with the species overlap algorithm (implemented in the Python package ETE 3) (*84*). We then searched lists of orthologs of each gene of the reference species, the well-characterized nodulating legume Medicago, in total 29,213 ortholog lists, for the presence or absence of orthologs in all remaining 36 species of our dataset. The following criteria were applied to identify candidates:

1) orthologs present in at least 66% of the nodulating species (10 of 15),

2) orthologs present in a fraction of nodulating species from each order of the NFN clade (at least 7 of 10 nodulating Fabales, at least 1 of 2 nodulating Rosales, at least 1 of 2 nodulating Fagales, 1 of 1 nodulating Cucurbitales),

3) (only for the predisposition hypothesis) orthologs absent from ALL outgroup species outside of the NFN clade, and

4) (only for the multiple-losses hypothesis) orthologs absent from more than 50% of the nonnodulating species (at least 7 of 13).

Instead of filtering for the presence of orthologs of all nodulating species (10/10 Fabales, 2/2 Rosales, 2/2 Fagales, and 1/1 Cucurbitales), we used relatively relaxed criteria (as defined in 1 and 2) for nodulating species to avoid missed potential candidates owing to erroneous gene annotations (e.g., false negatives from gene annotation pipelines). Each of the candidates resulting from these criteria (predisposition hypothesis: 31 candidates; multiple-losses hypothesis: 121 candidates) underwent a refined candidate analysis described below with stricter criteria for nodulating species.

Besides the candidates of the genome-wide presence-absence analysis of phylogeny-based orthologs, we collected 22 candidate genes (table S33) that have been reported to be involved in root nodule symbiosis mostly from the model legume organisms Medicago and *L. japonicus*. For each OrthoFinder gene family cluster containing one of these genes, we calculated maximum likelihood gene family trees (RaxML v.8.2.4 Model: CATWAG). On the basis of the topology of the gene family tree and its protein alignment (MAFFT v.7.222 L-INS-I (*85*), trimming: BMGE gap-rate cut-off 80%) (*86*), we manually selected the subtree that contained the gene of interest (orthogroup). Subsequently we realigned (MAFFT v.7.222 L-INS-i, trimming: BMGE gap-rate cutoff 20%) and recalculated the phylogenetic tree (RaXML v.8.2.4 Model: GAMMAJTT, 200 bootstraps) (*87*) of the orthologous group to improve the quality of the subtree. Trees were rooted manually. Starting from the gene of interest and traversing to the root of the tree, we marked all nodes as duplication or speciation events. If the two subclades of a node shared genes that were originating from the same species, this node was interpreted as a duplication, otherwise as a speciation event. On the basis of speciation nodes, we inferred the orthologs of the gene of interest. At duplication nodes, all genes in the subclade lacking the query gene were inferred as paralogs. In the case of a speciation node, all genes belonging to both subclades of that particular node were inferred as orthologs of the query gene unless they were annotated as paralogs at a previous node. All orthologs with incomplete gene models were removed as long as they had paralogs among the species they were derived from, keeping at least one ortholog per species. We defined gene models as incomplete or fragmented if more than 20% of the conserved amino acid sequence was absent. As conserved amino acid sequences, we used the trimmed alignments (BMGE 20% gap-rate cutoff). To avoid false conclusions for missing orthologs, we retested a potential absence of genes by a homolog search. In such a case and in the case of still-incomplete gene models, we searched the complete genome sequence of the corresponding species (tblastn v2.2.30+, default parameters) with the closest homolog from the gene tree for regions containing potential gene loci of putative orthologs. These regions were then used to predict gene models (fgenesh+) (*88*). The resulting gene models were included in the set of sequences of the orthologous group, and another round of alignment and tree calculation was performed (same settings and tools as last round). These identified sequences complementing the species gene annotation are provided as a separate fasta-formatted file in data S1. The resulting tree was then used for a final round of ortholog inference so that the final set of orthologous genes is the result of an iterative process with constant improvement of gene models and phylogenies. If a complete gene model was not detected, the fragmented model was used and the ortholog was annotated as "fragmented" for the respective species. Fragmented models were only annotated as complete if the different fragments merged to a complete model and the fragmentation could be explained by the fragmentation of genomic scaffolds.

For the prefiltered candidates from the genome-wide phylogeny-based ortholog presence-absence analysis, we used stricter criteria than in the fully automated approach that led to the prefiltered candidates. We only kept such candidates from the genome-wide phylogeny-based automated presence-absence pipeline for which orthologs of all nodulating species were present and orthologs in more than 50% of nonnodulating species (7 of 13) were absent. Orthologs absent from the orthofinder output were independently searched by microsynteny to exclude the possibility that the ortholog could not be found because the syntenic region of the ortholog was not in the corresponding genome assembly (for more detail see the "synteny analysis" section). The presence, absence, and fragmentation of orthologs for each of the 22 selected known symbiosis genes and each species is summarized in Fig. 3. The diversification times shown in the chronogram are based on estimates from Bell *et al*. (Outgroup, BEAST, 36 minimum age constraints treated as log-normal distributions) (*89*), Xi *et al*. (Malpighiales, BEAST, uncorrelated lognormal model) (*90*, *91*), and Li *et al*. (NFN clade, r8s, 1008 taxatree) (*13*). Phylogenetic trees for all candidate genes are provided in figs. S4 and S6 to S25.

Analysis of gene copy-number variation

OrthoFinder gene family clusters were used to identify nodes in the species tree of our dataset, where gene family expansions or contractions in fast-evolving gene families occurred. The number of genes for each species of each cluster were counted and analyzed with the software tool CAFE (*92*). Following instructions given in the CAFE manual, we removed gene family clusters with strong outliers in gene copy number. Therefore, we excluded 193 gene family clusters from the analysis for which the difference between maximum copy number and median copy number was greater than or equal to 50 copies, which meets the elbow criterion to identify the optimal number of clusters in a clustering problem (fig. S27). To avoid overestimation of gene family contractions, we only used gene family clusters that contained orthologs from at least 28 species. This enabled us to analyze gene families that lost all genes in zero up to nine species. We chose this cut-off because of the nine outgroup species in the dataset. In the extreme case that all these nine outgroups have a gene count of zero, we could still analyze gene families originating from the last common ancestor of the NFN clade. After applying this cut-off, a total of 10,237 gene families were kept for the gene family evolution analysis (automatic λ and μ estimation, significance level for fast-evolving families 5%). The results of the analysis are shown in Fig. 2.

From the Medicago Gene Expression Atlas, we obtained all differentially expressed probesets that were either up-regulated with a fold change of 2 or down-regulated with a fold change of 0.5 in root nodule tissue of different age (7, 10, 14, and 28 dpi) compared to untreated root tissue. We associated all of these 18,131 regulated probesets to 17,521 Medicago v4.0 gene IDs using the mapping file provided by MtGEA. For each of the 14 hypothesized independent gain-of–NFN symbiosis nodes (Fig. 2, blue boxes), we extracted all gene family clusters that showed expansions at these nodes. For each of these clusters, we counted

the number of transcriptionally regulated and not regulated Medicago genes, ignoring all genes that could not be mapped to probesets. To test whether a gene family cluster was enriched for Medicago genes differentially expressed in nodulating versus mock control roots, we performed Fisher's exact test on a significance level of 5%.

Synteny analysis

Genome-wide syntenic and collinear blocks were identified across the 37 selected genomes in this study. First, all versus all Blastp (E-value ≤ $1e^{-10}$) was performed on the translated protein sequences of the 37 sets of annotated gene models, resulting in a database of protein similarity. We then used the Multiple Collinearity Scan (Mcscan toolkit version 1.1, 2016, ≥5 homologous gene pairs/block) to identify conserved collinear blocks between the 37 genomes, creating a syntenic or collinear block database across all of the 37 species. To find all of the homologous syntenic blocks of interest, we first used Medicago genome as a reference, searching and locating the target genes along the syntenic blocks with the flanking genes surrounding up– or down–100 kb genomic regions as well as the counterparts from different genomes. For any other given genome, the optimal collinear block (the highest score if multiple duplicated blocks were found) was defined according to conservation of gene content (the largest number of orthologous gene pairs) and consistency of gene order. If a corresponding ortholog was present in the collinear block from other aligned genomes, this was called scenario-1 (indicating that synteny supports gene presence and consistent with the genome-wide gene family ortholog prediction); if the target orthologous gene was absent from the collinear blocks of the given genome, this was called scenario-2 (synteny supports gene absence). After this first round was finished, to complete any missing genes or blocks owing to weak alignment signals, we classified these identified orthologous genes as well as the corresponding collinear blocks and repeated the searching and locating process between genomes according to their evolutionary proximity in phylogenetic position (using the most closely related genome as query). By this process, we updated scenario-1 and scenario-2 as described above. In addition, for some species, no collinear block was identified around the possible target gene, and we manually revisited the protein similarity database and searched the genomic regions flanking the Medicago gene to confirm the gene presence or absence supported by synteny. Finally, if no synteny was identified, but the candidate ortholog gene was predicted from the genome-wide gene family analysis, we called this as scenario-3 (gene presence without synteny support, which indicates possible gene translocation and synteny erosion).

Detecting selection pressure on gene trees

For *NIN* and *RPG*, protein sequences were aligned using MAFFT v7.380 (*85*). The protein alignment served as matrix for codon alignment performed using the Perl script pal2nal v14 with the –nogap option enabled to remove all gapped positions. Codon alignments were then subjected to a maximum likelihood analysis using IQ-TREE v1.6.1 (*93*) with 10,000 Ultrafast bootstraps replicates (*94*). The best-fitted evolutionary model was previously investigated using ModelFinder (*95*). The unrooted tree obtained was controlled to fit with the evolutionary frame of species, and the NFN clade labeled as the foreground branch that was tested for being under positive selection. The latter was investigated using the branch-site model A implemented in the codeml module from the PAML package v4.9 g (*49*). An alternative hypothesis (NFN clade may have proportion of sites under positive selection) was compared to the null hypothesis (NFN clade may have different proportion of sites under neutral selection compared to the other clades). For the null model, the parameters were set as follows: "model = 2, NSites = 2, fix_kappa = 0, fix_omega = 1 and omega = 1," whereas the parameters for the alternative model were: "model = 2, NSites = 2, fix_kappa = 0, fix_omega = 0 and omega = 1.5". The two hypotheses were compared using the likelihood ratio test based on a χ^2 distribution with one degree of freedom. If the alternative hypothesis was validated, codon sites likely to fall under positive selection were identified using the Bayes Empirical Bayes procedure (*49*).

PCR validation of the absence of NIN in nonnodulating legumes

A nested PCR approach was undertaken using primers designed on an alignment of *NIN* gDNA from Medicago and *M. pudica*. These primers are designed to amplify ~120 bp on Exon 4, ~170 bp on Exon 5, and the intron in between. The size of this intron ranges from 133 bp in Medicago to 397 bp in *M. pudica*. For the first PCR, we used the degenerated primer pair NIN-Fwd-3 5′-GGAGAAAGTCMGGCGASAA and NIN-Rev-3 5′-GRAARCTGGCATAGAATGA. The Nested PCR was run with 0.2 µl of the PCR reaction and primers NIN-Fwd-2 5′-CGAACCAAGGCTGAGAAGAC and NIN-Rev-2 5′-ATCTGTATGGCACCCTCTGC. The first PCR run was 94°C 30 s, 45°C 1 min, 72°C 1 min for 35 cycles, and the second PCR run was 94°C 30 s, 50°C 1 min, 72°C 1 min for 35 cycles. For all species, the PCR was run on >2 samples. In addition, a PCR on the 28*S* was run to confirm the quality of the samples. All PCRs were run using a GoTaq DNA polymerase.

REFERENCES AND NOTES

1. P. C. Dos Santos, Z. Fang, S. W. Mason, J. C. Setubal, R. Dixon, Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* **13**, 162 (2012). doi: 10.1186/1471-2164-13-162; pmid: 22554235
2. P. H. Beatty, A. G. Good, Future prospects for cereals that fix nitrogen. *Science* **333**, 416–417 (2011). doi: 10.1126/science.1209467; pmid: 21778391
3. L. Hiltner, Über die Bedeutung der Wurzelknöllchen von *Alnus glutinosa* für die Stickstoffernährung dieser Pflanze. *Landw. Versuchsstat* **46**, 153–161 (1895).
4. H. Hellriegel, H. Wilfarth, *Untersuchungen über die Stickstoffnahrung der Gramineen und Leguminosen* (Buchdruckerei der Post Kayssler, Berlin, 1888).
5. D. E. Soltis *et al*., Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2647–2651 (1995). doi: 10.1073/pnas.92.7.2647; pmid: 7708699
6. C. Kistner, M. Parniske, Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci.* **7**, 511–518 (2002). doi: 10.1016/S1360-1385(02)02356-7; pmid: 12417152
7. J. J. Doyle, Phylogenetic perspectives on the origins of nodulation. *Mol. Plant Microbe Interact.* **24**, 1289–1295 (2011). doi: 10.1094/MPMI-05-11-0114; pmid: 21995796
8. J. I. Sprent, J. Ardley, E. K. James, Biogeography of nodulated legumes and their nitrogen-fixing symbionts. *New Phytol.* **215**, 40–56 (2017). doi: 10.1111/nph.14474; pmid: 28211601
9. K. Pawlowski, K. N. Demchenko, The diversity of actinorhizal symbiosis. *Protoplasma* **249**, 967–979 (2012). doi: 10.1007/s00709-012-0388-4; pmid: 22398987
10. S. M. Swensen, The evolution of actinorhizal symbioses: Evidence for multiple origins of the symbiotic association. *Am. J. Bot.* **83**, 1503–1512 (1996). doi: 10.1002/j.1537-2197.1996.tb13943.x
11. G. D. A. Werner, W. K. Cornwell, J. I. Sprent, J. Kattge, E. T. Kiers, A single evolutionary innovation drives the deep evolution of symbiotic N_2-fixation in angiosperms. *Nat. Commun.* **5**, 4087 (2014). doi: 10.1038/ncomms5087; pmid: 24912610
12. S. M. Swensen, D. R. Benson, in *Nitrogen-fixing Actinorhizal Symbioses*, K. Pawlowski, W. E. Newton, Eds. (Springer Netherlands, 2008), pp. 73–104.
13. H. L. Li *et al*., Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Sci. Rep.* **5**, 14023 (2015). doi: 10.1038/srep14023; pmid: 26354898
14. J. A. Eisen, C. M. Fraser, Phylogenomics: Intersection of evolution and genomics. *Science* **300**, 1706–1707 (2003). doi: 10.1126/science.1086292; pmid: 12805538
15. P. M. Delaux, G. Radhakrishnan, G. Oldroyd, Tracing the evolutionary path to nitrogen-fixing crops. *Curr. Opin. Plant Biol.* **26**, 95–99 (2015). doi: 10.1016/j.pbi.2015.06.003; pmid: 26123396
16. Materials and methods are available as supplementary materials.
17. N. D. Young *et al*., The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**, 520–524 (2011). doi: 10.1038/nature10625; pmid: 22089132
18. J. Schmutz *et al*., A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **46**, 707–713 (2014). doi: 10.1038/ng.3008; pmid: 24908249
19. J. Schmutz *et al*., Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183 (2010). doi: 10.1038/nature08670; pmid: 20075913
20. D. M. Emms, S. Kelly, OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015). doi: 10.1186/s13059-015-0721-2; pmid: 26243257
21. D. L. Stern, The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764 (2013). doi: 10.1038/nrg3483; pmid: 24105273
22. N. Shubin, C. Tabin, S. Carroll, Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823 (2009). doi: 10.1038/nature07891; pmid: 19212399
23. J. J. Doyle, Chasing unicorns: Nodulation origins and the paradox of novelty. *Am. J. Bot.* **103**, 1865–1868 (2016). doi: 10.3732/ajb.1600260; pmid: 27756731
24. J. F. Marsh *et al*., *Medicago truncatula NIN* is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiol.* **144**, 324–335 (2007). doi: 10.1104/pp.106.093021; pmid: 17369436
25. F. Clavijo *et al*., The *Casuarina NIN* gene is transcriptionally activated throughout *Frankia* root infection as well as in response to bacterial diffusible signals. *New Phytol.* **208**, 887–903 (2015). doi: 10.1111/nph.13506; pmid: 26096779
26. H. Gherbi *et al*., SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and *Frankia*bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4928–4932 (2008). doi: 10.1073/pnas.0710618105; pmid: 18316735
27. L. Schauser, A. Roussis, J. Stiller, J. Stougaard, A plant regulator controlling development of symbiotic root nodules. *Nature* **402**, 191–195 (1999). doi: 10.1038/46058; pmid: 10647012
28. A. van Zeijl *et al*., CRISPR/Cas9-mediated mutagenesis of four putative symbiosis genes of the tropical tree *Parasponia*

andersonii reveals novel phenotypes. *Front. Plant Sci.* **9**, 284 (2018). doi: 10.3389/fpls.2018.00284; pmid: 29559988
29. S. Svistoonoff *et al*., The independent acquisition of plant root nitrogen-fixing symbiosis in Fabids recruited the same genetic pathway for nodule organogenesis. *PLOS ONE* **8**, e64515 (2013). doi: 10.1371/journal.pone.0064515; pmid: 23741336
30. K. Markmann, G. Giczey, M. Parniske, Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. *PLOS Biol.* **6**, e68 (2008). doi: 10.1371/journal.pbio.0060068; pmid: 18318603
31. L. G. Nagy *et al*., Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nat. Commun.* **5**, 4471 (2014). doi: 10.1038/ncomms5471; pmid: 25034666
32. C. E. Conn *et al*., Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* **349**, 540–543 (2015). doi: 10.1126/science.aab1140; pmid: 26228149
33. J. He *et al*., The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics* **10**, 441 (2009). doi: 10.1186/1471-2105-10-441; pmid: 20028527
34. G. Criscuolo, V. T. Valkov, A. Parlati, L. M. Alves, M. Chiurazzi, Molecular characterization of the *Lotus japonicus* NRT1(PTR) and NRT2 families. *Plant Cell Environ.* **35**, 1567–1581 (2012). doi: 10.1111/j.1365-3040.2012.02510.x; pmid: 22458810
35. L. Aravind, H. Watanabe, D. J. Lipman, E. V. Koonin, Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11319–11324 (2000). doi: 10.1073/pnas.200346997; pmid: 11016957
36. R. Albalat, C. Cañestro, Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016). doi: 10.1038/nrg.2016.39; pmid: 27087500
37. P. M. Delaux *et al*., Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLOS Genet.* **10**, e1004487 (2014). doi: 10.1371/journal.pgen.1004487; pmid: 25032823
38. A. Y. Borisov *et al*., The *Sym35* gene required for root nodule development in pea is an ortholog of *Nin* from *Lotus japonicus*. *Plant Physiol.* **131**, 1009–1017 (2003). doi: 10.1104/pp.102.016071; pmid: 12644653
39. N. Azani *et al*., A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny – The Legume Phylogeny Working Group (LPWG). *Taxon* **66**, 44–77 (2017). doi: 10.12705/661.3
40. J. I. Sprent, Evolving ideas of legume evolution and diversity: A taxonomic perspective on the occurrence of nodulation. *New Phytol.* **174**, 11–25 (2007). doi: 10.1111/j.1469-8137.2007.02015.x; pmid: 17335493
41. J.-F. Arrighi *et al*., The *RPG* gene of *Medicago truncatula* controls *Rhizobium*-directed polar growth during infection. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9817–9822 (2008). doi: 10.1073/pnas.0710273105; pmid: 18621693
42. M. R. Chandler, Some observations on infection of *Arachis hypogaea* L. by *Rhizobium*. *J. Exp. Bot.* **29**, 749–755 (1978). doi: 10.1093/jxb/29.3.749
43. Z. Peng *et al*., Transcriptome profiles reveal gene regulation of peanut (*Arachis hypogaea* L.) nodulation. *Sci. Rep.* **7**, 40066 (2017). pmid: 28059169
44. D. W. Gorbet, J. C. Burton, A non-nodulating peanut. *Crop Sci.* **19**, 727–728 (1979). doi: 10.2135/cropsci1979.0011183X001900050045x
45. S. De Mita, A. Streng, T. Bisseling, R. Geurts, Evolution of a symbiotic receptor through gene duplications in the legume-rhizobium mutualism. *New Phytol.* **201**, 961–972 (2014). doi: 10.1111/nph.12549; pmid: 24400903
46. M. Parniske, Arbuscular mycorrhiza: The mother of plant root endosymbioses. *Nat. Rev. Microbiol.* **6**, 763–775 (2008). doi: 10.1038/nrmicro1987; pmid: 18794914
47. P. Favre *et al*., A novel bioinformatics pipeline to discover genes related to arbuscular mycorrhizal symbiosis based on their evolutionary conservation pattern among higher plants. *BMC Plant Biol.* **14**, 333 (2014). doi: 10.1186/s12870-014-0333-0; pmid: 25465219
48. A. Bravo, T. York, N. Pumplin, L. A. Mueller, M. J. Harrison, Genes conserved for arbuscular mycorrhizal symbiosis identified through phylogenomics. *Nat. Plants* **2**, 15208 (2016). doi: 10.1038/nplants.2015.208; pmid: 27249190
49. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007). doi: 10.1093/molbev/msm088; pmid: 17483113
50. S. B. Carroll, Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008). doi: 10.1016/j.cell.2008.06.030; pmid: 18614008
51. R. van Velzen *et al*., Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4700–E4709 (2018). pmid: 29717040
52. L. G. Nagy, Evolution: Complex multicellular life with 5,500 genes. *Curr. Biol.* **27**, R609–R612 (2017). doi: 10.1016/j.cub.2017.04.032; pmid: 28633034
53. M. A. O'Malley, J. G. Wideman, I. Ruiz-Trillo, Losing complexity: The role of simplification in macroevolution. *Trends Ecol. Evol.* **31**, 608–621 (2016). doi: 10.1016/j.tree.2016.04.004; pmid: 27212432
54. D. S. LeBauer, K. K. Treseder, Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology* **89**, 371–379 (2008). doi: 10.1890/06-2057.1; pmid: 18409427
55. J. Streeter, P. P. Wong, Inhibition of legume nodule formation and N_2-fixation by nitrate. *Crit. Rev. Plant Sci.* **7**, 1–23 (1988). doi: 10.1080/07352688809382257
56. E. T. Kiers, R. A. Rousseau, S. A. West, R. F. Denison, Host sanctions and the legume-rhizobium mutualism. *Nature* **425**, 78–81 (2003). doi: 10.1038/nature01931; pmid: 12955144
57. H. Fujita, S. Aoki, M. Kawaguchi, Evolutionary dynamics of nitrogen fixation in the legume-rhizobia symbiosis. *PLOS ONE* **9**, e93670 (2014). doi: 10.1371/journal.pone.0093670; pmid: 24691447
58. L. Carro, P. Pujic, M. E. Trujillo, P. Normand, *Micromonospora* is a normal occupant of actinorhizal nodules. *J. Biosci.* **38**, 685–693 (2013). doi: 10.1007/s12038-013-9359-y;pmid: 24287646
59. J. I. Sprent, *Legume Nodulation: A Global Perspective* (Wiley, 2009).
60. R. Zgadzaj *et al*., Root nodule symbiosis in *Lotus japonicus* drives the establishment of distinctive rhizosphere, root, and nodule bacterial communities. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E7996–E8005 (2016). doi: 10.1073/pnas.1616564113; pmid: 27864511
61. G. Marçais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011). doi: 10.1093/bioinformatics/btr011; pmid: 21217122
62. E. Veeckman, T. Ruttink, K. Vandepoele, Are we there yet? Reliably estimating the completeness of plant genome sequences. *Plant Cell* **28**, 1759–1768 (2016). doi: 10.1105/tpc.16.00349; pmid: 27512012
63. R. Arratia, D. Martin, G. Reinert, M. S. Waterman, Poisson process approximation for sequence repeats, and sequencing by hybridization. *J. Comput. Biol.* **3**, 425–463 (1996). doi: 10.1089/cmb.1996.3.425; pmid: 8891959
64. R. Luo *et al*., SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012). doi: 10.1186/2047-217X-1-18; pmid: 23587118
65. R. Kajitani *et al*., Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014). doi: 10.1101/gr.170720.113; pmid: 24755901
66. E. W. Myers *et al*., A whole-genome assembly of *Drosophila*. *Science* **287**, 2196–2204 (2000). doi: 10.1126/science.287.5461.2196; pmid: 10731133
67. M. Boetzer, C. V. Henkel, H. J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011). doi: 10.1093/bioinformatics/btq683; pmid: 21149342
68. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015). doi: 10.1093/bioinformatics/btv351; pmid: 26059717
69. A. Smit, R. Hubley, P. Green, RepeatMasker Open–4.0, 2013–2015; www.repeatmasker.org.
70. Y. Han, S. R. Wessler, MITE-Hunter: A program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, e199 (2010). doi: 10.1093/nar/gkq862; pmid: 20880995
71. D. Ellinghaus, S. Kurtz, U. Willhoeft, LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008). doi: 10.1186/1471-2105-9-18; pmid: 18194517
72. G. Gremme, S. Steinbiss, S. Kurtz, GenomeTools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **10**, 645–656 (2013). doi: 10.1109/TCBB.2013.68; pmid: 24091398
73. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004). doi: 10.1093/nar/gkh340; pmid: 15034147
74. M. S. Campbell *et al*., MAKER-P: A tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* **164**, 513–524 (2014). doi: 10.1104/pp.113.230144; pmid: 24306534
75. B. J. Haas *et al*., Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003). doi: 10.1093/nar/gkg770; pmid: 14500829
76. M. Stanke, R. Steinkamp, S. Waack, B. Morgenstern, AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309–W312 (2004). doi: 10.1093/nar/gkh379; pmid: 15215400
77. A. V. Lukashin, M. Borodovsky, GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Res.* **26**, 1107–1115 (1998). doi: 10.1093/nar/26.4.1107; pmid: 9461475
78. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004). doi: 10.1186/1471-2105-5-59; pmid: 15144565
79. M. G. Grabherr *et al*., Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011). doi: 10.1038/nbt.1883; pmid: 21572440
80. E. M. Zdobnov, R. Apweiler, InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001). doi: 10.1093/bioinformatics/17.9.847; pmid: 11590104
81. W. Li, A. Godzik, Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006). doi: 10.1093/bioinformatics/btl158; pmid: 16731699
82. Y. Yang, S. A. Smith, Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Mol. Biol. Evol.* **31**, 3081–3092 (2014). doi: 10.1093/molbev/msu245; pmid: 25158799
83. V. Lefort, R. Desper, O. Gascuel, FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* **32**, 2798–2800 (2015). doi: 10.1093/molbev/msv150; pmid: 26130081
84. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016). doi: 10.1093/molbev/msw046; pmid: 26921390
85. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013). doi: 10.1093/molbev/mst010; pmid: 23329690
86. G. Tan *et al*., Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst. Biol.* **64**, 778–791 (2015). doi: 10.1093/sysbio/syv033; pmid: 26031838
87. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). doi: 10.1093/bioinformatics/btu033; pmid: 24451623
88. V. Solovyev, in *Handbook of Statistical Genetics*, D. J. Balding, M. Bishop, C. Cannings, Eds. (Wiley, 2008), pp. 97–159.
89. C. D. Bell, D. E. Soltis, P. S. Soltis, The age and diversification of the angiosperms re-revisited. *Am. J. Bot.* **97**, 1296–1303 (2010). doi: 10.3732/ajb.0900346; pmid: 21616882
90. Z. Xi *et al*., Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17519–17524 (2012). doi: 10.1073/pnas.1205818109; pmid: 23045684
91. M. T. J. Johnson *et al*., Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLOS ONE* **7**, e50226 (2012). doi: 10.1371/journal.pone.0050226; pmid: 23185583
92. M. V. Han, G. W. C. Thomas, J. Lugo-Martinez, M. W. Hahn, Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013). doi: 10.1093/molbev/mst100; pmid: 23709260
93. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015). doi: 10.1093/molbev/msu300; pmid: 25371430
94. D. T. Hoang *et al*., MPBoot: Fast phylogenetic maximum parsimony tree inference and bootstrap approximation. *BMC Evol. Biol.* **18**, 11 (2018). doi: 10.1186/s12862-018-1131-3; pmid: 29390973

95. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017). doi: 10.1038/nmeth.4285; pmid: 28481363
96. M. Griesmann, Y. Chang, X. Liu, Y. Song, G. Haberer, M. B. Crook, B. Billault-Penneteau, D. Lauressergues, L. Imanishi, Y. P. Roswanjaya, W. Kohlen, P. Pujic, K. Battenberg, N. Alloisio, W. Sun, H. Hilhorst, M. G. Salgado, V. Hocher, H. Gherbi, S. Svistoonoff, J. J. Doyle, S. He, Y. Xu, W. Xian, Y. Fu, P. Normand, A. M. Berry, L. G. Wall, J. M. Ané, K. Pawlowski, X. Xu, H. Yang, M. Spannagl, K. F. X. Mayer, G. K. Wong, M. Parniske, P. M. Delaux, S. Cheng, Data supporting multiple independent losses of nitrogen-fixing root nodule symbiosis. GigaDB (2018); doi: 10.5524/100300

ACKNOWLEDGMENTS

The authors thank J. Sprent and G. Werner for initial discussions; R. Hagen and S. Singer for *C. fasciculata* leaves; S. Cannon for *C. canadensis* leaves; N. Séjalon-Delmas (Toulouse Botanical Garden), the International Center for Tropical Agriculture (CIAT), and USDA ARS-GRIN for seeds; and D. Lang for bioinformatics advice. **Funding:** This work was partly performed in the LRSV, which belongs to the TULIP LABEX (ANR-10-LABX-41), and was supported by grants from the Shenzhen Municipal Government of China to S.C. (JCYJ20150529150409546) and Y.C. (JCYJ20150831201643396); a grant from the Swedish Research Council Vetenskapsrådet (VR 2012-03061) to K.P.; the German Ministry of Education and Research grant 031A536 "de.NBI" and the Deutsche Forschungsgemeinschaft Project SFB924 to K.F.X.M. and M.P.; grants from the National Science Foundation (NSF 1237936 and 1546742) to J.-M.A.; grants from the French National Research Agency (ANR-12-BSV7-0007-02), the USDA (NIFA 2015-67014-22849), and ECOS-SUD (A13B03) to V.H., H.G., and S.S.; a grant from the National University of Quilmes, Argentina (PUNQ 1411/15) to L.G.W.; and by USDA grant NIFA590 CA-D-PLS-2173-H to A.M.B. W.K. is supported by NWO-VENI (863.15.010), and L.G.W. and L.I. are members of CONICET (Argentina). G.K.-S.W. is funded by Alberta Innovates Technology Futures (AITF) Innovates Centres of Research Excellence (iCORE). M.P. acknowledges funding by ERC Advanced grant 340904 (EvolvingNodules), which supported the work of M.G. and B.B.-P. and the sabbatical stay of K.P. at the LMU Munich. **Author contributions:** K.P., M.P., P.-M.D., and J.-M.A. defined the species to sequence and designed the project. K.P. coordinated the work on actinorhizal species, supervised M.G.S., isolated DNA for *Datisca glomerata* and *Begonia fuchsioides*, and extracted RNA of *B. fuchsioides*. M.G.S. prepared RNA of *D. glomerata*. P.-M.D. coordinated the consortium, provided DNA and RNA samples for *N. schottii*, conducted phylogenetic analyses of target genes, and supervised D.L. and J.K. D.L. extracted DNA for and conducted the experiment presented in fig. S5. J.K. conducted the PAML analysis on *NIN* and *RPG*. J.-M.A. supervised experiments on most legumes. M.B.C. extracted DNA and RNA of *C. fasciculata*, *M. pudica*, and *C. canadensis*. W.K. and Y.P.R. extracted DNA samples for *C. australe*, *Sophora flavescens*, and the *Chamaecrista* species used in fig. S5. L.G.W. collected *D. trinervis* plant material and supervised L.I. L.I. prepared DNA and RNA samples for *D. trinervis*. P.N. supervised experiments on *Alnus glutinosa*. N.A. and P.P. collected *A. glutinosa* material, designed extraction protocols, and extracted RNA (N.A.) and DNA (P.P.). H.G., V.H., and S.S. collected *C. glauca* material, supervised DNA extraction, and extracted RNA for *C. glauca*. K.F.X. M. and M.S. coordinated genome assembly and annotation. M.S. and G.H. supervised M.G. G.H. performed genome assembly, gene annotation, phylogenetic analyses, and prepared Fig. 1. M.P. initiated the consortium and supervised M.G. and B.B.-P. and the work on *Dryas drummondii*. M.G. conducted most of the phylogenetic analyses and prepared Figs. 2 and 3 and related supplementary texts and files. B.B.-P. collected material and extracted DNA and RNA from *D. drummondii*. S.C. and G.K.-S.W. designed the sequencing strategy and coordinated and supervised the bioinformatics and genomic analysis, including genome sequencing, assembly, and annotation at BGI-Shenzhen with H.Y., X.L., Q.G., X.F., and X.X. X.L. performed genome analyses. Y.C. performed samples management and data quality control and developed custom bioinformatics pipelines for genome assembly, annotation, and evaluation with S.X. S.X. managed the super computer clusters to perform assembly and annotation. Y.S. identified orthologs for Fig. 3 and performed most of the microsynteny analyses. Y.L. ran RNA data quality control, assembly, alignment, and annotation. S.H. finalized the synteny analysis and prepared Fig. 3. Y.X., J.Q., and Y.F. performed genome assembly, repeat, and gene annotation, respectively, for *M. pudica*, *A. glutinosa*, and *C. canadensis* (Y.X.); *N. schottii* and *D. trinervis* (J.G.); and *C. glauca* and *C. fasciculata* (Y.F.). H.H. provided access to the unpublished genome of *C. australe*. J.J.D. provided comments and suggestions on the possible evolutionary scenarios and contributed to writing the manuscript. A.M.B. and K.B. provided material and comments on the experimental strategy. S.C. wrote the supplementary texts with contributed sections from all co-authors. M.P. and P.-M.D. wrote the manuscript with inputs and comments from co-authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All raw reads have been deposited in the Sequence Read Archive (SRA) under project PRJNA374770; accession codes are summarized in the supplementary materials. All files have been deposited in GigaDB (*96*). All material used for genome sequencing is available on request to the responsible author as indicated in data S2. All materials used for genome sequencing are available on request to P.N. for *A. glutinosa*; K.P. for *D. glomerata* and *B. fuchioides*; S.S., H.G., or V.H. for *C. glauca*; J.-M.A. or M.B.C. for *C. canadensis*, *C. fasciculata*, and *M. pudica*; L.G.W. for *D. trivernis*; M.P. for *D. drummondii*; and P.-M.D. for *N. schottii* (also listed in data S2).

SUPPLEMENTARY MATERIALS

12 February 2018; accepted 16 May 2018
Published online 24 May 2018
10.1126/science.aat1743

RESEARCH ARTICLE SUMMARY

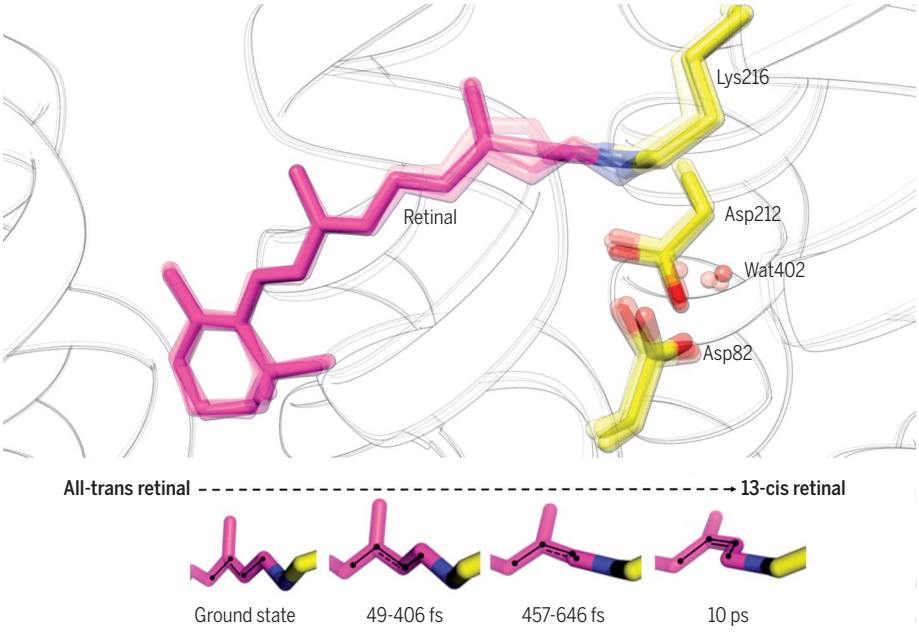
STRUCTURAL DYNAMICS

Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser

Przemyslaw Nogly, Tobias Weinert, Daniel James, Sergio Carbajo, Dmitry Ozerov, Antonia Furrer, Dardan Gashi, Veniamin Borin, Petr Skopintsev, Kathrin Jaeger, Karol Nass, Petra Båth, Robert Bosman, Jason Koglin, Matthew Seaberg, Thomas Lane, Demet Kekilli, Steffen Brünle, Tomoyuki Tanaka, Wenting Wu, Christopher Milne, Thomas White, Anton Barty, Uwe Weierstall, Valerie Panneels, Eriko Nango, So Iwata, Mark Hunter, Igor Schapiro, Gebhard Schertler, Richard Neutze, Jörg Standfuss*

INTRODUCTION : Retinal is a light-sensitive protein ligand that is used by all domains of life to process the information and energy content of light. Retinal-binding proteins are integral membrane proteins that drive vital biological processes, including light sensing for spatial orientation and circadian clock adjustment, as well as maintaining electrochemical gradients through ion transport. They also form the basis for optogenetic manipulation of neural cells. How the protein environment guides retinal isomerization on a subpicosecond time scale toward a single high-yield product is a fundamental outstanding question in photobiology.

RATIONALE: Light-induced isomerization of retinal is among the fastest reactions known in biology. It has been widely studied by spectroscopic techniques to probe the evolution of spectral intermediates over time. Using x-ray free-electron lasers (XFELs), it is now possible to observe ultrafast photochemical reactions and their induced molecular motions within proteins on scales of femtoseconds to milliseconds with near-atomic structural resolution. In this work, we used XFEL radiation to study the structural dynamics of retinal isomerization in the light-driven proton-pump bacteriorhodopsin (bR). The principal mechanism of isomerization in this prototypical retinal-binding protein has direct relevance for all other members of this important family of membrane proteins, and it provides insight into how protein environments catalyze photochemical reactions in general.



Time-resolved serial crystallography resolves ultrafast atomic motions of retinal and the surrounding protein following photoexcitation. Retinal evolves from an all-trans conformation in the ground state toward a twisted 13-cis retinal over the course of a few hundred femtoseconds. The complex counterion, formed by two aspartic acid residues (Asp) and a water molecule (Wat), responds to changes in the electronic structure of the chromophore on the same time scale as the formation of the excited state.

RESULTS: We collected high-resolution x-ray diffraction data from bR microcrystals injected across the femtosecond x-ray pulses of the Linac Coherent Light Source after excitation of the retinal chromophore by an optical laser pulse. X-ray diffraction images were sorted into temporal subgroups with a precision of about 200 fs. A series of 18 overlapping difference Fourier electron density maps reveal structural changes over the first picosecond of retinal photoexcitation. Complementary data for time delays of 10 ps and 8.33 ms allow us to resolve the later stages of the reaction. In combination with refined crystallographic structures at pump-probe delays corresponding to where the spectroscopically characterized I, J, K, and M intermediates form in solution, our time-resolved structural data reveal the trajectory of retinal isomerization and provide atomic details at key points along the reaction.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat0094

The aspartic acid residues of the retinal counterion and functional water molecules in close proximity to the retinal Schiff base respond collectively to the formation and decay of the excited state. This collective motion sets the stage for retinal isomerization, which proceeds via a twisted retinal configuration. Quantum mechanics/molecular mechanics simulations provide theoretical support for this structural evolution.

CONCLUSION: Our observations reveal how, concomitant with the formation of the earliest excited state, the retinal-binding pocket opens up in close proximity to the isomerizing bond. We propose that ultrafast charge transfer along retinal is a driving force for collective motions that contribute to the stereoselectivity and efficiency of retinal isomerization within a protein scaffold. Vibrational quake-like motions extending from retinal to the protein may also be a mechanism through which excess energy is released in a nonradiative fashion. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: joerg.standfuss@psi.ch
Cite this article as P. Nogly *et al.*, *Science* **361**, eaat0094 (2018). DOI: 10.1126/science.aat0094

RESEARCH ARTICLE

STRUCTURAL DYNAMICS

Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser

**Przemyslaw Nogly[1], Tobias Weinert[1,2], Daniel James[1], Sergio Carbajo[3], Dmitry Ozerov[4], Antonia Furrer[1], Dardan Gashi[5], Veniamin Borin[6], Petr Skopintsev[1], Kathrin Jaeger[1], Karol Nass[5,2], Petra Båth[7], Robert Bosman[7], Jason Koglin[3], Matthew Seaberg[3], Thomas Lane[3], Demet Kekilli[1], Steffen Brünle[1], Tomoyuki Tanaka[8,9], Wenting Wu[1], Christopher Milne[5], Thomas White[10], Anton Barty[10], Uwe Weierstall[11], Valerie Panneels[1], Eriko Nango[8,9], So Iwata[8,9], Mark Hunter[3], Igor Schapiro[6], Gebhard Schertler[1,12], Richard Neutze[7], Jörg Standfuss[1]*

Ultrafast isomerization of retinal is the primary step in photoresponsive biological functions including vision in humans and ion transport across bacterial membranes. We used an x-ray laser to study the subpicosecond structural dynamics of retinal isomerization in the light-driven proton pump bacteriorhodopsin. A series of structural snapshots with near-atomic spatial resolution and temporal resolution in the femtosecond regime show how the excited all-trans retinal samples conformational states within the protein binding pocket before passing through a twisted geometry and emerging in the 13-cis conformation. Our findings suggest ultrafast collective motions of aspartic acid residues and functional water molecules in the proximity of the retinal Schiff base as a key facet of this stereoselective and efficient photochemical reaction.

Organisms harvest light for its energy and information content. Seven-transmembrane-helix retinal proteins use the absorption of photons to achieve both of these purposes across all domains of life. Family members include visual rhodopsins (light receptors triggering vision) in animals, as well as bacteriorhodopsin and proteorhodopsins (proton pumps), halorhodopsins (anion pumps), channelrhodopsins (gated ion channels), and sensory rhodopsin (phototaxis receptors) in archaea and bacteria. Retinal-binding ion pumps and channels have found exciting applications in the optogenetic manipulation of neural cells (*1*).

[1]Division of Biology and Chemistry–Laboratory for Biomolecular Research, Paul Scherrer Institut, 5232 Villigen, Switzerland. [2]Photon Science Division–Swiss Light Source, Paul Scherrer Institut, 5232 Villigen, Switzerland. [3]Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. [4]Science IT, Paul Scherrer Institut, 5232 Villigen, Switzerland. [5]SwissFEL, Paul Scherrer Institut, 5232 Villigen, Switzerland. [6]Fritz Haber Center for Molecular Dynamics, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. [7]Department of Chemistry and Molecular Biology, University of Gothenburg, Box 462, SE- 40530 Gothenburg, Sweden. [8]RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan. [9]Department of Cell Biology, Graduate School of Medicine, Kyoto University, Yoshidakonoe- cho, Sakyo-ku, Kyoto 606-8501, Japan. [10]Center for Free-Electron Laser Science (CFEL), DESY, Notkestrasse 85, 22607 Hamburg, Germany. [11]Department of Physics, Arizona State University, Tempe, AZ 85287, USA. [12]Department of Biology, ETH Zürich, 8093 Zürich, Switzerland.
***Corresponding author. Email: joerg.standfuss@psi.ch**

Activation of seven-transmembrane-helix retinal proteins is initiated by the photochemical trans-to-cis (or cis-to-trans for visual rhodopsins) isomerization of the conjugated double bond system of the retinal chromophore. Photo-isomerization is one of the fastest reactions in biology, being completed within a few picoseconds after photon absorption. Retinal isomerization occurs stereoselectively at a specific double bond within the binding pocket of the protein host and with a quantum yield of up to 67% (*2*, *3*). In contrast, illumination of all-trans retinal in solution leads to formation of a mixture of different stereoisomers—13-cis, 11-cis, and 9-cis—with a quantum yield of only a few percent for each subproduct (*4*). How the protein scaffold rapidly guides retinal isomerization toward a single high-yield product is an outstanding and fundamental question in photobiology.

Bacteriorhodopsin (bR) is an archetypical proton pump that has long served as a system for understanding how the energy of a captured photon may be used to achieve unidirectional proton transport against a transmembrane proton concentration gradient. Chemical insights into retinal isomerization have emerged from ultrafast spectroscopy, which identified the I, J, and K intermediates rising and decaying in the ultrafast time regime (Fig. 1) (*5*–*8*). Hybrid quantum mechanics/molecular mechanics (QM/MM) simulations have provided further insight into the initial steps of retinal isomerization, including predictions of energy and charge redistributions (*9*). Despite these advances, theoretical and spectroscopic studies require complementary structural information to build a complete picture of the highly efficient photo-isomerization of retinal within proteins. Researchers have described reaction-cycle intermediates from x-ray structures after illuminating crystals with visible light at low temperatures or changing their pH (*10*). Such intermediate-trapping studies, however, have yielded conflicting results for the earliest available structural intermediate, K (*11*–*13*). Importantly, they cannot address the first events upon photo-isomerization because the low thermal energy barriers are not rate-limiting on the ultrafast time scale.

With the advent of x-ray free-electron lasers (XFELs), time-resolved serial femtosecond crystallography (TR-SFX) has emerged as a powerful method to study ultrafast structural changes in proteins (*14*). TR-SFX was validated against time-resolved Laue diffraction recorded at a synchrotron radiation source, using photoactive yellow protein (PYP) as a model system (*15*). The method has further provided insight into ultrafast structural changes during the photodissociation of carbon monoxide from the active site of myoglobin (*16*), isomerization of *p*-coumaric acid in PYP (*17*), and isomerization of hydroxybenzylidene imidazolinone in rsEGFP2 (reversibly photoswitchable green fluorescent protein 2) (*18*). TR-SFX has also tracked the evolution of protein and water molecule rearrangements in bR from nanoseconds to milliseconds, from which a coherent picture of proton pumping has emerged (*19*). We used this powerful new methodology to characterize the photo-isomerization of retinal and the immediate adaptations of its protein binding pocket.

X-ray diffraction data were collected to 1.50-Å resolution (table S1) from a continuous stream of light-adapted bR microcrystals grown in a lipidic cubic phase (LCP) (*20*, *21*) at the coherent x-ray imaging (CXI) beamline (*22*) of the Linac Coherent Light Source (LCLS) (*23*). About 1 million indexable diffraction patterns were collected in a pump-probe scheme using a sequence of four x-ray probe pulses between every optical pump laser pulse (fig. S1) (*24*). The x-ray diffraction images immediately following the optical laser pulse at four nominal delays were sorted into subgroups based on the LCLS timing tool signal (*25*) with a temporal resolution estimated to be on the order of 200 fs (*16*, *17*) (fig. S2). Sequential windows of data within the first picosecond [overlapping pump-probe delay (Δt) ranges with ~30,000 sorted diffraction patterns], together with data collected at Δt = 10 ps and Δt = 8.33 ms, yielded 20 snapshots of bR activation.

Trajectory of retinal isomerization

A long-distance overview of electron density changes ($F_{obs}^{light} - F_{obs}^{dark}$) in bR during the course of the first 10 ps is shown in movie S1. Changes are already visible for the earliest time point and are initially clustered at the retinal and its binding pocket. A close-up view of the retinal chromophore (Fig. 2 and movie S2) reveals negative density features that are visible above

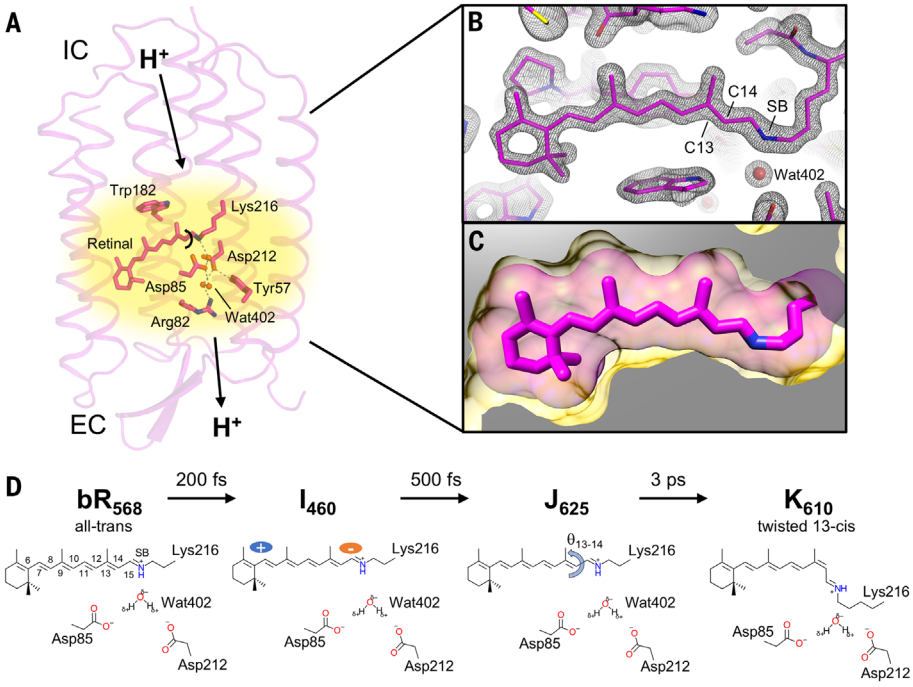


Fig. 1. Early events in bacteriorhodopsin (bR) activation. (**A**) Structural changes in the seven-transmembrane-helix proton-pump bR are initiated by the trans-to-cis isomerization of the covalently bound retinal chromophore. The arrows indicate the direction of H^+ transport from the cytoplasmic side (IC) to the extracellular side (EC) of the membrane. A hydrogen-bonding network including the complex counterion network and several water molecules is critical for maintaining the high affinity of the Schiff base (SB) link for protons. (**B**) Serial femtosecond crystallography allows us to resolve the retinal molecule and its binding pocket with near-atomic resolution ($2F_{obs} - F_{calc}$ electron density of the light-adapted resting state at 1.5-Å resolution shown in gray at 2σ) at ambient temperature and in the membrane-like environment of lipidic cubic phases. The atoms C13 and C14 are labeled to indicate the bond undergoing isomerization. (**C**) The retinal chromophore (sticks with van der Waals radii shown as translucent halos) is tightly encased within the binding pocket of the bR dark state (yellow). (**D**) Retinal isomerization is an ultrafast event progressing within picoseconds from the bR dark state through the I and J intermediates to the K intermediate. I forms within about 200 fs after excitation and is characterized by a rapid rise in the retinal dipole moment (*5*) and twisting of the polyene backbone (*6*). After about 500 fs, the J intermediate evolves, which initiates the isomerization reaction (*7*) and turns into the K state with isomerized 13-cis retinal after about 3 ps (*8*). The ultrafast dynamic interplay between retinal and its binding pocket increases quantum efficiency and guides stereoselectivity of retinal isomerization.

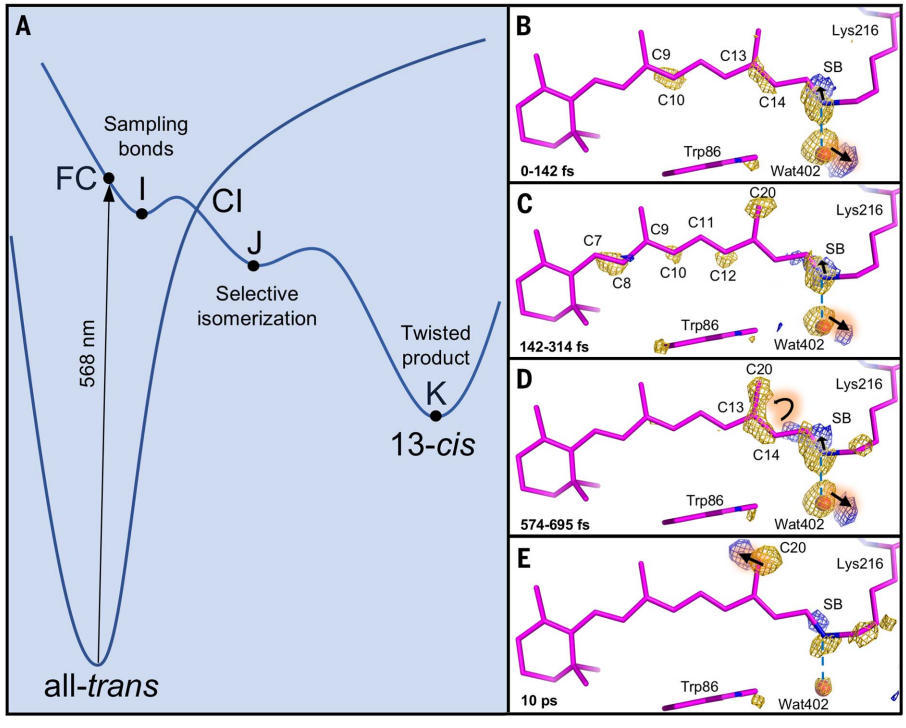
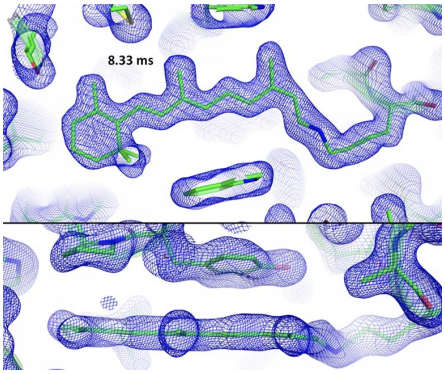


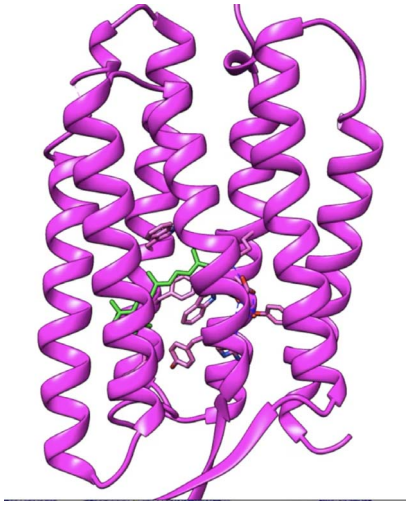
Fig. 2. Femtosecond structural dynamics along the retinal chromophore. (**A**) The potential energy diagram (energy, *y* axis; reaction coordinates, *x* axis) indicates how absorption of a photon promotes all-trans retinal to the Franck-Condon point (FC). The excitation energy is kinetically dissipated after passage through the conical intersection (CI). Trans-cis isomerization occurs along local energy minima correlating with the I, J, and K spectroscopic intermediates. According to the commonly accepted model, the transition of the excited-state intermediate I to the first ground-state product J occurs within about 0.5 ps and coincides with the onset of retinal isomerization (*7*, *27*, *59*). (**B** to **E**) The progression of difference Fourier electron density ($F_{obs}^{light} - F_{obs}^{dark}$, contoured at 4σ; gold, negative; blue, positive) displayed along a dark-state model allows us to follow the trans-cis transition and see how the double bond system is sampled immediately after photoactivation, before selective isomerization around the C13=C14 bond occurs. The process results in an energetically elevated 13-cis isomer that is still twisted by the protein binding pocket 10 ps after activation. All temporal snapshots of this trajectory are available as movie S2.

4σ (σ is the root mean square deviation of electron density) on the carbon atoms of the retinal polyene backbone (associated with C8, C10, C12, and C14) and corresponding weaker positive features (2.7σ to 4.4σ; table S3) for $\Delta t \leq 500$ fs, but which decay for later time delays. Theoretical analyses have argued that photon absorption induces a partial charge relocation in retinal from the Schiff base (SB) toward the β-ionone ring in the excited state and before the trans-to-cis isomerization (*9*). Measurements of transient changes in the excitonic coupling between the retinal dipole moment and nearby tryptophan residues imply that this charge redistribution occurs within 200 fs after photon absorption (*5*).

Simultaneously, nuclear rearrangements lead to an extension of double bonds and contraction of the single bonds (so-called inversion of the bond length alternation), which allows the planar geometry of retinal to distort in the excited state (*26*). Such distortion is consistent with the negative and (weaker) positive features of the difference Fourier electron density map (F_{obs}^{light} – F_{obs}^{dark}) observed around the C7 to C14 atoms of the retinal on this time scale (Fig. 2).



Movie 1. Retinal isomerization in refined structures with extrapolated electron density ($2F_{ext} - F_{calc}$, contoured at 1.2σ).



Movie 2. Dynamic view of retinal isomerization and response of the counterion network. The movie shows the transitions between four structural intermediates: ground state, Δt = 49 to 406 fs, Δt = 457 to 646 fs, and Δt = 10 ps. The morphs between the refined structures were generated with Chimera (*60*) and reveal a fast motion proceeding through the entire structure, followed by a slower rearrangement. A zoom-in to the region of the counterion network highlights the hydrogen-bond breakage between the SB and Wat^{402} (W 402). The detailed view of the retinal isomerization visualizes the structural rearrangements as the isomerization proceeds. Time scales in the movie do not correlate to real time scales, and the structures are linear morphs of atom motions. The 457 to 646 fs → 10 ps transition is shown more slowly than the other two transitions to highlight the longer time scale of the transition.

The retinal initially samples possible isomerization geometries for $\Delta t \leq 500$ fs, but the possible evolution is heavily constrained by the rigidity of the binding pocket. After this sampling of double bond geometries in the initial excited state, electron density changes focus around the C13=C14 double bond, suggesting a coordinated evolution throughout the trans-cis isomerization. The progress of retinal isomerization can be followed in extrapolated electron density maps (fig. S3 and Movie 1). Changes over time are best visualized by overlaying the difference Fourier electron density map with the resting and refined structures for the time ranges Δt = 49 to 406 fs, 457 to 646 fs, 10 ps, and 8.33 ms (table S2) with Fig. 3. These time windows correlate approximately with accumulation of the spectroscopically distinct I, J, K, and M intermediates identified in solution (*5*–*8*, *27*, *28*). In the earliest temporal windows, paired difference density features indicate that the retinal SB nitrogen is already displaced toward helix G, which is quantified by structural refinement as a 0.5-Å sideways motion of the SB but with the retinal geometry otherwise remaining similar to the all-trans conformation. For Δt = 457 to 646 fs, a stronger continuous negative difference is apparent along the C20–C13=C14 bonds, and two positive difference density peaks flank a negative feature on the SB nitrogen. This time point is modeled as a twisted retinal configuration with the C12–C13=C14–C15 torsion angle ~90° and the C15=Nζ bond lying in the plane of the membrane. Overall changes between the dark and Δt = 457 to 646 fs time delay twist the torsion angles of the C13=C14 and the neighboring C11=C12 and C15=Nζ bonds in opposite directions, indicating that the reaction proceeds according to the aborted bicycle-pedal model of retinal isomerization (*9*, *29*, *30*).

By Δt = 10 ps, only one positive feature remains in the difference Fourier electron density map that is complementary to the negative peak on the SB nitrogen atom, and paired positive and negative density features are associated with the retinal C20 methyl group. Structural refinement indicates that these changes represent the retinal having reached its twisted 13-cis configuration and the C20 methyl being twisted toward helix G. This sequence of events is entirely consistent with QM/MM simulations of the retinal excited state when the key water molecule (Wat^{402}), Lys^{216}, and the counterion network are included

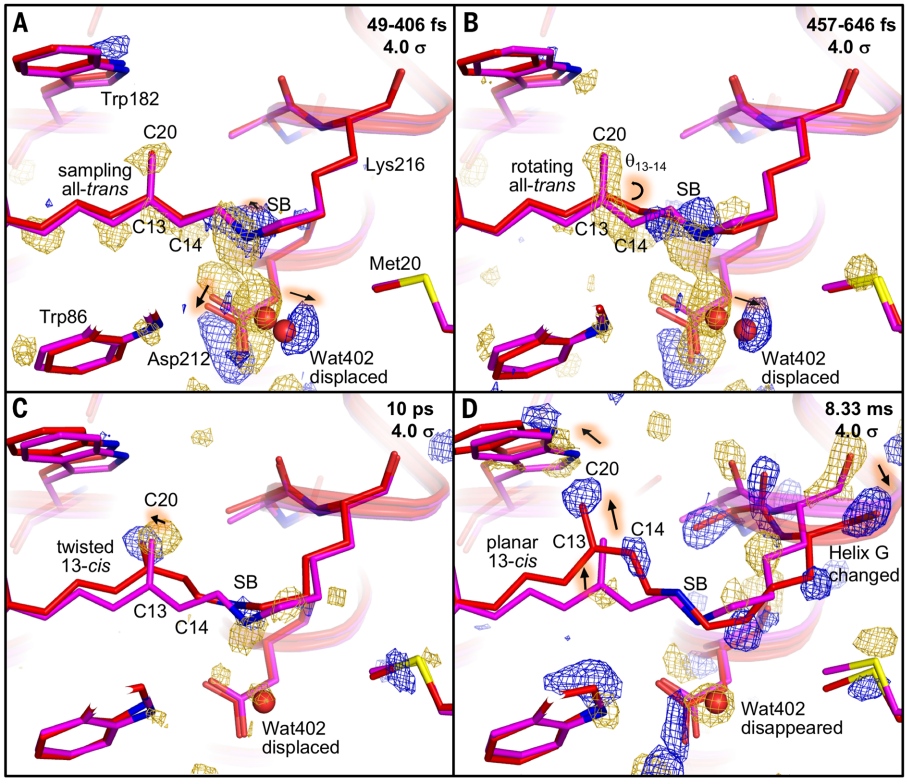


Fig. 3. Structural intermediates at pump-probe delays (Δt) corresponding to times when key spectral intermediates of the bR photocycle occur. (**A**) Δt = 49 to 406 fs, I intermediate. (**B**) Δt = 457 to 646 fs, J intermediate. (**C**) Δt = 10 ps, K intermediate. (**D**) Δt = 8.33 ms, M intermediate. The structure of the bR dark state is shown in magenta and overlaid with the bR reaction intermediates in red. The difference Fourier electron density map (F_{obs}^{light} – F_{obs}^{dark}, contoured at 4σ; gold, negative; blue, positive) and arrows indicate the structural rearrangements around the isomerizing C13=C14 bond and the retinal counterion. The corresponding extrapolated electron density maps ($2F_{ext}$ – F_{calc}) are available as fig. S3 and Movie 1.

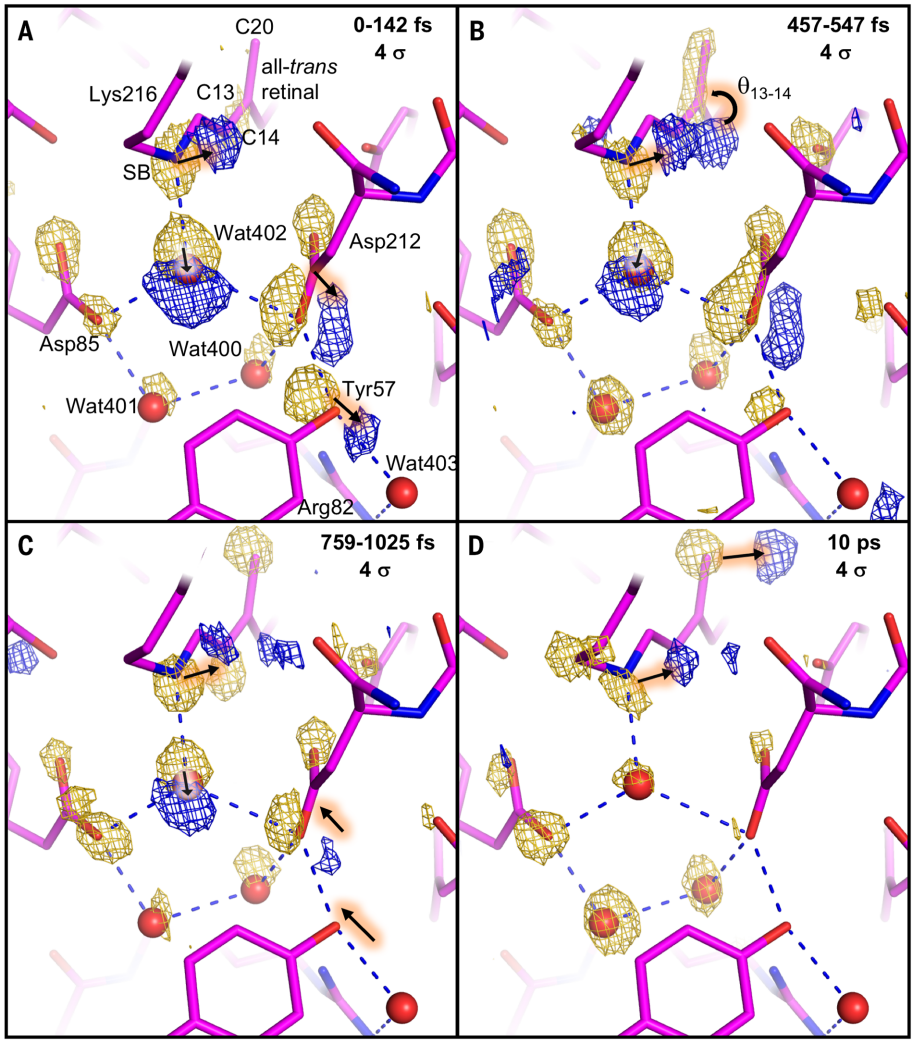


Fig. 4. Close-up view of the protonated SB and associated counterion network. (**A** to **D**) Progression of the difference Fourier electron density map ($F_{obs}^{light} - F_{obs}^{dark}$, contoured at 4σ; gold, negative; blue, positive) shows the coherent motions of the hydrogen-bonding network connecting the SB with Asp^{85}, Asp^{212}, and Tyr^{57} via water molecules Wat^{402}, Wat^{401}, and Wat^{400}. Arrows indicate correlated motions in retinal and the hydrogen-bonding network. The sticks represent the dark-state model of bR. All temporal snapshots are available as movie S3.

within the quantum chemical description [fig. S4 and (*9*)], providing theoretical support for the structural evolution that we observed. Lastly, the difference Fourier electron density map recorded for Δt = 8.33 ms displays all major features previously observed for Δt = 1.725 ms (*19*), including the planar 13-cis retinal characteristic of later reaction intermediates of the bR photocycle.

Ultrafast response of the counterion cluster

A key feature in the resting bR state is a hydrogen bond (H-bond) interaction of the SB with Wat^{402}, which participates in the H-bonding network with the counterions Asp^{85} and Asp^{212} and stabilizes the positive charge on the SB nitrogen (*31*). Mutation of either Asp^{85} or Asp^{212} to noncharged residues greatly reduces the rate of retinal isomerization (*32*), and both have been implicated in guiding stereoselectivity of the reaction (*33*).

Wat^{402} and the SB respond to the photon absorption of retinal even before the isomerization appears to begin, moving up to 0.6 Å further away from each other (to a distance of 3.25 Å) after light absorption (Δt = 49 to 406 fs). This separation increases to 4.05 Å once the retinal is isomerized (Δt = 10 ps). This movement is visualized as very strong, paired positive and negative features in the difference Fourier electron density map (from 6σ to 11σ; Fig. 4) associated with Wat^{402}, Asp^{212}, and Tyr^{57}, which together reveal an ultrafast collective motion of a group of polar residues connected through H-bond interactions (Movie 2). The driving force for this motion is the ultrafast redistribution of the positive charge away from the SB and toward the β-ionone ring. Quantum chemical calculation of the difference between the ground- and excited-state electron density (fig. S4) shows the localization of changes along the retinal polyene chain. The sudden polarization instantaneously pushes the water molecule away from the SB (up to a distance of 3.65 Å at Δt= 457 to 646 fs) and breaks the SB–Wat^{402} H-bond (which would otherwise oppose the SB reorientation during isomerization). Loss of the H-bond lowers the energy barrier for a specific isomerization pathway (*34*), allowing the retinal to isomerize about the C13=C14 bond. These collective motions are later damped, with the retinal being fully isomerized at Δt = 10 ps, while Wat^{402} loses its interaction with the SB, consistent with a changing H-bonding pattern in this region predicted by low-temperature Fourier transform infrared (*35*) and Raman spectroscopy (*36*). The rigidity of the large hydrophobic region of the retinal-binding pocket, which prevents unwanted C7=C8, C9=C10, or C11=12 cis products, is in stark contrast with the specific ultrafast motions associated with Wat^{402} and the counterion cluster. These findings implicate ultrafast collective motions as an essential ingredient for efficient isomerization within the protein scaffold.

Protein quake

The collective motions from the counterion region extend up to 12 Å away from the SB nitrogen within 600 fs (fig. S5) yet become dispersed on longer time scales. This dampening is evidenced by weaker paired difference densities associated with Tyr^{57} at Δt = 759 to 1025 fs, as well as the motions of Thr^{47} and Asp^{212}, which are no longer visible at Δt = 10 ps. Changes are further apparent when comparing the refined structural intermediates at Δt = 49 to 406 fs, 457 to 646 fs, and 10 ps (Movie 2). As such, these motions propagate away from the active site at 2 nm/ps, which is noticeably faster than the speed of sound in water (~1.5 nm/ps). These observations are consistent with the propagation of motions in the bacterial photoreaction center (*37*) and myoglobin (*16*, *38*, *39*) and the theory of protein quakes that suggests how excess energy in proteins can be dissipated in earthquake-like motions of collective structural deformations (*40*). In the case of photoactive proteins such as bR, these motions may be the mechanism through which excess energy from decay of the excited state is released in a nonradiative fashion.

Our ultrafast TR-SFX data recorded using an XFEL captures a fundamental photochemical reaction and one of the fastest processes in biology. In particular, our measurements uncover the initial structural rearrangements in the excited state that guide the selective isomerization process. Photo-excited retinal initially samples multiple isomerization pathways within its binding pocket, but the ultimate fate of the reaction is steered by the near-instantaneous charge redistribution along the retinal and the resulting collective motions of the SB counterion complex. Breaking of the SB–Wat^{402} H-bond enables the trans-to-cis isomerization about the C13=C14 double bond. Similar ultrafast collective motions may guide

the stereochemistry and enhance the quantum efficiency of retinal isomerization in such diverse processes as phototrophy in the seas, neuron stimulation in optogenetics, and the initiation of visual signals within the photoreceptors of our eyes.

Materials and methods

Purification and crystallization of bR

Bacteriorhodopsin was purified from purple membranes of *Halobacterium salinarum*. The purification and crystallization was performed as described previously (*24*, *41*) with a few modifications. The purple membranes were solubilized overnight in the presence of 1.7% β-octylglucoside (Anatrace) and 50 mM sodium phosphate buffer pH 6.9 (GERBU). After solubilization the pH was adjusted to 5.5 with 0.1 M HCl and the insoluble fraction was removed by 1-hour centrifugation at 150 000 × *g*. The size exclusion purification step was omitted following (*19*). The protein was concentrated to 40 to 80 mg ml^{-1} using Millipore centricolumns with 50 kDa cutoff. Lipidic Cubic Phase (LCP) for crystallization was obtained in a Hamilton syringe by mixing protein with monoolein (Nu-Chek) in a 42:58 ratio. Subsequently the LCP was slowly injected into another Hamilton syringe filled with crystallization buffer consisting of 100 mM Sorensen buffer pH 5.6 (GERBU) and 30% polyethylene glycol 2000 (Molecular Dimensions). The crystallization was carried out at 21°C and crystals between 15 μm and 50 μm along the longest edge and with a thickness of 1 to 5 μm were obtained within 3 to 6 days (average crystals size 35x35x3 μm). All purification and crystallization steps were performed under dim red light or in the dark.

Sample preparation for LCP-SFX

The crystallization buffer was removed from LCP containing crystals via a syringe coupler by slow pressing on the syringe plunger. Shortly before the experiment monoolein was added to bring the sample into a homogeneous and transparent LCP mesophase. The sample was supplemented with 5% of MAG 7.9 to prevent phase transition during injection into vacuum and 5% of paraffin was added to support smooth flow of the sample. The crystal density within the sample was homogenized using a custom made "3-way syringe coupler". It consisted of three syringe inlets and the simultaneous passage of sample from two syringes into a third one over several cycles ensured optimal crystal distribution throughout the sample volume for smooth jet operation. The stability of jetting was tested offline with a high-speed camera providing further evidence that the addition of paraffin improved stable jetting. Selected batches of the prepared crystalline sample were tested in advance to confirm high hit rates and diffraction quality using the Swiss Light Source set up for LCP injection (*42*). Before data collection the sample was light adapted at 300 mW for 5 min through a long-pass yellow filter (>515 nm). Sample was used for data collection no longer than 30 min after light exposure to prevent decay of the light adapted dark state. On average, 1 ml of final sample was used per 12-hour shift.

Data collection

The TR-SFX experiment was performed in June 2017 in the microfocus chamber of the CXI beamline at LCLS (*43*) with the integrated optical pump-probe laser set-up (*44*). The 50 fs XFEL pulses at 9.5 keV were delivered at 120 Hz into a vacuum chamber. In the experiment we used a modified version of the LCP injector (*20*) with a larger sample reservoir of 130 μl to allow a longer data collection and the higher flow rates (2.5 μl min^{-1} through a 50 μm diameter nozzle and, for a small amount of data, 5 μl min^{-1} through a 75 μm diameter nozzle) necessary for time-resolved crystallography (*24*). The LCP stream was aligned with the x-ray beam and an Ti:Sapphire-pump Optical Parametric Amplifier producing optical laser pulses with a 30 Hz repetition rate, 100 fs pulse length, 529 nm wavelength and 17 μJ energy in a focal spot of 95 μm ($1/e^2$). The arrival time of the optical laser pulse relative to the XFEL pulse, was monitored using the timing tool at LCLS (*25*, *45*) with nominal time delays Δt = −500 fs, 300 fs, 600 fs, 900 fs, 1100 fs, and 10 ps. As the exact timing is subject to jitter, data collected at these delays were sorted and binned after collection based on their timing tool signal (see Data sorting and binning). By combining all indexed patterns from the second XFEL pulse after the pump laser excitation (175 μm from the interaction region) we could further resolve Δt = 8.33 ms, but with lower occupancy due to this time point being excited further from the pump laser peak intensity (fig. S1). We obtained a high quality dark dataset (table S1) by merging patterns of the fourth XFEL pulse after light excitation from all datasets. For clarity, the 120 Hz XFEL, 30 Hz optical data collection scheme is further illustrated in fig. S1.

An important parameter for a successful TR-SFX experiment is the laser power density at sample position, which is critical for efficient activation but difficult to estimate. In our case we aligned the laser approximately 50 μm below the x-ray interaction region to exclude light contamination of the following pulses. This shift reduced the nominal laser power density of 2.4 TW/cm^2 to 0.324 TW/cm^2 [factor of 0.135 ($1/e^{-2}$)] at the x-ray interaction region. This value is in a similar range as in previously published ultrafast TR-SFX studies (*16*, *17*, *46*). Using an offline setup with a continuous laser we established that scattering from the 50 μm wide LCP jet (without protein or crystals) reduces the incident laser light by about 80%. Scattering from the LCP jet and the high optical density of the crystals further reduced the average light intensity any given bR molecule was exposed to. Taking the typical crystal dimensions of 35x35x3 μm^3 and the parameters above into account, we estimate 3.3 to 0.6 absorbed photons per retinal depending on the orientation of the crystal with respect to the laser pulse. This exposure is within the range where multiphoton effects are negligible in a crystallographic analysis based on the small spectroscopic effect under similar laser conditions (*46*–*48*). Moreover, our setup provided comparable results to our previous experiment using a nanosecond pump laser with much lower power density (0.06 GW cm^{-2}) at SACLA (*19*), with a twisted retinal at 10 ps and 16 ns (fig. S7). The two delays correspond to the lower and upper end of the temporal range where the K intermediate of bR is observed spectroscopically and the high similarity between the two structures indicates that productive photocycles can be induced using a femtosecond laser pump pulse. The similarity of structures for 10 ps/16 ns provides us with an internal control for the presented experiments in the femtosecond range.

Our choice of laser energy was guided by two previous TR-SFX experiments with very similar femtosecond pump laser parameters but different laser energies (fig. S9). In December 2015, we used 8 μJ and 12 μJ, with the former yielding nearly flat Fourier difference electron density maps while the latter provided very weak signal with a few peaks at 3.5 σ and the estimated activation levels of about 5%. In July 2014, we used 20 μJ which yielded about 13% activation levels (*24*). In the currently reported experiment we aimed for activation levels better than 10% and measured all time delays with a consistent 17 μJ in order to collect interpretable crystallographic data.

Data sorting and binning

Diffraction images were selected with Cheetah (*49*) followed by further sorting according to event codes corresponding to one of the four images collected after the optical laser pulse.

Light data in the femtosecond range were sorted by pump-probe delay time using the timing tool available at CXI with yttrium aluminum garnet (YAG) target (*45*). The spectrally encoded delay was determined by the standard matched-filter procedure implemented in psana, the LCLS-provided analysis software (*50*). Strong etalon effects caused by the thick YAG target necessitated one modification to this standard analysis. Before edge-position determination by matched filtering, the spectral signal was bandstop filtered using a 5th-order Butterworth filter with a cut-on and cut-off frequency of 45 fs^{-1} and 120 fs^{-1} respectively, effectively removing signal oscillations due to etalon. The data where initially sorted into 20 bins with 10,000 images each (see fig. S1). For calculating structural snapshots, we combined 3 overlapping bins which lead to 18 snapshots in the femtosecond range, which we further complement with data collected at Δt =10 ps and 8.33 ms (see table S1).

Data processing

The images were indexed using mosflm and dirax indexing algorithms on the peaks identified by Cheetah. Data were integrated (the integration radius was 3 pixels for the peak region and 4 to 7 pixels for the background annulus with integrating 1.2 nm^{-1} beyond the conservative resolution limit determined on a per image basis) and merged with CrystFEL (*51*) version 0.6.3. A geometry refinement using geoptimiser was carried out on a subset of images that diffracted beyond 2.0 Å in CrystFEL's conservative apparent resolution. This step improved high-resolution statistics after reindexing and reintegration of the data.

The indexing ambiguity in space group $P6_3$ was resolved using ambigator in CrystFEL (*52*). Data were scaled and merged using partialator in CrystFEL without the calculation of partialities (–model = unity), using one cycle of scaling and applying the option to extend the conservative resolution limit determined on the per image basis ("–pushres 1.2"). A small subset of patterns (11,758) with HPLC back pressures 25 bars above the average value for each data run was discarded in order to eliminate any source of potential light contamination in dark data due to occasional jet slowdowns owing to higher crystal density.

Data analysis

For the structure determination of the dark state, diffraction images from every fourth diffraction image after the optical laser pulse were used. The signal recorded by the timing tool for every first diffraction image after the optical laser pulse was used to sort the diffraction images after the experiment into bins with defined time ranges (see Data processing and binning). Each of the datasets (table S1) consists of three overlapping original bins (fig. S1). The time ranges used for data analysis were initially chosen to contain around 30,000 images (about 27,000 after indexing, see table S1) to ensure good quality data for analysis and to be comparable to the number of images collected at 10 ps. The resolution cutoff for the light data was established so that most datasets have at least a CC* of 0.5 at 1.50 Å. We used a dark dataset of 29,461 patterns collected without optical pump laser to cross-check our high quality dark dataset for light contamination and could not observe difference density peaks corresponding to the M-state above the noise level of 2.5σ.

Structure determination and refinement of the bR dark state

PDB entry with accession code 5J7A (*24*) with all water molecules and lipids removed was used as molecular replacement model in Phaser (*53*). The resting state structure was obtained with several cycles of building and refinement using COOT (*54*) and PHENIX (*55*) to a resolution of 1.5 Å (table S2).

Calculation of difference Fourier density maps

The difference Fourier density maps were calculated with phenix.fobs_minus_fobs_map (*55*) including the multiscaling feature and low resolution cutoff at 4.5 Å. The electron density peak minima and maxima in σ units (= root mean square electron density of the unit cell) were extracted in COOT (*56*). Table S3 lists peak values for regions of the density map which at any point in the ultrafast time range exceeded ± 4σ. The values below 2.5σ were set to 0 as being of too low significance.

Refinement of intermediate states

In a TR-SFX experiment, crystals subjected to optical laser light pulse contain a mixture containing light activated molecules and dark state molecules. We used extrapolated structure factors F_{ext} (*57*) to enable modeling of bacteriorhodopsin activated states, similar approach as used to model the subpicosecond structural states of PYP (*17*). The extrapolated structure factors were calculated using a linear approximation (*58*) as follows: $F_{ext} = [(F_{obs}^{light} - F_{obs}^{dark}) / \text{activated fraction}] + F_{obs}^{dark}$. The $2F_{ext} - F_{calc}$ maps calculated with phases of the dark state model showed distinct features in agreement with the $F_{obs}^{light} - F_{obs}^{dark}$ Fourier difference maps. Using this comparison, the light activated fraction of molecules in the data time ranges 49 to 406 fs, 457 to 646 fs, 10 ps, and 8.3 ms was estimated to be 25%, 18%, 16%, and 10% corresponding to fractions where the features associated exclusively with the dark state disappear from extrapolated maps. Reciprocal space refinement against the F_{ext} to 1.9 Å resolution in PHENIX (*55*) in combination with real space fitting in COOT (*56*) provided models (table S2) for the respective time points with all of them showing about 98% of residues in the Ramachandran favored region. A comparison of calculated ($F_{calc}^{refined} - F_{calc}^{dark}$) and experimental ($F_{obs}^{light} - F_{obs}^{dark}$) difference densities (fig. S8) was used as additional control to validate the refined structures. Comparison of the bR 8.33 ms structure obtained from the refinement against the extrapolated structure factors with the 1.725 ms structure (*19*) obtained by an alternative conformation refinement approach shows a high level of similarity between the structures derived using the two different approaches (fig. S6).

QM/MM calculations

We used a hybrid quantum mechanics/molecular mechanics (QM/MM) method to study the photochemical reaction of bR. The key step in setting up such a QM/MM calculation is the partitioning of the total system in QM and MM subsystems. The QM subsystem is described using a quantum chemical method, while the MM subsystem is computed by a classical force field. Following the crystallographic results, we included the retinal chromophore and the Lys^{216} sidechain, as well as the counter-ion sidechains Arg^{82}, Asp^{85}, Asp^{212} and the water molecules Wat^{402}, Wat^{404}, Wat^{406} to the QM part, while the rest of the protein was treated at the MM level. The MM subsystem is treated using an AMBER force field and interacts with the QM subsection through electrostatic embedding. In all the simulations we have performed, the protein backbone was kept fixed in the crystallographic position and only the QM section and all sidechains, which have at least one atom within 5 Å from the retinal Schiff base and Lys^{216} moiety was allowed to be moved freely in the geometry optimization.

To optimize the geometry in the ground state the QM subsystem was described using the pure-GGA BP86 functional with Grimme's corrections for dispersion interactions and the cc-pVDZ basis set. The difference between the ground and excited state density was computed using CAM-B3LYP and the cc-pVTZ basis set.

REFERENCES AND NOTES

1. B. R. Rost, F. Schneider-Warme, D. Schmitz, P. Hegemann, Optogenetic Tools for Subcellular Applications in Neuroscience. *Neuron* **96**, 572–603 (2017). doi: 10.1016/j.neuron.2017.09.047; pmid: 29096074
2. P. Hamm *et al*., Femtosecond spectroscopy of the photoisomerisation of the protonated Schiff base of all-trans retinal. *Chem. Phys. Lett.* **263**, 613–621 (1996). doi: 10.1016/S0009-2614(96)01269-9
3. O. P. Ernst *et al*., Microbial and animal rhodopsins: Structures, functions, and molecular mechanisms. *Chem. Rev.* **114**, 126–163 (2014). doi: 10.1021/cr4003769; pmid: 24364740
4. K. A. Freedman, R. S. Becker, Comparative investigation of the photoisomerization of the protonated and unprotonated n-butylamine Schiff bases of 9-cis-, 11-cis-, 13-cis-, and all-trans-retinals. *J. Am. Chem. Soc.* **108**, 1245–1251 (1986). doi: 10.1021/ja00266a020
5. S. Schenkl, F. van Mourik, G. van der Zwan, S. Haacke, M. Chergui, Probing the ultrafast charge translocation of photoexcited retinal in bacteriorhodopsin. *Science* **309**, 917–920 (2005). doi: 10.1126/science.1111482; pmid: 16081732
6. T. Kobayashi, T. Saito, H. Ohtani, Real-time spectroscopy of transition states in bacteriorhodopsin during retinal isomerization. *Nature* **414**, 531–534 (2001). doi: 10.1038/35107042; pmid: 11734850
7. J. Herbst, K. Heyne, R. Diller, Femtosecond infrared spectroscopy of bacteriorhodopsin chromophore isomerization. *Science* **297**, 822–825 (2002). doi: 10.1126/science.1072144; pmid: 12161649
8. G. H. Atkinson, T. L. Brack, D. Blanchard, G. Rumbles, Picosecond time-resolved resonance Raman spectroscopy of the initial trans to cis isomerization in the bacteriorhodopsin photocycle. *Chem. Phys.* **131**, 1–15 (1989). doi: 10.1016/0301-0104(89)87077-6
9. P. Altoè, A. Cembran, M. Olivucci, M. Garavelli, Aborted double bicycle-pedal isomerization with hydrogen bond breaking is the primary event of bacteriorhodopsin proton pumping. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20172–20177 (2010). doi: 10.1073/pnas.1007000107; pmid: 21048087
10. C. Wickstrand, R. Dods, A. Royant, R. Neutze, Bacteriorhodopsin: Would the real structural intermediates please stand up? *Biochim. Biophys. Acta* **1850**, 536–553 (2015). doi: 10.1016/j.bbagen.2014.05.021; pmid: 24918316
11. K. Edman *et al*., High-resolution X-ray structure of an early intermediate in the bacteriorhodopsin photocycle. *Nature* **401**, 822–826 (1999). doi: 10.1038/44623; pmid: 10548112
12. B. Schobert, J. Cupp-Vickery, V. Hornak, S. Smith, J. Lanyi, Crystallographic structure of the K intermediate of bacteriorhodopsin: Conservation of free energy after photoisomerization of the retinal. *J. Mol. Biol.* **321**, 715–726 (2002). doi: 10.1016/S0022-2836(02)00681-2; pmid: 12206785
13. Y. Matsui *et al*., Specific damage induced by X-ray radiation and structural changes in the primary photoreaction of bacteriorhodopsin. *J. Mol. Biol.* **324**, 469–481 (2002). doi: 10.1016/S0022-2836(02)01110-5; pmid: 12445782
14. R. Neutze, K. Moffat, Time-resolved structural studies at synchrotrons and X-ray free electron lasers: Opportunities and challenges. *Curr. Opin. Struct. Biol.* **22**, 651–659 (2012). doi: 10.1016/j.sbi.2012.08.006; pmid: 23021004
15. J. Tenboer *et al*., Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242–1246 (2014). doi: 10.1126/science.1259357; pmid: 25477465
16. T. R. M. Barends *et al*., Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* **350**, 445–450 (2015). doi: 10.1126/science.aac5492; pmid: 26359336
17. K. Pande *et al*., Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* **352**, 725–729 (2016). doi: 10.1126/science.aad5081; pmid: 27151871
18. N. Coquelle *et al*., Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nat. Chem.* **10**, 31–37 (2018). doi: 10.1038/nchem.2853; pmid: 29256511
19. E. Nango *et al*., A three-dimensional movie of structural changes in bacteriorhodopsin. *Science* **354**, 1552–1557 (2016). doi: 10.1126/science.aah3497; pmid: 28008064
20. U. Weierstall *et al*., Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5**, 3309 (2014). doi: 10.1038/ncomms4309; pmid: 24525480

21. E. M. Landau, J. P. Rosenbusch, Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14532–14535 (1996). doi: 10.1073/pnas.93.25.14532; pmid: 8962086
22. S. Boutet, G. J. Williams, The Coherent X-ray Imaging (CXI) instrument at the Linac Coherent Light Source (LCLS). *New J. Phys.* **12**, 035024 (2010). doi: 10.1088/1367-2630/12/3/035024
23. P. Emma *et al*., First lasing and operation of an ångstrom-wavelength free-electron laser. *Nat. Photonics* **4**, 641–647 (2010). doi: 10.1038/nphoton.2010.176
24. P. Nogly *et al*., Lipidic cubic phase injector is a viable crystal delivery system for time-resolved serial crystallography. *Nat. Commun.* **7**, 12314 (2016). doi: 10.1038/ncomms12314; pmid: 27545823
25. M. Harmand *et al*., Achieving few-femtosecond time-sorting at hard X-ray free-electron lasers. *Nat. Photonics* **7**, 215–218 (2013). doi: 10.1038/nphoton.2013.11
26. R. González-Luque *et al*., Computational evidence in favor of a two-state, two-mode model of the retinal chromophore photoisomerization. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9379–9384 (2000). doi: 10.1073/pnas.97.17.9379; pmid: 10944211
27. S. Schenkl *et al*., Insights into excited-state and isomerization dynamics of bacteriorhodopsin from ultrafast transient UV absorption. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4101–4106 (2006). doi: 10.1073/pnas.0506303103; pmid: 16537491
28. G. H. Atkinson, L. Ujj, Y. Zhou, Vibrational Spectrum of the J-625 Intermediate in the Room Temperature Bacteriorhodopsin Photocycle. *J. Phys. Chem. A* **104**, 4130–4139 (2000). doi: 10.1021/jp9918306
29. A. Warshel, Bicycle-pedal model for the first step in the vision process. *Nature* **260**, 679–683 (1976). doi: 10.1038/260679a0; pmid: 1264239
30. A. Warshel, Z. T. Chu, Nature of the Surface Crossing Process in Bacteriorhodopsin: Computer Simulations of the Quantum Dynamics of the Primary Photochemical Event. *J. Phys. Chem. B* **105**, 9857–9871 (2001). doi: 10.1021/jp010704a
31. M. Sheves, N. Friedman, A. Albeck, M. Ottolenghi, Primary photochemical event in bacteriorhodopsin: Study with artificial pigments. *Biochemistry (Mosc.)* **24**, 1260–1265 (1985). doi: 10.1021/bi00326a031
32. L. Song, M. A. El-Sayed, J. K. Lanyi, Protein catalysis of the retinal subpicosecond photoisomerization in the primary process of bacteriorhodopsin photosynthesis. *Science* **261**, 891–894 (1993). doi: 10.1126/science.261.5123.891; pmid: 17783735
33. T. Yamato, T. Kakitani, Molecular Dynamics Simulation of the Excited-State Dynamics of Bacteriorhodopsin. *Photochem. Photobiol.* **66**, 735–740 (1997). doi: 10.1111/j.1751-1097.1997.tb03217.x
34. F. Gai, K. C. Hasson, J. C. McDonald, P. A. Anfinrud, Chemical dynamics in proteins: The photoisomerization of retinal in bacteriorhodopsin. *Science* **279**, 1886–1891 (1998). doi: 10.1126/science.279.5358.1886; pmid: 9506931
35. M. Shibata, H. Kandori, FTIR studies of internal water molecules in the Schiff base region of bacteriorhodopsin. *Biochemistry* **44**, 7406–7413 (2005). doi: 10.1021/bi050122+; pmid: 15895984
36. S. Shim, J. Dasgupta, R. A. Mathies, Femtosecond time-resolved stimulated Raman reveals the birth of bacteriorhodopsin's J and K intermediates. *J. Am. Chem. Soc.* **131**, 7592–7597 (2009). doi: 10.1021/ja809137x; pmid: 19441850
37. D. Arnlund *et al*., Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser. *Nat. Methods* **11**, 923–926 (2014). doi: 10.1038/nmeth.3067; pmid: 25108686
38. L. U. L. Brinkmann, J. S. Hub, Ultrafast anisotropic protein quake propagation after CO photodissociation in myoglobin. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10565–10570 (2016). doi: 10.1073/pnas.1603539113; pmid: 27601659
39. M. Levantino *et al*., Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. *Nat. Commun.* **6**, 6772 (2015). doi: 10.1038/ncomms7772; pmid: 25832715
40. A. Ansari *et al*., Protein states and proteinquakes. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5000–5004 (1985). doi: 10.1073/pnas.82.15.5000; pmid: 3860839
41. P. Nogly *et al*., Lipidic cubic phase serial millisecond crystallography using synchrotron radiation. *IUCrJ* **2**, 168–176 (2015). doi: 10.1107/S2052252514026487; pmid: 25866654
42. T. Weinert *et al*., Serial millisecond crystallography for routine room-temperature structure determination at synchrotrons. *Nat. Commun.* **8**, 542 (2017). doi: 10.1038/s41467-017-00630-4; pmid: 28912485
43. M. Liang *et al*., The Coherent X-ray Imaging instrument at the Linac Coherent Light Source. *J. Synchrotron Radiat.* **22**, 514–519 (2015). doi: 10.1107/S160057751500449X; pmid: 25931062
44. M. P. Minitti *et al*., Optical laser systems at the Linac Coherent Light Source. *J. Synchrotron Radiat.* **22**, 526–531 (2015). doi: 10.1107/S1600577515006244; pmid: 25931064
45. M. R. Bionta *et al*., Spectral encoding of x-ray/optical relative delay. *Opt. Express* **19**, 21855–21865 (2011). doi: 10.1364/OE.19.021855; pmid: 22109037
46. B. Schmidt *et al*., Excited-state dynamics of bacteriorhodopsin probed by broadband femtosecond fluorescence spectroscopy. *Biochim. Biophys. Acta Bioenerg.* **1706**, 165–173 (2005). doi: 10.1016/j.bbabio.2004.10.008; pmid: 15620377
47. A. C. Florean *et al*., Control of retinal isomerization in bacteriorhodopsin in the high-intensity regime. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10896–10900 (2009). doi: 10.1073/pnas.0904589106; pmid: 19564608
48. V. I. Prokhorenko *et al*., Coherent control of retinal isomerization in bacteriorhodopsin. *Science* **313**, 1257–1261 (2006). doi: 10.1126/science.1130747; pmid: 16946063
49. A. Barty *et al*., *Cheetah*: Software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J. Appl. Crystallogr.* **47**, 1118–1131 (2014). doi: 10.1107/S1600576714007626; pmid: 24904246
50. D. Damiani *et al*., Linac Coherent Light Source data analysis using psana. *J. Appl. Crystallogr.* **49**, 672–679 (2016). doi: 10.1107/S1600576716004349
51. T. A. White *et al*., CrystFEL: A software suite for snapshot serial crystallography. *J. Appl. Crystallogr.* **45**, 335–341 (2012). doi: 10.1107/S0021889812002312
52. W. Brehm, K. Diederichs, Breaking the indexing ambiguity in serial crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 101–109 (2014). doi: 10.1107/S1399004713025431; pmid: 24419383
53. A. J. McCoy *et al*., Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007). doi: 10.1107/S0021889807021206; pmid: 19461840
54. P. Emsley, K. Cowtan, *Coot*: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004). doi: 10.1107/S0907444904019158; pmid: 15572765
55. P. D. Adams *et al*., *PHENIX*: Building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002). doi: 10.1107/S0907444902016657; pmid: 12393927
56. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). doi: 10.1107/S0907444910007493; pmid: 20383002
57. M. Schmidt, in *Ultrashort Laser Pulses in Biology and Medicine*, M. Braun, P. Gilch, W. Zinth, Eds. (Springer, 2008), pp. 201–241.
58. U. K. Genick *et al*., Structure of a protein photocycle intermediate by millisecond time-resolved crystallography. *Science* **275**, 1471–1475 (1997). doi: 10.1126/science.275.5305.1471; pmid: 9045611
59. R. A. Mathies, C. H. Brito Cruz, W. T. Pollard, C. V. Shank, Direct observation of the femtosecond excited-state cis-trans isomerization in bacteriorhodopsin. *Science* **240**, 777–779 (1988). doi: 10.1126/science.3363359; pmid: 3363359
60. E. F. Pettersen *et al*., UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254

ACKNOWLEDGMENTS

We thank the SwissFEL and particularly R. Abela for their constant logistic and financial support in implementing injector-based serial crystallography. We thank F. Dworkowski for help with auxiliary LCP absorption and scattering experiments. We are grateful to the staff at the X06SA beamline at the Swiss Light Source for their assistance in pretesting crystals. Data collection was carried out at the LCLS at the SLAC National Accelerator Laboratory during the LP41 beamtime in June 2017. **Funding:** Use of the LCLS, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under contract no. DE-AC02-76SF00515. T.T., E.N., and S.I. acknowledge support from the X-ray Free Electron Laser Priority Strategy Program (MEXT) and the Platform Project for Supporting Drug Discovery and Life Science Research from the Japan Agency for Medical Research and Development. Th.W. and A.B. acknowledge funding from the Helmholtz Association via Programme Oriented Funds. I.S. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 678169 "PhotoMutant"). The work was financially supported by grant FP7-PEOPLE-2011-ITN 317079 NanoMem (to G.S. and R.N.) and Horizon2020 grant XPROBE 637295 (to R.N). We further acknowledge the Swiss National Science Foundation for grants 310030_153145 and 310030B_173335 (to G.S.), PZ00P3_174169 (to P.N.), and 31003A_179351 and 31003A_159558 (to J.S.). P.N. acknowledges support from the European Commission's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 290605 (PSI-FELLOW/COFUND). This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 701647. We acknowledge support from the Data Analysis Service (142-004) project of the Swiss Universities SUC-P2 program. R.N. acknowledges funding from the Swedish Research Council (2015–00560 and 349-2011-6485), the Swedish Foundation for Strategic Research (SRL10-0036), and the Knut and Alice Wallenberg Foundation (KAW 2014.0275). G.S. acknowledges support through the NCCR MUST/ETH FAST program of the ETH Zürich. **Author contributions:** P.N., T.W., and J.S. conceived of the research, with suggestions for pump-probe experiments from C.M. and ultrafast biology from R.N. P.N., A.F., W.W., K.J., and V.P. prepared protein and microcrystals. P.N., A.F., K.J., and P.B. prepared samples during data collection. Crystal injection was optimized by P.N. and D.J. The LCP injector with a large reservoir was designed by U.W. and operated by D.J., D.G., and P.S. during data collection. Data processing during the beamtime was performed by T.W., D.O., K.N., and A.B., implementing suggestions from Th.W. Final sorting of data in the femtosecond domain was done by T.W., D.O., and A.B. using code developed by T.L. Progress during the beamtime was recorded by V.P. and R.B. The probe laser was installed and aligned by S.C. The CXI endstation was controlled by M.H., J.K., and M.S. during the beamtime. QM/MM simulations were contributed by V.B., I.S., P.N., and T.W. performed postbeamtime serial data processing, data analysis, and structural refinement. D.K. and S.B. assisted with data interpretation. S.I. and G.S. supported the project. M.H. coordinated efforts at the LCLS, E.N. at SACLA, I.S. at the HUJI, R.N. at the UG, and J.S. at the PSI. The manuscript was written by R.N., P.N., and J.S. with direct contributions from most authors. All authors read and acknowledged the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 6G7H (dark state), 6G7I (Δt = 49 to 406 fs), 6G7J (Δt = 457 to 646 fs), 6G7K (Δt = 10 ps), and 6G7L (Δt = 8.33 ms). Scripts used for correcting the jitter in the subpicosecond time delays are available online (https://gist.github.com/tjlane/1d0c7d7d1f93ec8fcadd53593fbe25b2) or upon request to T.L., SLAC National Accelerator Laboratory. Other software used for data evaluation is freely available as detailed in the methods.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/eaat0094/suppl/DC1
Figs. S1 to S9
Tables S1 to S3
Reference (*61*)
Movies S1 to S3

16 January 2018; accepted 29 May 2018
Published online 14 June 2018
10.1126/science.aat0094

RESEARCH ARTICLE SUMMARY

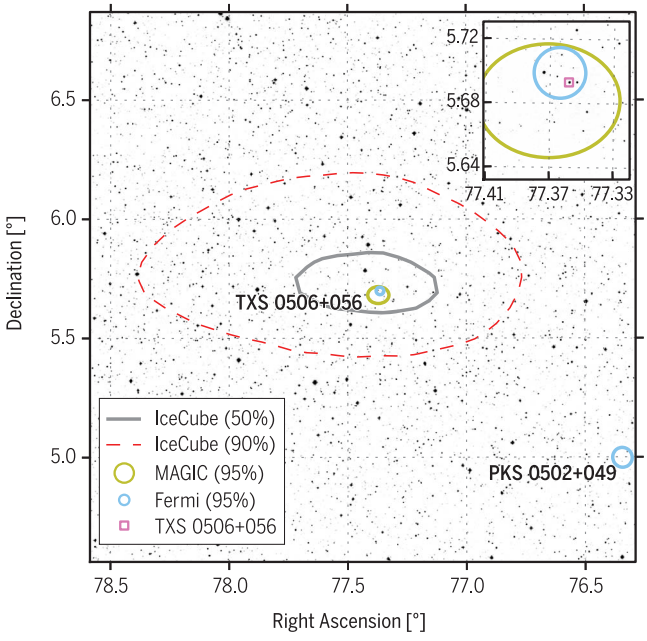
NEUTRINO ASTROPHYSICS

Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A

The IceCube Collaboration, *Fermi*-LAT, MAGIC, *AGILE*, ASAS-SN, HAWC, H.E.S.S., *INTEGRAL*, Kanata, Kiso, Kapteyn, Liverpool Telescope, Subaru, *Swift/NuSTAR*, VERITAS, and VLA/17B-403 teams*†

INTRODUCTION: Neutrinos are tracers of cosmic-ray acceleration: electrically neutral and traveling at nearly the speed of light, they can escape the densest environments and may be traced back to their source of origin. High-energy neutrinos are expected to be produced in blazars: intense extragalactic radio, optical, x-ray, and, in some cases, γ-ray sources characterized by relativistic jets of plasma pointing close to our line of sight. Blazars are among the most powerful objects in the Universe and are widely speculated to be sources of high-energy cosmic rays. These cosmic rays generate high-energy neutrinos and γ-rays, which are produced when the cosmic rays accelerated in the jet interact with nearby gas or photons. On 22 September 2017, the cubic-kilometer IceCube Neutrino Observatory detected a ~290-TeV neutrino from a direction consistent with the flaring γ-ray blazar TXS 0506+056. We report the details of this observation and the results of a multiwavelength follow-up campaign.

RATIONALE: Multimessenger astronomy aims for globally coordinated observations of cosmic rays, neutrinos, gravitational waves, and electromagnetic radiation across a broad range of wavelengths. The combination is expected to yield crucial information on the mechanisms energizing the most powerful astrophysical sources. That the production of neutrinos is accompanied by electromagnetic radiation from the source favors the chances of a multiwavelength identification. In particular, a measured association of high-energy neutrinos with a flaring source of γ-rays would elucidate the mechanisms and conditions for acceleration of the highest-energy cosmic rays. The discovery of an extraterrestrial diffuse flux of high-energy neutrinos, announced by IceCube in 2013, has characteristic properties that hint at contributions from extragalactic sources, although the individual sources remain as yet unidentified. Continuously monitoring the entire sky for astrophysical neutrinos, IceCube provides real-time triggers for observatories around the world measuring γ-rays, x-rays, optical, radio, and gravitational waves, allowing for the potential identification of even rapidly fading sources.



Multimessenger observations of blazar TXS 0506+056. The 50% and 90% containment regions for the neutrino IceCube-170922A (dashed red and solid gray contours, respectively), overlain on a V-band optical image of the sky. Gamma-ray sources in this region previously detected with the *Fermi* spacecraft are shown as blue circles, with sizes representing their 95% positional uncertainty and labeled with the source names. The IceCube neutrino is coincident with the blazar TXS 0506+056, whose optical position is shown by the pink square. The yellow circle shows the 95% positional uncertainty of very-high-energy γ-rays detected by the MAGIC telescopes during the follow-up campaign. The inset shows a magnified view of the region around TXS 0506+056 on an R-band optical image of the sky.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat1378

RESULTS: A high-energy neutrino-induced muon track was detected on 22 September 2017, automatically generating an alert that was distributed worldwide within 1 min of detection and prompted follow-up searches by telescopes over a broad range of wavelengths. On 28 September 2017, the *Fermi* Large Area Telescope Collaboration reported that the direction of the neutrino was coincident with a cataloged γ-ray source, 0.1° from the neutrino direction. The source, a blazar known as TXS 0506+056 at a measured redshift of 0.34, was in a flaring state at the time with enhanced γ-ray activity in the GeV range. Follow-up observations by imaging atmospheric Cherenkov telescopes, notably the Major Atmospheric Gamma Imaging Cherenkov (MAGIC) telescopes, revealed periods where the detected γ-ray flux from the blazar reached energies up to 400 GeV. Measurements of the source have also been completed at x-ray, optical, and radio wavelengths. We have investigated models associating neutrino and γ-ray production and find that correlation of the neutrino with the flare of TXS 0506+056 is statistically significant at the level of 3 standard deviations (sigma). On the basis of the redshift of TXS 0506+056, we derive constraints for the muon-neutrino luminosity for this source and find them to be similar to the luminosity observed in γ-rays.

CONCLUSION: The energies of the γ-rays and the neutrino indicate that blazar jets may accelerate cosmic rays to at least several PeV. The observed association of a high-energy neutrino with a blazar during a period of enhanced γ-ray emission suggests that blazars may indeed be one of the long-sought sources of very-high-energy cosmic rays, and hence responsible for a sizable fraction of the cosmic neutrino flux observed by IceCube. ■

The list of author affiliations is available in the full article online.
*The full lists of participating members for each team and their affiliations are provided in the supplementary materials.
†Email: analysis@icecube.wisc.edu
Cite this article as IceCube Collaboration *et al.*, *Science* **361**, eaat1378 (2018). DOI: 10.1126/science.aat1378

RESEARCH ARTICLE

NEUTRINO ASTROPHYSICS

Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A

The IceCube Collaboration, *Fermi*-LAT, MAGIC, *AGILE*, ASAS-SN, HAWC, H.E.S.S., *INTEGRAL*, Kanata, Kiso, Kapteyn, Liverpool Telescope, Subaru, *Swift/NuSTAR*, VERITAS, and VLA/17B-403 teams*†

Previous detections of individual astrophysical sources of neutrinos are limited to the Sun and the supernova 1987A, whereas the origins of the diffuse flux of high-energy cosmic neutrinos remain unidentified. On 22 September 2017, we detected a high-energy neutrino, IceCube-170922A, with an energy of ~290 tera–electron volts. Its arrival direction was consistent with the location of a known γ-ray blazar, TXS 0506+056, observed to be in a flaring state. An extensive multiwavelength campaign followed, ranging from radio frequencies to γ-rays. These observations characterize the variability and energetics of the blazar and include the detection of TXS 0506+056 in very-high-energy γ-rays. This observation of a neutrino in spatial coincidence with a γ-ray–emitting blazar during an active phase suggests that blazars may be a source of high-energy neutrinos.

Since the discovery of a diffuse flux of high-energy astrophysical neutrinos (*1*, *2*), IceCube has searched for its sources. The only nonterrestrial neutrino sources identified previously are the Sun and the supernova 1987A, producing neutrinos with energies millions of times lower than the high-energy diffuse flux, such that the mechanisms and the environments responsible for the high-energy cosmic neutrinos are still to be ascertained (*3*, *4*). Many candidate source types exist, with active galactic nuclei (AGN) among the most prominent (*5*), in particular the small fraction of them designated as radio-loud (*6*). In these AGNs, the central supermassive black hole converts gravitational energy of accreting matter and/or the rotational energy of the black hole into powerful relativistic jets, within which particles can be accelerated to high energies. If a number of these particles are protons or nuclei, their interactions with the radiation fields and matter close to the source would give rise to a flux of high-energy pions that eventually decay into photons and neutrinos (*7*). In blazars (*8*)—AGNs that have one of the jets pointing close to our line of sight—the observable flux of neutrinos and radiation is expected to be greatly enhanced owing to relativistic Doppler boosting. Blazar electromagnetic (EM) emission is known to be highly variable on time scales from minutes to years (*9*).

Neutrinos travel largely unhindered by matter and radiation. Even if high-energy photons (TeV and above) are unable to escape the source owing to intrinsic absorption, or are absorbed by interactions with the extragalactic background light (EBL) (*10*, *11*), high-energy neutrinos may escape and travel unimpeded to Earth. An association of observed astrophysical neutrinos with blazars would therefore imply that high-energy protons or nuclei up to energies of at least tens of PeV are produced in blazar jets, suggesting that they may be the birthplaces of the most energetic particles observed in the Universe, the ultrahigh-energy cosmic rays (*12*). If neutrinos are produced in correlation with photons, the coincident observation of neutrinos with electromagnetic flares would greatly increase the chances of identifying the source(s). Neutrino detections must therefore be combined with the information from broadband observations across the electromagnetic spectrum (multimessenger observations).

To take advantage of multimessenger opportunities, the IceCube neutrino observatory (*13*) has established a system of real-time alerts that rapidly notify the astronomical community of the direction of astrophysical neutrino candidates (*14*). From the start of the program in April 2016 through October 2017, 10 public alerts have been issued for high-energy neutrino candidate events with well-reconstructed directions (*15*).

We report the detection of a high-energy neutrino by IceCube and the multiwavelength/multi-instrument observations of a flaring γ-ray blazar, TXS 0506+056, which was found to be positionally coincident with the neutrino direction (*16*). Chance coincidence of the IceCube-170922A event with the flare of TXS 0506+056 is statistically disfavored at the level of 3σ in models evaluated below, associating neutrino and γ-ray production.

The neutrino alert

IceCube is a neutrino observatory with more than 5000 optical sensors embedded in 1 km^3 of the Antarctic ice-sheet close to the Amundsen-Scott South Pole Station. The detector consists of 86 vertical strings frozen into the ice 125 m apart, each equipped with 60 digital optical modules (DOMs) at depths between 1450 and 2450 m. When a high-energy muon-neutrino interacts with an atomic nucleus in or close to the detector array, a muon is produced moving through the ice at superluminal speed and creating Cherenkov radiation detected by the DOMs. On 22 September 2017 at 20:54:30.43 Coordinated Universal Time (UTC), a high-energy neutrino-induced muon track event was detected in an automated analysis that is part of IceCube's real-time alert system. An automated alert was distributed (*17*) to observers 43 s later, providing an initial estimate of the direction and energy of the event. A sequence of refined reconstruction algorithms was automatically started at the same time, using the full event information. A representation of this neutrino event with the best-fitting reconstructed direction is shown in Fig. 1. Monitoring data from IceCube indicate that the observatory was functioning normally at the time of the event.

A Gamma-ray Coordinates Network (GCN) Circular (*18*) was issued ~4 hours after the initial notice, including the refined directional information (offset 0.14° from the initial direction; see Fig. 2). Subsequently, further studies were performed to determine the uncertainty of the directional reconstruction arising from statistical and systematic effects, leading to a best-fitting right ascension (RA) $77.43^{+0.95}_{-0.65}$ and declination (Dec) $+5.72^{+0.50}_{-0.30}$ (degrees, J2000 equinox, 90% containment region). The alert was later reported to be in positional coincidence with the known γ-ray blazar TXS 0506+056 (*16*), which is located at RA 77.36° and Dec +5.69° (J2000) (*19*), 0.1° from the arrival direction of the high-energy neutrino.

The IceCube alert prompted a follow-up search by the Mediterranean neutrino telescope ANTARES (Astronomy with a Neutrino Telescope and Abyss environmental RESearch) (*20*). The sensitivity of ANTARES at the declination of IceCube-170922A is about one-tenth that of IceCube's (*21*), and no neutrino candidates were found in a ±1 day period around the event time (*22*).

An energy of 23.7 ± 2.8 TeV was deposited in IceCube by the traversing muon. To estimate the parent neutrino energy, we performed simulations of the response of the detector array, considering that the muon-neutrino might have interacted outside the detector at an unknown distance. We assumed the best-fitting power-law energy spectrum for astrophysical high-energy muon neutrinos, $dN/dE \propto E^{-2.13}$ (*2*), where N is the number of neutrinos as a function of energy E. The simulations yielded a most probable neutrino energy of 290 TeV, with a 90% confidence level (CL)

*The full lists of participating members for each team and their affiliations are provided in the supplementary materials.
†Email: analysis@icecube.wisc.edu

lower limit of 183 TeV, depending only weakly on the assumed astrophysical energy spectrum (*25*).

The vast majority of neutrinos detected by IceCube arise from cosmic-ray interactions within Earth's atmosphere. Although atmospheric neutrinos are dominant at energies below 100 TeV, their spectrum falls steeply with energy, allowing astrophysical neutrinos to be more easily identified at higher energies. The muon-neutrino astrophysical spectrum, together with simulated data, was used to calculate the probability that a neutrino at the observed track energy and zenith angle in IceCube is of astrophysical origin. This probability, the so-called signalness of the event (*14*), was reported to be 56.5% (*17*). Although IceCube can robustly identify astrophysical neutrinos at PeV energies, for individual neutrinos at several hundred TeV, an atmospheric origin cannot be excluded. Electromagnetic observations are valuable to assess the possible association of a single neutrino to an astrophysical source.

Following the alert, IceCube performed a complete analysis of relevant data prior to 31 October 2017. Although no additional excess of neutrinos was found from the direction of TXS 0506+056 near the time of the alert, there are indications at the 3σ level of high-energy neutrino

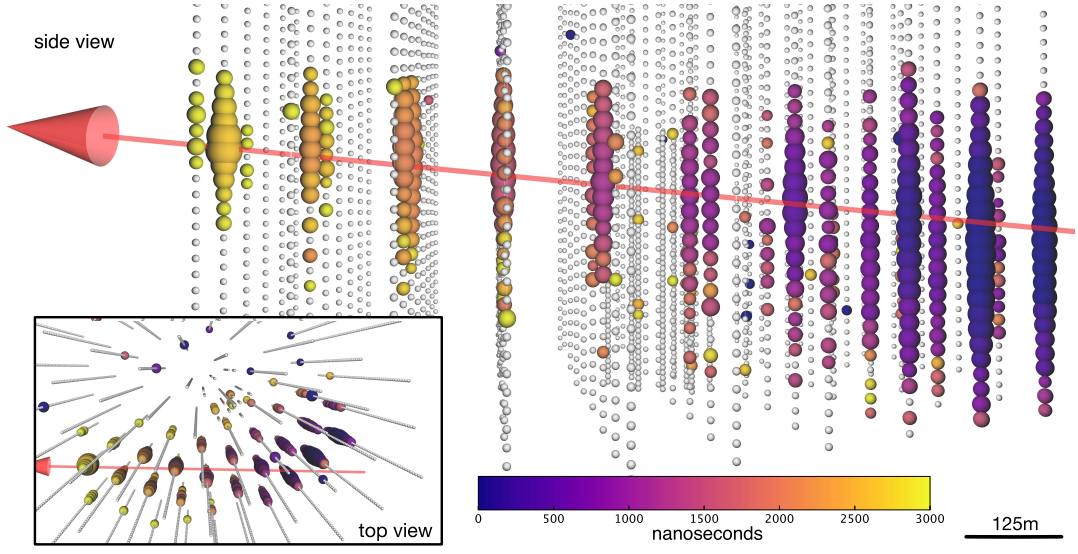


Fig. 1. Event display for neutrino event IceCube-170922A. The time at which a DOM observed a signal is reflected in the color of the hit, with dark blues for earliest hits and yellow for latest. Times shown are relative to the first DOM hit according to the track reconstruction, and earlier and later times are shown with the same colors as the first and last times, respectively. The total time the event took to cross the detector is ~3000 ns. The size of a colored sphere is proportional to the logarithm of the amount of light observed at the DOM, with larger spheres corresponding to larger signals. The total charge recorded is ~5800 photoelectrons. Inset is an overhead perspective view of the event. The best-fitting track direction is shown as an arrow, consistent with a zenith angle $5.7^{+0.50}_{-0.30}$ degrees below the horizon.

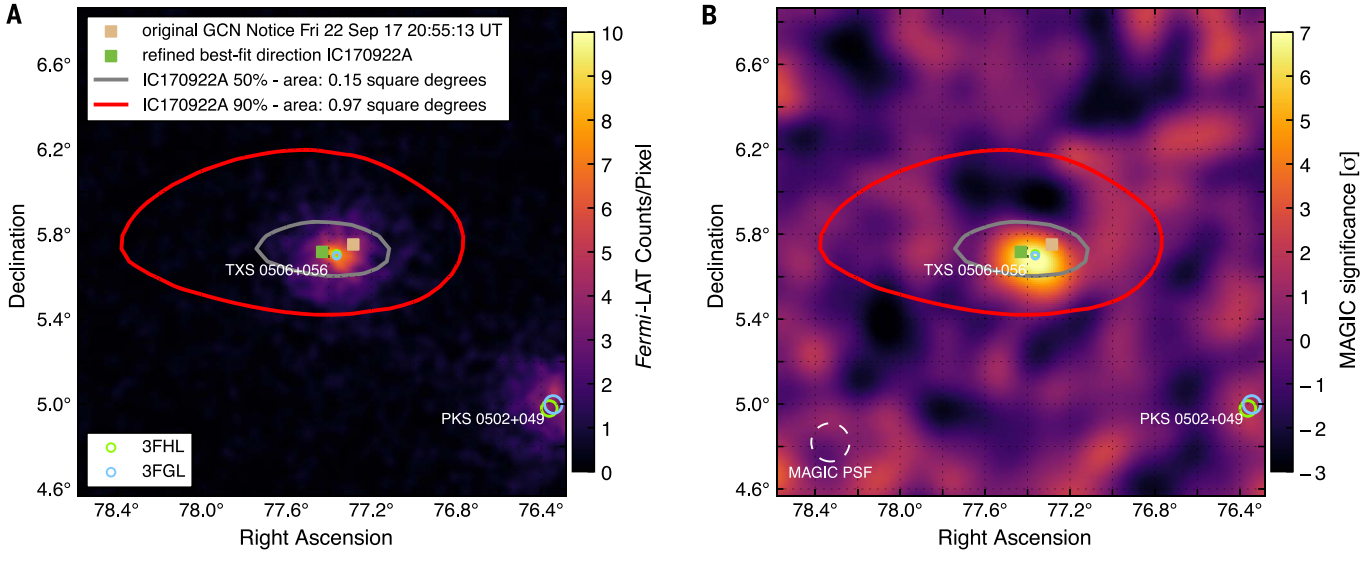


Fig. 2. *Fermi*-LAT and MAGIC observations of IceCube-170922A's location. Sky position of IceCube-170922A in J2000 equatorial coordinates overlaying the γ-ray counts from *Fermi*-LAT above 1 GeV (**A**) and the signal significance as observed by MAGIC (**B**) in this region. The tan square indicates the position reported in the initial alert, and the green square indicates the final best-fitting position from follow-up reconstructions (*18*). Gray and red curves show the 50% and 90% neutrino containment regions, respectively, including statistical and systematic errors. *Fermi*-LAT data are shown as a photon counts map in 9.5 years of data in units of counts per pixel, using detected photons with energy of 1 to 300 GeV in a 2° by 2° region around TXS0506+056. The map has a pixel size of 0.02° and was smoothed with a 0.02°-wide Gaussian kernel. MAGIC data are shown as signal significance for γ-rays above 90 GeV. Also shown are the locations of a γ-ray source observed by *Fermi*-LAT as given in the *Fermi*-LAT Third Source Catalog (3FGL) (*23*) and the Third Catalog of Hard *Fermi*-LAT Sources (3FHL) (*24*) source catalogs, including the identified positionally coincident 3FGL object TXS 0506+056. For *Fermi*-LAT catalog objects, marker sizes indicate the 95% CL positional uncertainty of the source.

emission from that direction in data prior to 2017, as discussed in a companion paper (*26*).

High-energy γ-ray observations of TXS 0506+056

On 28 September 2017, the *Fermi* Large Area Telescope (LAT) Collaboration reported that the direction of origin of IceCube-170922A was consistent with a known γ-ray source in a state of enhanced emission (*16*). *Fermi*-LAT is a pair-conversion telescope aboard the *Fermi Gamma-ray Space Telescope* sensitive to γ-rays with energies from 20 MeV to greater than 300 GeV (*27*). Since August 2008, it has operated continuously, primarily in an all-sky survey mode. Its wide field of view of ~2.4 steradian provides coverage of the entire γ-ray sky every 3 hours. The search for possible counterparts to IceCube-170922A was part of the *Fermi*-LAT collaboration's routine multiwavelength, multimessenger program.

Inside the error region of the neutrino event, a positional coincidence was found with a previously cataloged γ-ray source, 0.1° from the best-fitting neutrino direction. TXS 0506+056 is a blazar of BL Lacertae (BL Lac) type. Its redshift of $z = 0.3365 \pm 0.0010$ was measured only recently based on the optical emission spectrum in a study triggered by the observation of IceCube-170922A (*28*).

TXS 0506+056 is a known *Fermi*-LAT γ-ray source, appearing in three catalogs of *Fermi* sources (*23*, *24*, *29*) at energies above 0.1, 50, and 10 GeV, respectively. An examination of the *Fermi* All-Sky Variability Analysis (FAVA) (*30*) photometric light curve for this object showed that TXS 0506+056 had brightened considerably in the GeV band starting in April 2017 (*16*). Independently, a γ-ray flare was also found by *Fermi*'s Automated Science Processing [ASP (*25*)]. Such flaring is not unusual for a BL Lac object and would not have been followed up as extensively if the neutrino were not detected.

Figure 3 shows the *Fermi*-LAT light curve and the detection time of the neutrino alert. The light curve of TXS 0506+056 from August 2008 to October 2017 was calculated in bins of 28 days for the energy range above 0.1 GeV. An additional light curve with 7-day bins was calculated for the period around the time of the neutrino alert. The γ-ray flux of TXS 0506+056 in each time bin was determined through a simultaneous fit of this source and the other *Fermi*-LAT sources in a 10° by 10° region of interest along with the Galactic and isotropic diffuse backgrounds, using a maximum-likelihood technique (*25*). The integrated γ-ray flux of TXS 0506+056 for $E > 0.1$ GeV, averaged over all *Fermi*-LAT observations spanning 9.5 years, is $(7.6 \pm 0.2) \times 10^{-8}$ cm^{-2} s^{-1}. The highest flux observed in a single 7-day light curve bin was $(5.3 \pm 0.6) \times 10^{-7}$ cm^{-2} s^{-1}, measured in the week 4 to 11 July 2017. Strong flux variations were observed during the γ-ray flare, the most prominent being a flux increase from $(7.9 \pm 2.9) \times 10^{-8}$ cm^{-2} s^{-1} in the week 8 to 15 August 2017 to $(4.0 \pm 0.5) \times 10^{-7}$ cm^{-2} s^{-1} in the week 15 to 22 August 2017.

The Astro-Rivelatore Gamma a Immagini Leggero (*AGILE*) γ-ray telescope (*31*) confirmed the elevated level of γ-ray emission at energies above 0.1 GeV from TXS 0506+056 in a 13-day window (10 to 23 September 2017). The *AGILE* measured flux of $(5.3 \pm 2.1) \times 10^{-7}$ cm^{-2} s^{-1} is consistent with the *Fermi*-LAT observations in this time period.

High-energy γ-ray observations are shown in Figs. 3 and 4. Details on the *Fermi*-LAT and *AGILE* analyses can be found in (*25*).

Very-high-energy γ-ray observations of TXS 0506+056

Following the announcement of IceCube-170922A, TXS 0506+056 was observed by several ground-based Imaging Atmospheric Cherenkov Telescopes (IACTs). A total of 1.3 hours of observations in the direction of the blazar TXS 0506+056 were taken using the High-Energy Stereoscopic System (H.E.S.S.) (*32*), located in Namibia, on 23 September 2017 [Modified Julian Date (MJD)

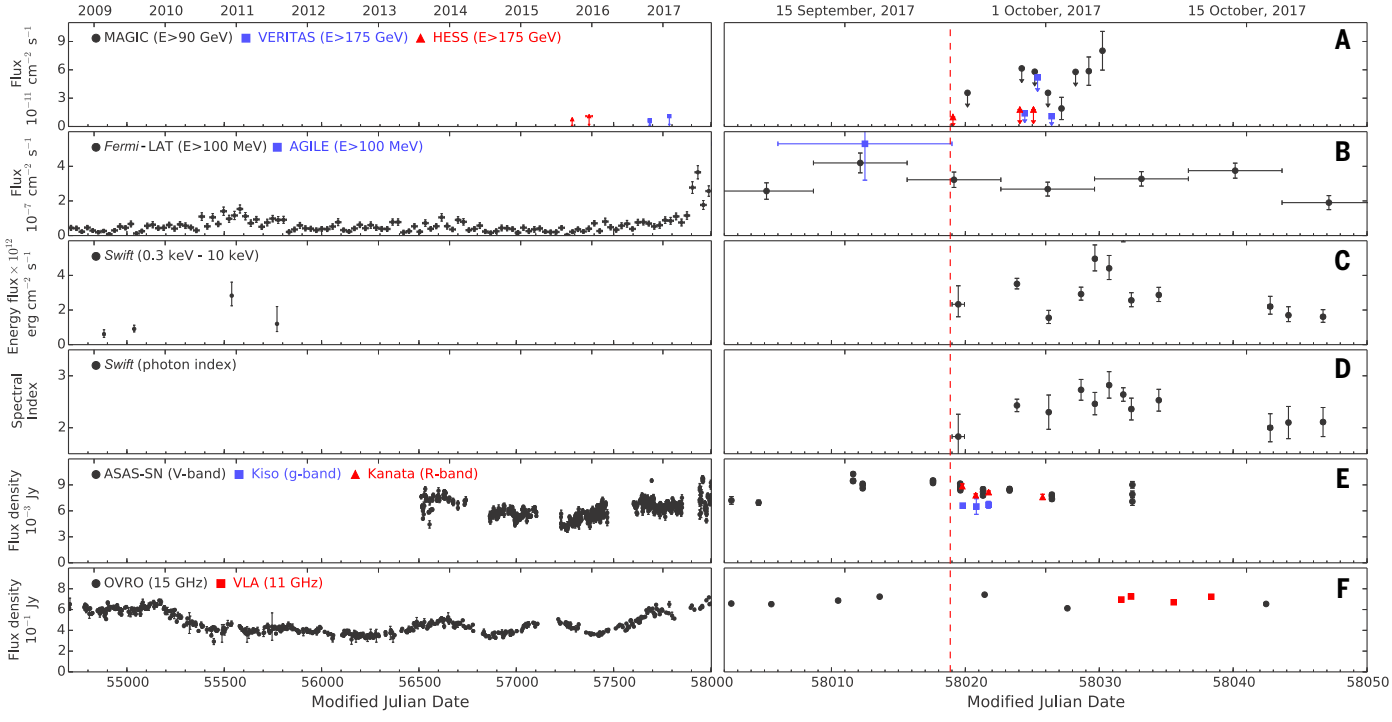


Fig. 3. Time-dependent multiwavelength observations of TXS 0506+056 before and after IceCube-170922A. Significant variability of the electromagnetic emission can be observed in all displayed energy bands, with the source being in a high-emission state around the time of the neutrino alert. From top to bottom: (**A**) VHE γ-ray observations by MAGIC, H.E.S.S., and VERITAS; (**B**) high-energy γ-ray observations by *Fermi*-LAT and *AGILE*; (**C** and **D**) x-ray observations by *Swift* XRT; (**E**) optical light curves from ASAS-SN, Kiso/KWFC, and Kanata/HONIR; and (**F**) radio observations by OVRO and VLA. The red dashed line marks the detection time of the neutrino IceCube-170922A. The left set of panels shows measurements between MJD 54700 (22 August 2008) and MJD 58002 (6 September 2017). The set of panels on the right shows an expanded scale for time range MJD 58002 to MJD 58050 (24 October 2017). The *Fermi*-LAT light curve is binned in 28-day bins on the left panel, while finer 7-day bins are used on the expanded panel. A VERITAS limit from MJD 58019.40 (23 September 2017) of 2.1×10^{-10} cm^{-2} s^{-1} is off the scale of the plot and not shown.

58019], ~4 hours after the circulation of the neutrino alert. A 1-hour follow-up observation of the neutrino alert under partial cloud coverage was performed using the Very Energetic Radiation Imaging Telescope Array System (VERITAS) γ-ray telescope array (*33*), located in Arizona, USA, later on the same day, ~12 hours after the IceCube detection. Both telescopes made additional observations on subsequent nights, but neither detected γ-ray emission from the source [see Fig. 3 and (*25*)]. Upper limits at 95% CL on the γ-ray flux were derived accordingly (assuming the measured spectrum, see below): 7.5×10^{-12} cm^{-2} s^{-1} during the H.E.S.S. observation period and 1.2×10^{-11} cm^{-2} s^{-1} during the VERITAS observations, both for energies E >175 GeV.

The Major Atmospheric Gamma Imaging Cherenkov (MAGIC) Telescopes (*34*) observed TXS 0506+056 for 2 hours on 24 September 2017 (MJD 58020) under nonoptimal weather conditions and then for a period of 13 hours from 28 September to 4 October 2017 (MJD 58024–58030) under good conditions. MAGIC consists of two 17-m telescopes, located at the Roque de los Muchachos Observatory on the Canary Island of La Palma (Spain).

No γ-ray emission from TXS 0506+056 was detected in the initial MAGIC observations on 24 September 2017, and an upper limit was derived on the flux above 90 GeV of 3.6×10^{-11} cm^{-2} s^{-1} at 95% CL (assuming a spectrum $dN/dE \propto E^{-3.9}$). However, prompted by the *Fermi*-LAT detection of enhanced γ-ray emission, MAGIC performed another 13 hours of observations of the region starting 28 September 2017. Integrating the data, MAGIC detected a significant very-high-energy (VHE) γ-ray signal (*35*) corresponding to 374 ± 62 excess photons, with observed energies up to about 400 GeV. This represents a 6.2σ excess over expected background levels (*25*). The day-by-day light curve of TXS 0506+056 for energies above 90 GeV is shown in Fig. 3. The probability that a constant flux is consistent with the data is less than 1.35%. The measured differential photon spectrum (Fig. 4) can be described over the energy range of 80 to 400 GeV by a simple power law, $dN/dE \propto E^{\gamma}$, with a spectral index $\gamma = -3.9 \pm 0.4$ and a flux normalization of $(2.0 \pm 0.4) \times 10^{-10}$ TeV^{-1} cm^{-2} s^{-1} at E = 130 GeV. Uncertainties are statistical only. The estimated systematic uncertainties are <15% in the energy scale, 11 to 18% in the flux normalization, and ±0.15 for the power-law slope of the energy spectrum (*34*). Further observations after 4 October 2017 were prevented by the full Moon.

An upper limit to the redshift of TXS 0506+056 can be inferred from VHE γ-ray observations using limits on the attenuation of the VHE flux due to interaction with the EBL. Details on the method are available in (*25*). The obtained upper limit ranges from 0.61 to 0.98 at a 95% CL, depending on the EBL model used. These upper limits are consistent with the measured redshift of $z = 0.3365$ (*28*).

No γ-ray source above 1 TeV at the location of TXS 0506+056 was found in survey data of the High Altitude Water Cherenkov (HAWC) γ-ray observatory (*36*), either close to the time of the neutrino alert or in archival data taken since November 2014 (*25*).

VHE γ-ray observations are shown in Figs. 3 and 4. All measurements are consistent with the observed flux from MAGIC, considering the differences in exposure, energy range, and observation periods.

Radio, optical, and x-ray observations

The Karl G. Jansky Very Large Array (VLA) (*37*) observed TXS 0506+056 starting 2 weeks after the alert in several radio bands from 2 to 12 GHz (*38*), detecting significant radio flux variability and some spectral variability of this source. The source is also in the long-term blazar monitoring program of the Owens Valley Radio Observatory (OVRO) 40-m telescope at 15 GHz (*39*). The light curve shows a gradual increase in radio emission during the 18 months preceding the neutrino alert.

Optical observations were performed by the All-Sky Automated Survey for Supernovae (ASAS-SN) (*40*), the Liverpool Telescope (*41*), the

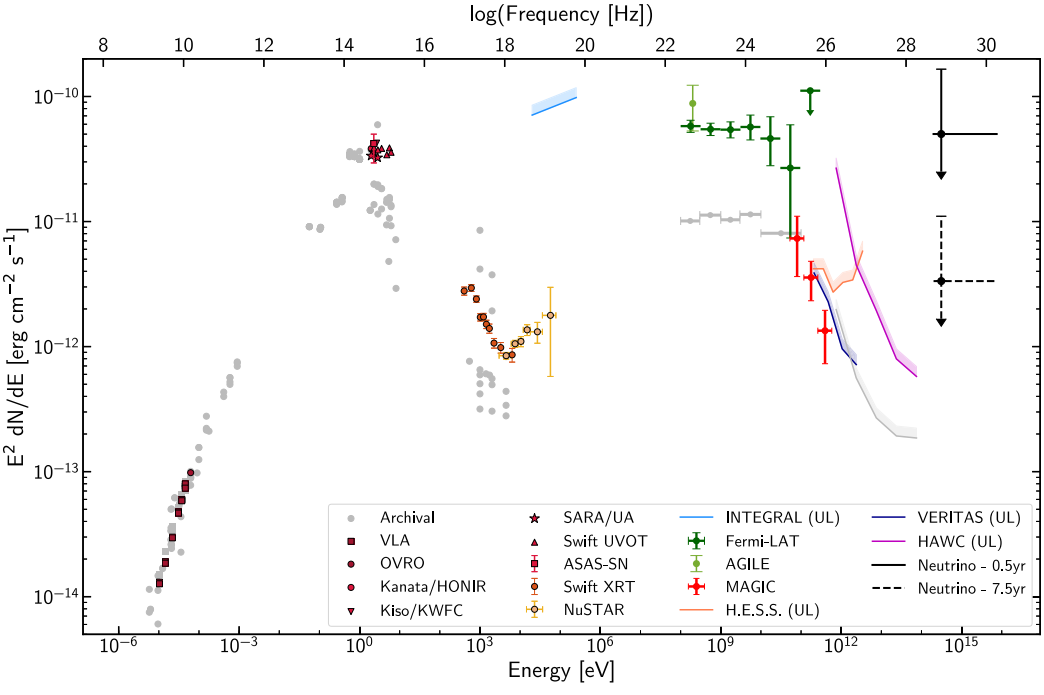


Fig. 4. Broadband spectral energy distribution for the blazar TXS 0506+056. The SED is based on observations obtained within 14 days of the detection of the IceCube-170922A event. The $E^2 dN/dE$ vertical axis is equivalent to a νF_ν scale. Contributions are provided by the following instruments: VLA (*38*), OVRO (*39*), Kanata Hiroshima Optical and Near-InfraRed camera (HONIR) (*52*), Kiso, and the Kiso Wide Field Camera (KWFC) (*43*), Southeastern Association for Research in Astronomy Observatory (SARA/UA) (*53*), ASAS-SN (*54*), *Swift* Ultraviolet and Optical Telescope (UVOT) and XRT (*55*), *NuSTAR* (*56*), *INTEGRAL* (*57*), *AGILE* (*58*), *Fermi*-LAT (*16*), MAGIC (*35*), VERITAS (*59*), H.E.S.S. (*60*), and HAWC (*61*). Specific observation dates and times are provided in (*25*). Differential flux upper limits (shown as colored bands and indicated as "UL" in the legend) are quoted at the 95% CL, while markers indicate significant detections. Archival observations are shown in gray to illustrate the historical flux level of the blazar in the radio-to-keV range as retrieved from the ASDC SED Builder (*62*), and in the γ-ray band as listed in the *Fermi*-LAT 3FGL catalog (*23*) and from an analysis of 2.5 years of HAWC data. The γ-ray observations have not been corrected for absorption owing to the EBL. SARA/UA, ASAS-SN, and Kiso/KWFC observations have not been corrected for Galactic attenuation. The electromagnetic SED displays a double-bump structure, one peaking in the optical-ultraviolet range and the second one in the GeV range, which is characteristic of the nonthermal emission from blazars. Even within this 14-day period, there is variability observed in several of the energy bands shown (see Fig. 3), and the data are not all obtained simultaneously. Representative $\nu_\mu + \bar{\nu}_\mu$ neutrino flux upper limits that produce on average one detection like IceCube-170922A over a period of 0.5 (solid black line) and 7.5 years (dashed black line) are shown, assuming a spectrum of $dN/dE \propto E^{-2}$ at the most probable neutrino energy (311 TeV).

Kanata Telescope (42), the Kiso Schmidt Telescope (43), the high-resolution spectrograph (HRS) on the Southern African Large Telescope (SALT) (44), the Subaru telescope Faint Object Camera and Spectrograph (FOCAS) (45), and the X-SHOOTER instrument on the Very Large Telescope (VLT) (46). The *V* band flux of the source is the highest observed in recent years, and the spectral energy distribution has shifted toward blue wavelengths. Polarization was detected by Kanata in the *R* band at the level of 7%. Redshift determination for BL Lac objects is difficult owing to the nonthermal continuum from the nucleus outshining the spectral lines from the host galaxies. Attempts were made using optical spectra from the Liverpool, Subaru, and VLT telescopes to measure the redshift of TXS 0506+056, but only limits could be derived [see, e.g., (47)]. The redshift of TXS 0506+056 was later determined to be $z = 0.3365 \pm 0.0010$ using the Gran Telescopio Canarias (28).

X-ray observations were made by the X-Ray Telescope (XRT) on the *Neil Gehrels Swift Observatory* (0.3 to 10 keV) (48), *MAXI* Gas Slit Camera (GSC) (2 to 10 keV) (49), Nuclear Spectroscopic Telescope Array (*NuSTAR*) (3 to 79 keV) (50), and the INTernational Gamma-Ray Astrophysics Laboratory (*INTEGRAL*) (20 to 250 keV) (51), with detections by *Swift* and *NuSTAR* . In a 2.1 square degree region around the neutrino alert, *Swift* identified nine x-ray sources, including TXS 0506+056.

Swift monitored the x-ray flux from TXS 0506+056 for 4 weeks after the alert, starting 23 September 2017 00:09:16 UT, finding clear evidence for spectral variability (see Fig. 3D). The strong increase in flux observed at VHE energies over several days up until MJD 58030 (4 October 2017) correlates well with an increase in the x-ray emission during this period of time. The spectrum of TXS 0506+056 observed in the week after the flare is compatible with the sum of two power-law spectra, a soft spectrum with index -2.8 ± 0.3 in the soft x-ray band covered by *Swift* XRT, and a hard spectrum with index -1.4 ± 0.3 in the hard x-ray band covered by *NuSTAR* (25). Extrapolated to 20 MeV, the *NuSTAR* hard-spectrum component connects smoothly to the plateau (index -2) component observed by the *Fermi*-LAT between 0.1 and 100 GeV and the soft VHE γ-ray component observed by MAGIC (compare Fig. 4). Taken together, these observations provide a mostly complete, contemporaneous picture of the source emissions from 0.3 keV to 400 GeV, more than nine orders of magnitude in photon energy.

Figures 3 and 4 summarize the multiwavelength light curves and the changes in the broadband spectral energy distribution (SED), compared to archival observations. Additional details about the radio, optical, and x-ray observations can be found in (25).

Chance coincidence probability

Data obtained from multiwavelength observations of TXS 0506+056 can be used to constrain the blazar-neutrino chance coincidence probability. This coincidence probability is a measure of the likelihood that a neutrino alert like IceCube-170922A is correlated by chance with a flaring blazar, considering the large number of known γ-ray sources and the modest number of neutrino alerts.

Given the large number of potential neutrino source classes available, no a priori blazar-neutrino coincidence model was selected ahead of the alert. After the observation, however, several correlation scenarios were considered and tested to quantify the a posteriori significance of the observed coincidence. Testing multiple models is important as the specific assumptions about the correlation between neutrinos and γ-rays have an impact on the chance coincidence probability. In each case, the probability to obtain, by chance, a degree of correlation at least as high as that observed for IceCube-170922A was calculated using simulated neutrino alerts and the light curves of *Fermi*-LAT γ-ray sources. Given the continuous all-sky monitoring of the *Fermi*-LAT since 2008, all tests utilized 28-day binned γ-ray light curves above 1 GeV from 2257 extragalactic *Fermi*-LAT sources, derived in the same manner as used for the analysis of TXS 0506+056 γ-ray data.

To calculate the chance probabilities, a likelihood ratio test is used that allows different models of blazar-neutrino flux correlation to be evaluated in a consistent manner. All models assume that at least some of the observed γ-ray flux is produced in the same hadronic interactions that would produce high-energy neutrinos within the source. Our first model assumes that the neutrino flux is linearly correlated with the high-energy γ-ray energy flux (4). In this scenario, neutrinos are more likely to be produced during periods of bright, hard γ-ray emission. In the second model, the neutrino flux is modeled as strongly tied to variations in the observed γ-ray flux, regardless of the average flux of γ-rays. Here, a weak or a strong γ-ray source is equally likely to be a neutrino source if the neutrino is temporally correlated with variability in the γ-ray light curve. Third, we consider a correlation of the neutrino flux with the VHE γ-ray flux. Because hadronic acceleration up to a few PeV is required to explain the detected neutrino energy, VHE γ-ray sources are potential progenitors. Full details and results from these analyses are presented in (25).

The neutrino IceCube-170922A was found to arrive in a period of flaring activity in high-energy γ-rays. Prior to IceCube-170922A, nine public alerts had been issued by the IceCube real-time system. Additionally, 41 archival events have been identified among the IceCube data recorded since 2010, before the start of the real-time program in April 2016, which would have caused alerts if the real-time alert system had been in place. These events were also tested for coincidence with the γ-ray data.

Chance coincidence of the neutrino with the flare of TXS 0506+056 is disfavored at the 3σ level in any scenario where neutrino production is linearly correlated with γ-ray production or with γ-ray flux variations. This includes look-elsewhere corrections for all 10 alerts issued previously by IceCube and the 41 archival events. One of the neutrino events that would have been sent as an alert and had a good angular resolution (<5°) is in a spatial correlation with the γ-ray blazar 3FGL J1040.4+0615. However, this source was not in a particularly bright emission state at the detection time of the corresponding neutrino. Therefore, a substantially lower test statistic would be obtained in the chance correlation tests defined in this paper (25).

We have investigated how typical the blazar TXS 0506+056 is among those blazars that might have given rise to a coincident observation similar to the one reported here. A simulation that assumes that the neutrino flux is linearly correlated with the blazar γ-ray energy flux shows that in 14% of the signal realizations, we would find a neutrino coincident with a similarly bright γ-ray state as that observed for TXS 0506+056 (25). The detection of a single neutrino does not allow us to probe the details of neutrino production models or measure the neutrino–to–γ-ray production ratio. Further observations will be needed to unambiguously establish a correlation between high-energy neutrinos and blazars, as well as to understand the emission and acceleration mechanism in the event of a correlation.

Discussion

Blazars have often been suggested as potential sources of high-energy neutrinos. The calorimetric high-energy output of certain candidate blazars is high enough to explain individual observed IceCube events at 100-TeV to 1-PeV energies (63). Spatial coincidences between catalogs of blazars and neutrinos have been examined in (64), while (65) investigated one shower-like event with several thousand square degrees angular uncertainty observed in time coincidence with a blazar outburst. A track-like event, IceCube-160731, has been previously connected to a flaring γ-ray source (66). However, the limited evidence for a flaring source in the multiwavelength coverage did not permit an identification of the source type of the potential counterpart (66).

Owing to the precise direction of IceCube-170922A, combined with extensive multiwavelength observations, a chance correlation between a high-energy neutrino and the potential counterpart can be rejected at the 3σ level. Considering the association between IceCube-170922A and TXS 0506+056, γ-ray blazars are strong candidate sources for at least a fraction of the observed astrophysical neutrinos. Earlier studies of the cross-correlation between IceCube events and the γ-ray blazar population observed by *Fermi*-LAT demonstrated that these blazars can only produce a fraction of the observed astrophysical neutrino flux above 10 TeV (4). Although these limits constrain the contribution from blazars to the diffuse neutrino background, the potential association of one or two high-energy neutrinos to blazars over the total observing time of IceCube is fully compatible with the constraint.

Adopting standard cosmological parameters (67) $H_0 = 67.8, \Omega_m = 0.308, \Omega_\lambda = 0.692$, where H_0 is the Hubble constant, Ω_m is the matter

density, and Ω_λ is the dark energy density, the observed redshift of $z = 0.3365$ implies an isotropic γ-ray luminosity between 0.1 and 100 GeV of 1.3×10^{47} erg s^{-1} in the ±2 weeks around the arrival time of the IceCube neutrino, and a luminosity of 2.8×10^{46} erg s^{-1}, averaged over all *Fermi*-LAT observations. Observations in the optical, x-ray, and VHE γ-ray bands show typical characteristics of blazar flares: strong variability on time scales of a few days and an indication of a shift of the synchrotron emission peak toward higher frequencies. VHE γ-ray emission is found to change by a factor of ~4 within just 3 days. Similarly, the high-energy γ-ray energy band shows flux variations up to a factor of ~5 from one week to the next.

No other neutrino event that would have passed the selection criteria for a high-energy alert was observed from this source since the start of IceCube observations in May 2010. The muon neutrino fluence for which we would expect to detect one high-energy alert event with IceCube in this period of time is 2.8×10^{-3} erg cm^{-2}. A power-law neutrino spectrum is assumed in this calculation with an index of −2 between 200 TeV and 7.5 PeV, the range between the 90% CL lower and upper limits for the energy of the observed neutrino [see (*25*) for details].

The fluence can be expressed as an integrated energy flux if we assume a time period during which the source was emitting neutrinos. For a source that emits neutrinos only during the ~6-month period corresponding to the duration of the high-energy γ-ray flare, the corresponding average integrated muon neutrino energy flux would be 1.8×10^{-10} erg cm^{-2} s^{-1}. Alternatively, the average integrated energy flux of a source that emits neutrinos over the whole observation period of IceCube (i.e., 7.5 years) would be 1.2×10^{-11} erg cm^{-2} s^{-1}. These two benchmark cases are displayed in Fig. 4. In an ensemble of faint sources with a summed expectation of order 1, we would anticipate observing a neutrino even if the individual expectation value is ≪1. This is expressed by the downward arrows on the neutrino flux points in Fig. 4.

The two cases discussed above correspond to average isotropic muon neutrino luminosities of 7.2×10^{46} erg s^{-1} for a source that was emitting neutrinos in the ~6-month period of the high-energy γ-ray flare, and 4.8×10^{45} erg s^{-1} for a source that emitted neutrinos throughout the whole observation period. This is similar to the luminosity observed in γ-rays and thus broadly consistent with hadronic source scenarios (*68*).

A neutrino flux that produces a high-energy alert event can, over time, produce many lower-energy neutrino-induced muons in IceCube. A study of neutrino emission from TXS 0506+056 prior to the high-energy γ-ray flare, based on the investigation of these lower-energy events, is reported in a companion paper (*26*).

REFERENCES AND NOTES

1. M. G. Aartsen *et al*., Evidence for high-energy extraterrestrial neutrinos at the IceCube detector. *Science* **342**, 1242856 (2013). doi: 10.1126/science.1242856; pmid: 24264993
2. M. G. Aartsen *et al*., Observation and characterization of a cosmic muon neutrino flux from the Northern Hemisphere using six years of IceCube data. *Astrophys. J.* **833**, 3 (2016). doi: 10.3847/0004-637X/833/1/3
3. M. G. Aartsen *et al*., All-sky search for time-integrated neutrino emission from astrophysical sources with 7 yr of IceCube data. *Astrophys. J.* **835**, 151 (2017). doi: 10.3847/1538-4357/835/2/151
4. M. G. Aartsen *et al*., the contribution of *FERMI* -2LAC blazars to diffuse TeV–PeV neutrino flux. *Astrophys. J.* **835**, 45 (2017). doi: 10.3847/1538-4357/835/1/45
5. F. W. Stecker, C. Done, M. H. Salamon, P. Sommers, High-energy neutrinos from active galactic nuclei. *Phys. Rev. Lett.* **66**, 2697–2700 (1991). doi: 10.1103/PhysRevLett.66.2697; pmid: 10043593
6. K. Mannheim, High-energy neutrinos from extragalactic jets. *Astropart. Phys.* **3**, 295–302 (1995). doi: 10.1016/0927-6505(94)00044-4
7. M. Petropoulou, S. Dimitrakoudis, P. Padovani, A. Mastichiadis, E. Resconi, Photohadronic origin of γ-ray BL Lac emission: Implications for IceCube neutrinos. *Mon. Not. R. Astron. Soc.* **448**, 2412–2429 (2015). doi: 10.1093/mnras/stv179
8. C. M. Urry, P. Padovani, Unified schemes for radio-loud active galactic nuclei. *Publ. Astron. Soc. Pac.* **107**, 803 (1995). doi: 10.1086/133630
9. M.-H. Ulrich, L. Maraschi, C. M. Urry, Variability of active galactic nuclei. *Annu. Rev. Astron. Astrophys.* **35**, 445–502 (1997). doi: 10.1146/annurev.astro.35.1.445
10. M. G. Hauser, E. Dwek, The cosmic infrared background: Measurements and implications. *Annu. Rev. Astron. Astrophys.* **39**, 249–307 (2001). doi: 10.1146/annurev.astro.39.1.249
11. F. W. Stecker, O. C. de Jager, M. H. Salamon, TeV gamma rays from 3C 279 - A possible probe of origin and intergalactic infrared radiation fields. *Astrophys. J.* **390**, L49 (1992). doi: 10.1086/186369
12. K. A. Olive *et al*., Review of particle physics. *Chin. Phys. (Beijing) C* **38**, 090001 (2014). doi: 10.1088/1674-1137/38/9/090001
13. M. G. Aartsen *et al*., The IceCube Neutrino Observatory: Instrumentation and online systems. *J. Instrum.* **12**, P03012 (2017). doi: 10.1088/1748-0221/12/03/P03012
14. M. G. Aartsen *et al*., The IceCube realtime alert system. *Astropart. Phys.* **92**, 30–41 (2017). doi: 10.1016/j.astropartphys.2017.05.002
15. GCN/AMON Notices, https://gcn.gsfc.nasa.gov/amon.html; accessed: 26 April 2018.
16. Y. T. Tanaka, S. Buson, D. Kocevski, *The Astronomer's Telegram* **10791** (2017).
17. IceCube Collaboration, *GRB Coordinates Network/AMON Notices* 50579430_130033 (2017).
18. IceCube Collaboration, *GRB Coordinates Network, Circular Service* **21916** (2017).
19. G. E. Lanyi *et al*., The celestial reference frame at 24 and 43 GHz. I. Astrometry. *Astron. J.* **139**, 1695–1712 (2010). doi: 10.1088/0004-6256/139/5/1695
20. M. Ageron *et al*., ANTARES: The first undersea neutrino telescope. *Nucl. Instrum. Methods Phys. Res. A* **656**, 11–38 (2011). doi: 10.1016/j.nima.2011.06.103
21. A. Albert *et al*., First all-flavor neutrino pointlike source search with the ANTARES neutrino telescope. *Phys. Rev. D* **96**, 082001 (2017). doi: 10.1103/PhysRevD.96.082001
22. D. Dornic, A. Coleiro, *The Astronomer's Telegram* **10773** (2017).
23. F. Acero *et al*., *FERMI* Large Area Telescope Third Source Catalog. *Astrophys. J.* **218** (suppl.), 23 (2015). doi: 10.1088/0067-0049/218/2/23
24. M. Ajello *et al*., 3FHL: The Third Catalog of Hard *Fermi* -LAT Sources. *Astrophys. J.* **232** (suppl.), 18 (2017). doi: 10.3847/1538-4365/aa8221
25. Materials and methods are available as supplementary materials.
26. IceCube Collaboration, Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert. *Science* **361**, 147–151 (2018).
27. W. B. Atwood *et al*., The Large Area Telescope on the *FERMI Gamma-Ray Space Telescope* mission. *Astrophys. J.* **697**, 1071–1102 (2009). doi: 10.1088/0004-637X/697/2/1071
28. S. Paiano, R. Falomo, A. Treves, R. Scarpa, The redshift of the BL Lac object TXS 0506+056. *Astrophys. J.* **854**, L32 (2018). doi: 10.3847/2041-8213/aaad5e
29. M. Ackermann *et al*., 2FHL: The second catalog of hard *FERMI* -LAT sources. *Astrophys. J.* **222** (suppl.), 5 (2016). doi: 10.3847/0067-0049/222/1/5
30. S. Abdollahi *et al*., The second catalog of flaring gamma-ray sources from the Fermi All-sky Variability Analysis. *Astrophys. J.* **846**, 34 (2017). doi: 10.3847/1538-4357/aa8092
31. M. Tavani *et al*., The *AGILE* mission. *Astron. Astrophys.* **502**, 995–1013 (2009). doi: 10.1051/0004-6361/200810527
32. F. Aharonian *et al*., Observations of the Crab Nebula with HESS. *Astron. Astrophys.* **457**, 899–915 (2006). doi: 10.1051/0004-6361:20065351
33. J. Holder *et al*., The first VERITAS telescope. *Astropart. Phys.* **25**, 391–401 (2006). doi: 10.1016/j.astropartphys.2006.04.002
34. J. Aleksić *et al*., The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula. *Astropart. Phys.* **72**, 76–94 (2016). doi: 10.1016/j.astropartphys.2015.02.005
35. R. Mirzoyan, *The Astronomer's Telegram* **10817** (2017).
36. A. U. Abeysekara *et al*., Observation of the Crab Nebula with the HAWC Gamma-Ray Observatory. *Astrophys. J.* **843**, 39 (2017). doi: 10.3847/1538-4357/aa7555
37. R. A. Perley, C. J. Chandler, B. J. Butler, J. M. Wrobel, The expanded Very Large Array: A new telescope for new science. *Astrophys. J.* **739**, L1 (2011). doi: 10.1088/2041-8205/739/1/L1
38. A. J. Tetarenko, G. R. Sivakoff, A. E. Kimball, J. C. A. Miller-Jones, *The Astronomer's Telegram* **10861** (2017).
39. J. L. Richards *et al*., Blazars in the *FERMI* Era: The OVRO 40 m telescope monitoring program. *Astrophys. J.* **194** (suppl.), 29 (2011). doi: 10.1088/0067-0049/194/2/29
40. C. S. Kochanek *et al*., The All-Sky Automated Survey for Supernovae (ASAS-SN) Light Curve Server v1.0. *Publ. Astron. Soc. Pac.* **129**, 104502 (2017). doi: 10.1088/1538-3873/aa80d9
41. I. A. Steele *et al*., *Ground-based Telescopes*, J. M. Oschmann Jr., Ed. (2004), vol. 5489 of *Proc. SPIE*, pp. 679–692.
42. H. Akitaya *et al*., *Ground-based and Airborne Instrumentation for Astronomy V* (2014), vol. 9147 of *Proc. SPIE*, p. 91474O.
43. S. Sako *et al*., *Ground-based and Airborne Instrumentation for Astronomy IV* (2012), vol. 8446 of *Proc. SPIE*, p. 84466L.
44. L. A. Crause *et al*., *Ground-based and Airborne Instrumentation for Astronomy V* (2014), vol. 9147 of *Proc. SPIE*, p. 91476T.
45. N. Kashikawa *et al*., FOCAS: The Faint Object Camera and Spectrograph for the Subaru Telescope. *Publ. Astron. Soc. Jpn.* **54**, 819–832 (2002). doi: 10.1093/pasj/54.6.819
46. J. Vernet *et al*., X-shooter, the new wide band intermediate resolution spectrograph at the ESO Very Large Telescope. *Astron. Astrophys.* **536**, A105 (2011). doi: 10.1051/0004-6361/201117752
47. A. Coleiro, S. Chaty, *The Astronomer's Telegram* **10840** (2017).
48. D. N. Burrows *et al*., The Swift X-Ray Telescope. *Space Sci. Rev.* **120**, 165–195 (2005). doi: 10.1007/s11214-005-5097-2
49. M. Matsuoka *et al*., The MAXI mission on the ISS: Science and instruments for monitoring All-Sky X-Ray Images. *Publ. Astron. Soc. Jpn.* **61**, 999–1010 (2009). doi: 10.1093/pasj/61.5.999
50. F. A. Harrison *et al*., The *Nuclear Spectroscopic Telescope Array* (*NuSTAR*) high-energy x-ray mission. *Astrophys. J.* **770**, 103 (2013). doi: 10.1088/0004-637X/770/2/103
51. C. Winkler *et al*., The INTEGRAL mission. *Astron. Astrophys.* **411**, L1–L6 (2003). doi: 10.1051/0004-6361:20031288
52. M. Yamanaka, *et al*., *The Astronomer's Telegram* **10844** (2017).
53. W. Keel, M. Santander, *The Astronomer's Telegram* **10831** (2017).
54. A. Franckowiak, *et al*., *The Astronomer's Telegram* **10794** (2017).
55. A. Keivani, *et al*., *The Astronomer's Telegram* **10792** (2017).
56. D. B. Fox, *et al*., *The Astronomer's Telegram* **10845** (2017).
57. V. Savchenko *et al*., *GRB Coordinates Network, Circular Service* **21917** (2017).
58. F. Lucarelli, *et al*., *The Astronomer's Telegram* **10801** (2017).
59. R. Mukherjee, *The Astronomer's Telegram* **10833** (2017).
60. M. de Naurois, H.E.S.S. Collaboration, *The Astronomer's Telegram* **10787** (2017).
61. I. Martinez, I. Taboada, M. Hui, R. Lauer, *The Astronomer's Telegram* **10802** (2017).
62. G. Stratta *et al*., The ASDC SED Builder Tool description and tutorial. arXiv:1103.0749 [atro-ph.IM] (3 March 2011).
63. F. Krauß *et al*., TANAMI blazars in the IceCube PeV-neutrino fields. *Astron. Astrophys.* **566**, L7 (2014). doi: 10.1051/0004-6361/201424219
64. P. Padovani, E. Resconi, P. Giommi, B. Arsioli, Y. L. Chang, Extreme blazars as counterparts of IceCube astrophysical neutrinos. *Mon. Not. R. Astron. Soc.* **457**, 3582–3592 (2016). doi: 10.1093/mnras/stw228

65. M. Kadler *et al*., Coincidence of a high-fluence blazar outburst with a PeV-energy neutrino event. *Nat. Phys.* **12**, 807–814 (2016). doi: 10.1038/nphys3715
66. F. Lucarelli *et al*., *AGILE* detection of a candidate gamma-ray precursor to the ICECUBE-160731 neutrino event. *Astrophys. J.* **846**, 121 (2017). doi: 10.3847/1538-4357/aa81c8
67. P. A. R. Ade *et al*., *Planck* 2015 results. *Astron. Astrophys.* **594**, A13 (2016). doi: 10.1051/0004-6361/201525830
68. T. K. Gaisser, F. Halzen, T. Stanev, Particle astrophysics with high energy neutrinos. *Phys. Rep.* **258**, 173–236 (1995). doi: 10.1016/0370-1573(95)00003-Y

ACKNOWLEDGMENTS

MAGIC: We thank the Instituto de Astrofsica de Canarias for the excellent working conditions at the Observatorio del Roque de los Muchachos in La Palma. ***AGILE:*** We thank ASI personnel involved in the operations and data center of the *AGILE* mission. **ASAS-SN:** We thank Las Cumbres Observatory and its staff for their continued support of ASAS-SN. **HAWC:** Thanks to S. Delay, L. Daz, and E. Murrieta for technical support. **H.E.S.S:** We appreciate the excellent work of the technical support staff in Berlin, Zeuthen, Heidelberg, Palaiseau, Paris, Saclay, Tübingen and in Namibia in the construction and operation of the equipment. **VERITAS:** We acknowledge the excellent work of the technical support staff at the Fred Lawrence Whipple Observatory and at the collaborating institutions in the construction and operation of the instrument. **VLA/17B-403 team:** We thank the NRAO for granting us DDT VLA time to observe this source and the NRAO staff for rapidly executing the observations.

Funding

IceCube Collaboration: The IceCube collaboration gratefully acknowledge the support from the following agencies and institutions: USA—U.S. National Science Foundation-Office of Polar Programs, U.S. National Science Foundation-Physics Division, Wisconsin Alumni Research Foundation, Center for High Throughput Computing (CHTC) at the University of Wisconsin–Madison, Open Science Grid (OSG), Extreme Science and Engineering Discovery Environment (XSEDE), U.S. Department of Energy National Energy Research Scientific Computing Center, Particle astrophysics research computing center at the University of Maryland, Institute for Cyber-Enabled Research at Michigan State University, and Astroparticle physics computational facility at Marquette University; Belgium—Funds for Scientific Research (FRS-FNRS and FWO), FWO Odysseus and Big Science programmes, and Belgian Federal Science Policy Office (Belspo); Germany—Bundesministerium für Bildung und Forschung (BMBF), Deutsche Forschungsgemeinschaft (DFG), Helmholtz Alliance for Astroparticle Physics (HAP), Initiative and Networking Fund of the Helmholtz Association, Deutsches Elektronen Synchrotron (DESY), and High Performance Computing cluster of the RWTH Aachen; Sweden—Swedish Research Council, Swedish Polar Research Secretariat, Swedish National Infrastructure for Computing (SNIC), and Knut and Alice Wallenberg Foundation; Australia—Australian Research Council; Canada—Natural Sciences and Engineering Research Council of Canada, Calcul Québec, Compute Ontario, Canada Foundation for Innovation, WestGrid, and Compute Canada; Denmark—Villum Fonden, Danish National Research Foundation (DNRF); New Zealand—Marsden Fund; Japan—Japan Society for Promotion of Science (JSPS) and Institute for Global Prominent Research (IGPR) of Chiba University; Korea—National Research Foundation of Korea (NRF); Switzerland—Swiss National Science Foundation (SNSF). ***Fermi*-LAT collaboration:** The *Fermi*-LAT Collaboration acknowledges generous ongoing support from a number of agencies and institutes that have supported both the development and the operation of the LAT as well as scientific data analysis. These include the National Aeronautics and Space Administration and the Department of Energy in the United States, the Commissariat à l'Energie Atomique and the Centre National de la Recherche Scientifique /Institut National de Physique Nucléaire et de Physique des Particules in France, the Agenzia Spaziale Italiana and the Istituto Nazionale di Fisica Nucleare in Italy, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), High Energy Accelerator Research Organization (KEK) and Japan Aerospace Exploration Agency (JAXA) in Japan, and the K. A. Wallenberg Foundation, the Swedish Research Council and the Swedish National Space Board in Sweden. Additional support for science analysis during the operations phase is gratefully acknowledged from the Istituto Nazionale di Astrofisica in Italy and the Centre National d'Études Spatiales in France. This work performed in part under DOE Contract DE-AC02-76SF00515. **MAGIC collaboration:** The financial support of the German BMBF and MPG, the Italian INFN and INAF, the Swiss National Fund SNF, the ERDF under the Spanish MINECO (FPA2015-69818-P, FPA2012-36668, FPA2015-68378-P, FPA2015-69210-C6-2-R, FPA2015-69210-C6-4-R, FPA2015-69210-C6-6-R, AYA2015-71042-P, AYA2016-76012-C3-1-P, ESP2015-71662-C2-2-P, CSD2009-00064), and the Japanese JSPS and MEXT is gratefully acknowledged. This work was also supported by the Spanish Centro de Excelencia "Severo Ochoa" SEV-2012-0234 and SEV-2015-0548, and Unidad de Excelencia "Mara de Maeztu" MDM-2014-0369, by the Croatian Science Foundation (HrZZ) Project IP-2016-06-9782 and the University of Rijeka Project 13.12.1.3.02, by the DFG Collaborative Research Centers SFB823/C4 and SFB876/C3, the Polish National Research Centre grant UMO-2016/22/M/ST9/00382 and by the Brazilian MCTIC, CNPq and FAPERJ. ***AGILE:*** *AGILE* is an ASI space mission developed with scientific and programmatic support from INAF and INFN. Research partially supported through the ASI grant no. I/028/12/2. Part of this work is based on archival data, software or online services provided by the ASI Space Science Data Center (SSDC, previously known as ASDC). **ASAS-SN:** ASAS-SN is funded in part by the Gordon and Betty Moore Foundation through grant GBMF5490 to the Ohio State University, NSF grant AST-1515927, the Mt. Cuba Astronomical Foundation, the Center for Cosmology and AstroParticle Physics (CCAPP) at OSU, and the Chinese Academy of Sciences South America Center for Astronomy (CASSACA). A.F. was supported by the Initiative and Networking Fund of the Helmholtz Association. J.F.B is supported by NSF grant PHY-1714479. S.D. acknowledges Project 11573003 supported by NSFC. J.L.P. is supported by FONDECYT grant 1151445 and by the Ministry of Economy, Development, and Tourism's Millennium Science Initiative through grant IC120009, awarded to The Millennium Institute of Astrophysics (MAS). This research was made possible through the use of the AAVSO Photometric All-Sky Survey (APASS), funded by the Robert Martin Ayers Sciences Fund. **HAWC:** HAWC acknowledges the support from: the US National Science Foundation (NSF) the US Department of Energy Office of High-Energy Physics; the Laboratory Directed Research and Development (LDRD) program of Los Alamos National Laboratory; Consejo Nacional de Ciencia y Tecnologa (CONACyT), México (grants 271051, 232656, 260378, 179588, 239762, 254964, 271737, 258865, 243290, 132197, 281653) (Cátedras 873, 1563), Laboratorio Nacional HAWC de rayos gamma; L'OREAL Fellowship for Women in Science 2014; Red HAWC, México; DGAPA-UNAM (grants IG100317, IN111315, IN111716-3, IA102715, 109916, IA102917, IN112218); VIEP-BUAP; PIFI 2012, 2013, PROFOCIE 2014, 2015; Royal Society grant Newton Advanced Fellowship 180385; the University of Wisconsin Alumni Research Foundation; the Institute of Geophysics, Planetary Physics, and Signatures at Los Alamos National Laboratory; Polish Science Centre grant DEC-2014/13/B/ST9/945; Coordinación de la Investigación Cientfica de la Universidad Michoacana. **H.E.S.S.:** 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 Knut and Alice Wallenberg Foundation, the National Science Centre, Poland grant no. 2016/22/M/ST9/00382, 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 and by the University of Amsterdam. This work benefited from services provided by the H.E.S.S. Virtual Organisation, supported by the national resource providers of the EGI Federation. ***INTEGRAL:*** *INTEGRAL* is an ESA space mission, with its instruments and science data center funded by the ESA member states (specifically the PI countries: Denmark, France, Germany, Italy, Switzerland, Spain), and with additional participation of Russia and the USA. The *INTEGRAL* SPI instrument was provided through Co-PI institutes IRAP (Toulouse/France) and MPE (Garching/Germany), the SPI project was coordinated and managed by CNES (Toulouse/France). The *INTEGRAL* IBIS instrument was provided through Co-PI institutes IAPS (Rome/Italy) and CEA (Saclay/France). The SPI-ACS detector system has been provided by MPE Garching/Germany. The SPI team is grateful to ASI, CEA, CNES, DLR, ESA, INTA, NASA, and OSTC for their support. The Italian *INTEGRAL* team acknowledges the support of ASI/INAF agreement n. 2013-025-R.1. J.R. acknowledges support from the European Union's Horizon 2020 Programme under the AHEAD project (grant no. 654215). R.D. acknowledges the German *INTEGRAL* support through DLR grants 50 OG 1101 and 1601. **Kanata, Kiso and Subaru observing teams:** Observations with the Kanata and Kiso Schmidt telescopes were supported by the Optical and Near-infrared Astronomy Inter-University Cooperation Program and the Grants-in-Aid of the Ministry of Education, Science, Culture, and Sport JP23740143, JP25800103, JP16H02158, JP17K14253, JP17H04830, JP26800103, JP24103003. This work was also based in part on data collected at Subaru Telescope, which is operated by the National Astronomical Observatory of Japan. **Kapteyn:** The Jacobus Kapteyn telescope is operated at the Observatorio del Roque de los Muchachos on the Spanish island of La Palma by the SARA consortium, whose member institutions (listed at http://saraobservatory.org) fund its operation. Refitting for remote operations and instrumentation were funded by the National Science Foundation under grant 1337566 to Texas A&M University–Commerce. **Liverpool Telescope:** The Liverpool Telescope is operated on the island of La Palma by Liverpool John Moores University in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofisica de Canarias with financial support from the UK Science and Technology Facilities Council. ***Swift*/*NuSTAR*:** A.K. and and D.F.C. acknowledge support from the National Aeronautics and Space Administration *Swift* Guest Investigator Program under grant NNX17AI95G. The *Swift* team at the Mission Operations Center (MOC) at Penn State acknowledges support from NASA contract NAS5-00136. *Swift* is supported at the University of Leicester by the UK Space Agency. **VERITAS:** This research is supported by grants from the U.S. Department of Energy Office of Science, the U.S. National Science Foundation and the Smithsonian Institution, and by NSERC in Canada. **VLA/17B-403** The National Radio Astronomy Observatory (NRAO) is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. A.J.T. is supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Post-Graduate Doctoral Scholarship (PGSD2-490318-2016). A.J.T. and G.R.S. are supported by NSERC Discovery Grants (RGPIN-402752-2011 and RGPIN-06569-2016). J.C.A.M.J. is the recipient of an Australian Research Council Future Fellowship (FT140101082).

Author contributions: All authors meet the journal's authorship criteria. **IceCube:** The IceCube Collaboration designed, constructed and now operates the the IceCube Neutrino Observatory. Data processing and calibration, Monte Carlo simulations of the detector and of theoretical models, and data analyses were performed by a large number of collaboration members, who also discussed and approved the scientific results presented here. The paper was reviewed by the entire collaboration before publication, and all authors approved the final version of the manuscript. ***Fermi*-LAT:** The *Fermi*-LAT contact authors and internal reviewers are S.B., A.F., Y.T., K.B., E.C., and M.W. **MAGIC:** E.B. is the MAGIC multimessenger contact and PI of the neutrino follow-up program. K.S. is co-convener of the MAGIC transient working group. L.F. and M.P. are the main analyzers of the MAGIC data. A.M. and E.P. derived an upper limit to the redshift inferred from MAGIC data. ***AGILE*:** All coauthors contributed to the scientific results presented in the paper. F.L. and M.T. wrote the part of the manuscript related to the *AGILE* results. **ASAS-SN:** A.F., B.J.S., and S.H. installed an automatic follow up to IceCube triggers which provided additional early data on this event. K.Z.S., C.S.K., J.F.B., T.A.T., T.W.S.H., S.D., J.L.P., and B.J.S. built the telescopes and developed the data processing pipelines. **HAWC:** T.W. is convener of the HAWC extragalactic working group. M.H. is the HAWC multimessenger contact. I.T., R.L., and I.M.C. were the main analyzers of the HAWC data. **H.E.S.S.:** A.T. is convener of the H.E.S.S. extragalactic working group. F.S. is the H.E.S.S. multimessenger contact and PI of the neutrino follow-up program. C.H. is the main analyzer of the H.E.S.S. data presented here. S.O. provided a cross-check of the presented analysis. ***INTEGRAL:*** V.S. performed the *INTEGRAL* analysis. C.F. is the PI of *INTEGRAL* Science Data Center. R.D. is the co-PI of the SPI instrument. E.K. is *INTEGRAL* Project Scientist. P.L. and P.U. are co-PIs of *INTEGRAL*/IBIS instrument. S.M. is responsible for the *INTEGRAL*/IBAS. All of the collaborators provided contribution to the text. **Kanata, Kiso and Subaru observing teams:** Y.T.T., Y.U., and K.O. developed the follow-up strategy to search for IceCube counterparts. T.N. and M.K. conducted the near-infrared imaging and polarimetric observations of the TXS 0506+056 using the HONIR instrument on the Kanata telescope, which were processed by the data reduction system developed by R.I. The reduced

images were mainly examined by H.M. and H.N. M.Y. reduced the polarimetric data. K.S.K. supervised all of the above. T.M. conducted the optical imaging observations of TXS 0506+056 with the KWFC instrument on the Kiso Schmidt telescope and reduced the data. Y.M. conducted optical spectroscopic observations of the TXS 0506 +056 with the FOCAS spectrograph on the 8.2 m Subaru telescope. The data are reduced and examined by M.Y. and T.M. **Kapteyn:** W.C.K. obtained and reduced the optical observations at the Kapteyn telescope. **Liverpool Telescope:** I.S. and C.C. obtained, reduced, and analyzed the Liverpool Telescope spectra. ***Swift*/*NuSTAR*:** A.K. led reduction of *Swift* XRT data, and J.J.D. led reduction of *NuSTAR* data. D.B.F. carried out the joint *Swift* XRT + *NuSTAR* analysis, and A.K. managed author contributions to this section. **VERITAS:** The construction, operation, and maintenance of the VERITAS telescopes, as well as the tools to analyze the VERITAS data, are the work of the the VERITAS Collaboration as a whole. The VERITAS Collaboration contacts for this paper are M.S. and D.A.W. **VLA/17B-403:** GRS wrote the Director's Discretionary Time proposal for the VLA observations. A.J.T. performed the VLA data reduction and analyses in consultation with the rest of the team. A.J.T. and G.R.S. wrote the VLA-related text in consultation with the rest of the team. G.R.S. contributed to writing the entire paper. **Competing interests:** All collaborations declare no competing interests.

Data and materials availability: IceCube: All IceCube data related to the the results presented in this paper are provided in the supplementary materials (*25*). ***Fermi*-LAT:** The *Fermi*-LAT data are available from the *Fermi* Science Support Center http://fermi.gsfc.nasa.gov/ssc and https://www-glast.stanford.edu/pub_data/1483/. **MAGIC:** The MAGIC data and analysis results are accessible at https://magic.mpp.mpg.de/public/public-data/ ***AGILE*:** The *AGILE* data are available at www.ssdc.asi.it/mmia/index.php?mission=agilemmia. Data analysis software and calibrations are available at http://agile.ssdc.asi.it/publicsoftware.html. **ASAS-SN:** The ASAS-SN light curves are available at https://asas-sn.osu.edu **HAWC:** The HAWC data are available at https://data.hawc-observatory.org/datasets/ic170922/index.php. **H.E.S.S.:** The H.E.S.S. data are available at https://www.mpi-hd.mpg.de/hfm/HESS/pages/publications/auxiliary/auxinfo_TXS0506.html ***INTEGRAL*:** The *INTEGRAL* data and analysis software are available at www.isdc.unige.ch/. **Kanata, Kiso and Subaru observing teams:** Data taken with the Kiso, Kanata, and Subaru telescopes are available in the archive SMOKA https://smoka.nao.ac.jp/, operated by the Astronomy Data Center, National Astronomical Observatory of Japan. The Subaru data were taken in the open-use program S16B-071I. **Kapteyn:** The Kapteyn data were taken from (*53*) **Liverpool Telescope:** The Liverpool Telescope data are available in the telescope archive at http://telescope.livjm.ac.uk/cgi-bin/lt_search. ***Swift*/*NuSTAR*:** The *Swift* data are available at www.swift.ac.uk/archive/obs.php. The initial tiling observations, shortly after the IceCube trigger, are targetIDs 10308-10326. The monitoring ObsIDs observed from 23 September to 23 October are 00010308001, 00083368001-006, 00010308008-013. The *NuSTAR* data are available at https://heasarc.gsfc.nasa.gov/FTP/nustar/data/obs/03/9//90301618002/ under ObsID 90301618002. **VERITAS:** The VERITAS data are available at https://veritas.sao.arizona.edu/veritas-science/veritas-results-mainmenu-72/490-ic-result **VLA/17B-403 team:** The VLA data are available through the NRAO Science Data Archive https://archive.nrao.edu/archive/advquery.jsp under Project Code 17B-403.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/eaat1378/suppl/DC1
Full Author List
Materials and Methods
Tables S1 to S10
Figs. S1 to S7
References (*69*–*116*)

9 February 2018; accepted 8 June 2018
10.1126/science.aat1378

RESEARCH ARTICLE

NEUTRINO ASTROPHYSICS

Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert

IceCube Collaboration*†

A high-energy neutrino event detected by IceCube on 22 September 2017 was coincident in direction and time with a gamma-ray flare from the blazar TXS 0506+056. Prompted by this association, we investigated 9.5 years of IceCube neutrino observations to search for excess emission at the position of the blazar. We found an excess of high-energy neutrino events, with respect to atmospheric backgrounds, at that position between September 2014 and March 2015. Allowing for time-variable flux, this constitutes 3.5σ evidence for neutrino emission from the direction of TXS 0506+056, independent of and prior to the 2017 flaring episode. This suggests that blazars are identifiable sources of the high-energy astrophysical neutrino flux.

The origin of the highest-energy cosmic rays is believed to be extragalactic (*1*), but their acceleration sites remain unidentified. High-energy neutrinos are expected to be produced in or near the acceleration sites when cosmic rays interact with matter and ambient light, producing charged mesons that decay into neutrinos and other particles. Unlike cosmic rays, neutrinos can travel through the Universe unimpeded by interactions with other particles and undeflected by magnetic fields, providing a means to identify and study the extreme environments producing cosmic rays (*2*). Blazars, a class of active galactic nuclei with powerful relativistic jets pointed close to our line of sight (*3*), are prominent candidate sources of such high-energy neutrino emission (*4*–*9*). The electromagnetic emission of blazars is observed to be highly variable on time scales from minutes to years (*10*).

The IceCube Neutrino Observatory (*11*) is a high-energy neutrino detector occupying an instrumented volume of 1 km^3 within the Antarctic ice sheet at the Amundsen-Scott South Pole Station. The detector consists of an array of 86 vertical strings, nominally spaced 125 m apart and descending to a depth of approximately 2450 m in the ice. The bottom 1 km of each string is equipped with 60 optical sensors that record Cherenkov light emitted by relativistic charged particles passing through the optically transparent ice. When high-energy muon neutrinos interact with the ice, they can create relativistic muons that travel many kilometers, creating a track-like series of Cherenkov photons recorded when they pass through the array. This allows the reconstruction of the original neutrino direction with a median angular uncertainty of 0.5° for a neutrino energy of ~30 TeV (or 0.3° at 1 PeV) (*12*, *13*).

IceCube discovered the existence of a diffuse flux of high-energy astrophysical neutrinos in 2013 (*14*, *15*). Measurements of the energy spectrum have since been refined (*16*, *17*), indicating that the neutrino spectrum extends above several PeV. However, analyses of neutrino observations have not succeeded in identifying individual sources of high-energy neutrinos (*12*, *18*). This suggests that the sources are distributed across the sky and that even the brightest individual sources contribute only a small fraction of the total observed flux.

Recently, the detection of a high-energy neutrino by IceCube, together with observations in gamma rays and at other wavelengths, indicates that a blazar, TXS 0506+056, located at right ascension (RA) 77.3582° and declination (Dec) +5.69314° (J2000 equinox) (*19*) may be an individually identifiable source of high-energy neutrinos (*20*). The neutrino-candidate event, IceCube-170922A, was detected on 22 September 2017, selected by the Extremely High Energy (EHE) online event filter (*21*), and reported as a public alert (*22*). EHE alerts are currently sent at a rate of about four per year, and are based on well-reconstructed, high-energy muon-track events. The selection threshold is set so that approximately half of the events are estimated to be astrophysical neutrinos, the rest being atmospheric background events. After the alert was sent, further studies refined the directional reconstruction, with best-fitting coordinates of RA $77.43^{+0.95}_{-0.65}$ and Dec $+5.72^{+0.50}_{-0.30}$ (degrees, J2000, 90% containment region). The most probable neutrino energy was estimated to be 290 TeV, with a 90% confidence level lower limit of 183 TeV (*20*).

It was soon determined that the direction of IceCube-170922A was consistent with the location of TXS 0506+056 and coincident with a state of enhanced gamma-ray activity observed since April 2017 (*23*) by the Large Area Telescope (LAT) on the Fermi Gamma-ray Space Telescope (*24*). Follow-up observations of the blazar led to the detection of gamma rays with energies up to 400 GeV by the Major Atmospheric Gamma Imaging Cherenkov (MAGIC) Telescopes (*25*, *26*). IceCube-170922A and the electromagnetic observations are described in detail in (*20*). The significance of the spatial and temporal coincidence of the high-energy neutrino and the blazar flare is estimated to be at the 3σ level (*20*). On the basis of this result, we consider the hypothesis that the blazar TXS 0506+056 has been a source of high-energy neutrinos beyond that single event.

Searching for neutrino emission

IceCube monitors the whole sky and has maintained essentially continuous observations since 5 April 2008. Searches for neutrino point sources using two model-independent methods, a time-integrated and a time-dependent unbinned maximum likelihood analysis, have previously been published for the data collected between 2008 and 2015 (*12*, *18*, *27*). Here, we analyze the same 7-year data sample supplemented with additional data collected from May 2015 until October 2017 (*21*). The data span 9.5 years and consist of six distinct periods, corresponding to changing detector configurations, data-taking conditions, and improved event selections (Table 1).

The northern sky, where TXS 0506+056 is located, is observed through Earth by IceCube. Approximately 70,000 neutrino-induced muon tracks are recorded each year from this hemisphere of the sky after passing the final event selection criteria. Fewer than 1% of these events originate from astrophysical neutrinos; the vast majority are background events caused by neutrinos of median energy ~1 TeV created in cosmic ray interactions in the atmosphere over other locations on Earth. However, for an astrophysical muon-neutrino flux where the differential number of neutrinos with energy E scales as $dN/dE \sim E^{-2}$, the distribution of muon energies is different than for the background atmospheric neutrino flux, which scales as $\sim E^{-3.7}$ (*17*). This allows for further discriminating power in point source searches besides directional-only excesses.

A high-significance point source detection (*12*, *18*) can require as few as two or three, or as many as 30, signal events to stand out from the background, depending on the energy spectrum and the clustering of events in time. To search for a neutrino signal at the coordinates of TXS 0506+056, we apply the standard time-integrated analysis (*28*) and time-dependent analysis (*29*) that have been used in past searches (*12*, *18*, *27*). The time-integrated analysis uses an unbinned maximum likelihood ratio method to search for an excess number of events consistent with a point source at a specified location, given the angular distance and angular uncertainty of each event. Energy information is included in the definition of the likelihood, assuming a power-law energy spectrum $E^{-\gamma}$, with the spectral index γ

*IceCube Collaboration authors and affiliations are listed in the supplementary materials.
†Email: analysis@icecube.wisc.edu

as a fitted parameter. The model parameters are correlated and are expressed as a pair, (Φ_{100}, γ), where Φ_{100} is the flux normalization at 100 TeV. The time-dependent analysis uses the same formulation of the likelihood but searches for clustering in time as well as space by introducing an additional time profile. It is performed separately for two different generic profile shapes: a Gaussian-shaped time window and a box-shaped time window. Each analysis varies the central time of the window, T_0, and the duration T_W (from seconds to years) of the potential signal to find the four parameters (Φ_{100}, γ, T_0, T_W) that maximize the likelihood ratio, which is defined as the test statistic *TS*. (For the Gaussian time window, T_W represents twice the standard deviation.) The test statistic includes a factor that corrects for the look-elsewhere effect arising from all of the possible time windows that could be chosen (*30*).

For each analysis method (time-integrated and time-dependent), a robust significance estimate is obtained by performing the identical analysis on trials with randomized datasets. These are produced by randomizing the event times and recalculating the RA coordinates within each data-taking period. The resultant *P* value is defined as the fraction of randomized trials yielding a value of *TS* greater than or equal to the one obtained for the actual data.

Because the detector configuration and event selections changed as shown in Table 1, the time-dependent analysis is performed by operating on each data-taking period separately. (A flare that spans a boundary between two periods could be partially detected in either period, but with reduced significance.) An additional look-elsewhere correction then needs to be applied for a result in an individual data segment, given by the ratio of the total 9.5-year observation time to the observation time of that data segment (*30*).

Table 1. IceCube neutrino data samples. Six data-taking periods make up the full 9.5-year data sample. Sample numbers correspond to the number of detector strings that were operational. During the first three periods, the detector was still under construction. The last three periods correspond to different data-taking conditions and/or event selections with the full 86-string detector.

Sample	Start	End
IC40	5 April 2008	20 May 2009
IC59	20 May 2009	31 May 2010
IC79	31 May 2010	13 May 2011
IC86a	13 May 2011	16 May 2012
IC86b	16 May 2012	18 May 2015
IC86c	18 May 2015	31 October 2017

Neutrinos from the direction of TXS 0506+056

The results of the time-dependent analysis performed at the coordinates of TXS 0506+056 are shown in Fig. 1 for each of the six data periods. One of the data periods, IC86b from 2012 to 2015, contains a significant excess, which is identified by both time-window shapes. The excess consists of 13 ± 5 events above the expectation from the atmospheric background. The significance depends on the energies of the events, their proximity to the coordinates of TXS 0506+056, and their clustering in time. This is illustrated in Fig. 2, which shows the time-independent weight of individual events in the likelihood analysis during the IC86b data period.

The Gaussian time window is centered at 13 December 2014 [modified Julian day (MJD) 57004] with an uncertainty of ±21 days and a duration $T_W = 110^{+35}_{-24}$ days. The best-fitting parameters for the fluence $J_{100} = \int \Phi_{100}(t)dt$ and the spectral index are given by $E^2J_{100} = 2.1^{+0.9}_{-0.7} \times 10^{-4}$ TeV cm^{-2} at 100 TeV and $\gamma = 2.1 \pm 0.2$, respectively. The joint uncertainty on these parameters is shown in Fig. 3 along with a skymap showing the result of the time-dependent analysis performed at the location of TXS 0506+056 and in its vicinity during the IC86b data period.

The box-shaped time window is centered 13 days later with duration $T_W = 158$ days (from MJD 56937.81 to MJD 57096.21, inclusive of contributing events at boundary times). For the box-shaped time window, the uncertainties are discontinuous and not well defined, but the uncertainties for the Gaussian window show that it is consistent with the box-shaped time window fit. Despite the different window shapes, which lead to different weightings of the events as a function of time, both windows identify the same time interval as significant. For the box-shaped time window, the best-fitting parameters are similar to those of the Gaussian window, with fluence at 100 TeV and spectral index given by $E^2J_{100} = 2.2^{+1.0}_{-0.8} \times 10^{-4}$ TeV cm^{-2} and $\gamma = 2.2 \pm 0.2$. This fluence corresponds to an average flux over 158 days of $\Phi_{100} = 1.6^{+0.7}_{-0.6} \times 10^{-15}$ TeV^{-1} cm^{-2} s^{-1}.

When we estimate the significance of the time-dependent result by performing the analysis at the coordinates of TXS 0506+056 on randomized datasets, we allow in each trial a new fit for all the parameters: Φ_{100}, γ, T_0, T_W. We find that the fraction of randomized trials that result in a more significant excess than the real data is 7×10^{-5} for the box-shaped time window and 3×10^{-5} for the Gaussian time window. This fraction, once corrected for the ratio of the total observation time to the IC86b observation time (9.5 years/3 years), results in *P* values of 2×10^{-4} and 10^{-4}, respectively, corresponding to 3.5σ and 3.7σ. Because there is no a priori reason to prefer one of the generic time windows over the other, we take the more significant one and include a trial factor of 2 for the final significance, which is then 3.5σ.

Outside the 2012–2015 time period, the next most significant excess is found using the Gaussian window in 2017 and includes the IceCube-170922A event. This time window is centered at 22 September 2017 with duration $T_W = 19$ days, $\gamma = 1.7 \pm 0.6$, and fluence $E^2J_{100} = 0.2^{+0.4}_{-0.2} \times 10^{-4}$ TeV cm^{-2} at 100 TeV. No other event besides the IceCube-170922A event contributes significantly to the best fit. As a consequence, the uncertainty on the best-fitting window location and width spans the entire IC86c period, because any window containing IceCube-170922A yields a similar value of the test statistic. Following the trial correction procedure for different observation periods as described above, the significance of this excess

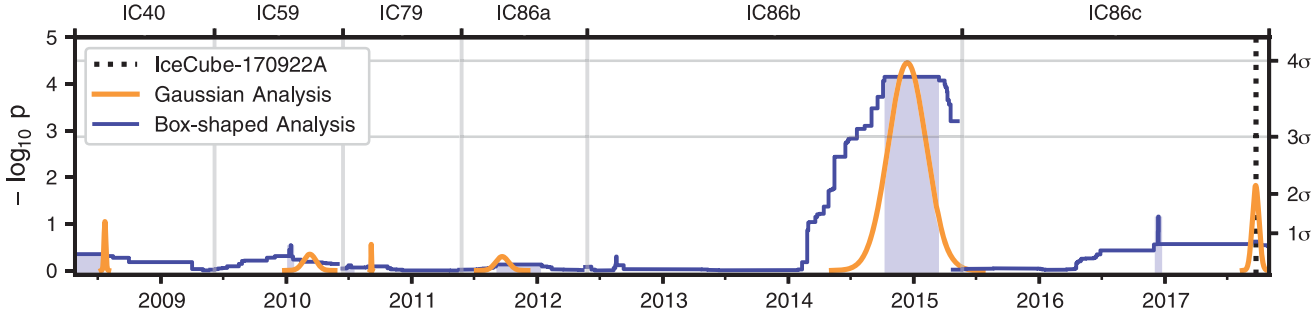


Fig. 1. Time-dependent analysis results. The orange curve corresponds to the analysis using the Gaussian-shaped time profile. The central time T_0 and width T_W are plotted for the most significant excess found in each period, with the *P* value of that result indicated by the height of the peak. The blue curve corresponds to the analysis using the box-shaped time profile. The curve traces the outer edge of the superposition of the best-fitting time windows (durations T_W) over all times T_0, with the height indicating the significance of that window. In each period, the most significant time window forms a plateau, shaded in blue. The large blue band centered near 2015 represents the best-fitting 158-day time window found using the box-shaped time profile. The vertical dotted line in IC86c indicates the time of the IceCube-170922A event.

is 1.4σ. If the IceCube-170922A event is removed, no excess remains during this time period. This agrees with the result of the rapid-response analysis (*31*) that is part of the IceCube alert program, which found no other potential astrophysical neutrinos from the same region of the sky during ±7 days centered on the time of IceCube-170922A.

We performed a time-integrated analysis at the coordinates of TXS 0506+056 using the full 9.5-year data sample. The best-fitting parameters for the flux normalization and the spectral index are $\Phi_{100} = 0.8^{+0.5}_{-0.4} \times 10^{-16}$ TeV^{-1} cm^{-2} s^{-1} and $\gamma = 2.0 \pm 0.3$, respectively. The joint uncertainty on these parameters is shown in Fig. 4A. The *P* value, based on repeating the analysis at the same coordinates with randomized datasets, is 0.002% (4.1σ), but this is an a posteriori significance estimate because it includes the IceCube-170922A event, which motivated performing the analysis at the coordinates of TXS 0506+056. An unbiased significance estimate including the event would need to take into account the look-elsewhere effect related to all other possible directions in the sky that could be analyzed. It is expected that there will be two or three directions somewhere in the northern sky with this significance or greater, resulting from the chance alignment of neutrinos (*12*). Here, we are interested in determining whether there is evidence of time-integrated neutrino emission from TXS 0506+056 besides the IceCube-170922A event.

If we remove the final data period IC86c, which contains the event, and perform the analysis again using only the first 7 years of data, we find best-fitting parameters that are nearly unchanged: $\Phi_{100} = 0.9^{+0.6}_{-0.5} \times 10^{-16}$ TeV^{-1} cm^{-2} s^{-1} and $\gamma = 2.1 \pm 0.3$, respectively. The joint uncertainty on these parameters is shown in Fig. 4B. The *P* value, using only the first 7 years of data, is 1.6% (2.1σ), based on repeating the analysis at the same coordinates with randomized datasets. These results indicate that the time-integrated fit is dominated by the same excess as found in the time-dependent analysis above, having similar values for the spectral index and total fluence ($E^2J_{100} = 2.0 \times 10^{-4}$ TeV cm^{-2} at 100 TeV over the 7-year period). This excess is not significant in the time-integrated analysis because of the additional background during the rest of the 7-year period.

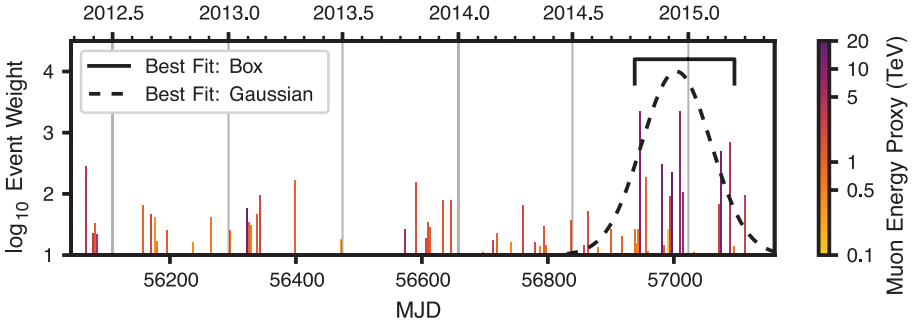


Fig. 2. Time-independent weight of individual events during the IC86b period. Each vertical line represents an event observed at the time indicated by calendar year (top) or MJD (bottom). Overlapping lines are shifted by 1 to 2 days for visibility. The height of each line indicates the event weight: the product of the event's spatial term and energy term in the unbinned likelihood analysis evaluated at the location of TXS 0506+056 and assuming the best-fitting spectral index $\gamma = 2.1$ (*30*). The color for each event indicates an approximate value in units of TeV of the reconstructed muon energy (muon energy proxy), which the analysis compares with expected muon energy distributions under different hypotheses. [A distribution for the true neutrino energy of a single event can also be inferred from the event's muon energy (*30*).] The dashed curve and the solid bracket indicate the best-fitting Gaussian and box-shaped time windows, respectively. The distribution of event weights and times outside of the best-fitting time windows is compatible with background.

Blazars as neutrino sources

The signal identified during the 5-month period in 2014–2015 consists of an estimated 13 ± 5 muon-neutrino events that are present in addition to the expected background. The analysis is unbinned, but the mean background at the declination of TXS 0506+056 is useful for comparison purposes; it is 5.8 events in a search bin of radius 1° during a 158-day time window. (We use the duration of the box-shaped time window result for convenience to calculate averages during the flare.) The significance of the excess is due to both the number of events and their energy distribution, with higher-energy events increasing the significance and leading to the best-fitting spectral index of 2.1, in contrast to the lower-energy atmospheric neutrino background with spectral index ~3.7. At this declination in the sky, the 68% central energy range in which IceCube is most sensitive to point sources with $E^{-2.1}$ spectra is between 32 TeV and 3.6 PeV. Assuming that the muon-neutrino fluence ($E^2J_{100} = 2.1^{+1.0}_{-0.7} \times 10^{-4}$ TeV cm^{-2}) is one-third of the total neutrino fluence, then the all-flavor neutrino energy fluence is $4.2^{+2.0}_{-1.4} \times 10^{-3}$ erg cm^{-2} over this energy range. With the recent measurement (*32*) of the redshift of TXS 0506+056 as $z = 0.3365 \pm 0.0010$, this energy fluence implies that the isotropic neutrino luminosity is $1.2^{+0.6}_{-0.4} \times 10^{47}$ erg s^{-1} averaged over 158 days. This is higher than the isotropic gamma-ray luminosity during the same period, which is similar to the long-term luminosity between 0.1 GeV and 100 GeV of 0.28×10^{47} erg s^{-1} averaged over all Fermi-LAT observations of TXS 0506+056 (*20*). Gamma rays are expected to

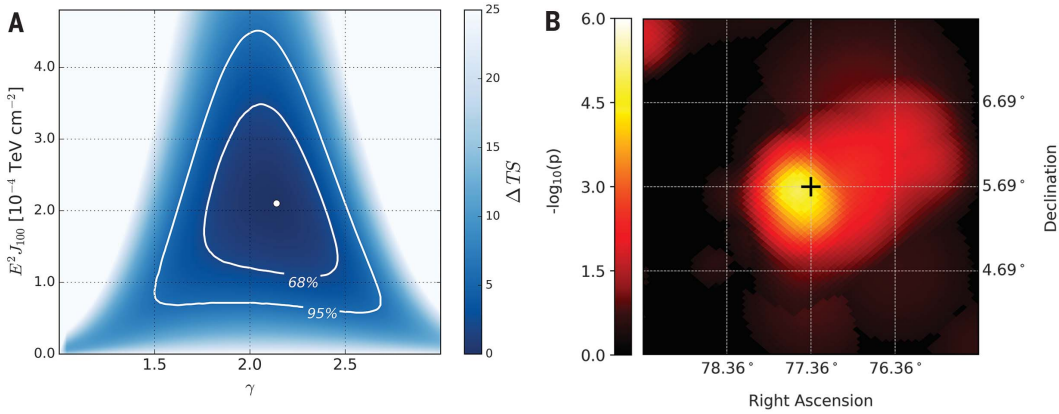


Fig. 3. Time-dependent analysis results for the IC86b data period (2012–2015). (**A**) Change in test statistic, Δ*TS*, as a function of the spectral index parameter γ and the fluence at 100 TeV given by E^2J_{100}. The analysis is performed at the coordinates of TXS 0506+056, using the Gaussian-shaped time window and holding the time parameters fixed (T_0 = 13 December 2014, T_W = 110 days). The white dot indicates the best-fitting values. The contours at 68% and 95% confidence level assuming Wilks' theorem (*36*) are shown in order to indicate the statistical uncertainty on the parameter estimates. Systematic uncertainties are not included. (**B**) Skymap showing the *P* value of the time-dependent analysis performed at the coordinates of TXS 0506+056 (cross) and at surrounding locations. The analysis is performed on the IC86b data period, using the Gaussian-shaped time window. At each point, the full fit for (Φ, γ, T_0, T_W) is performed. The *P* value shown does not include the look-elsewhere effect related to other data periods. An excess of events is detected, consistent with the position of TXS 0506+056.

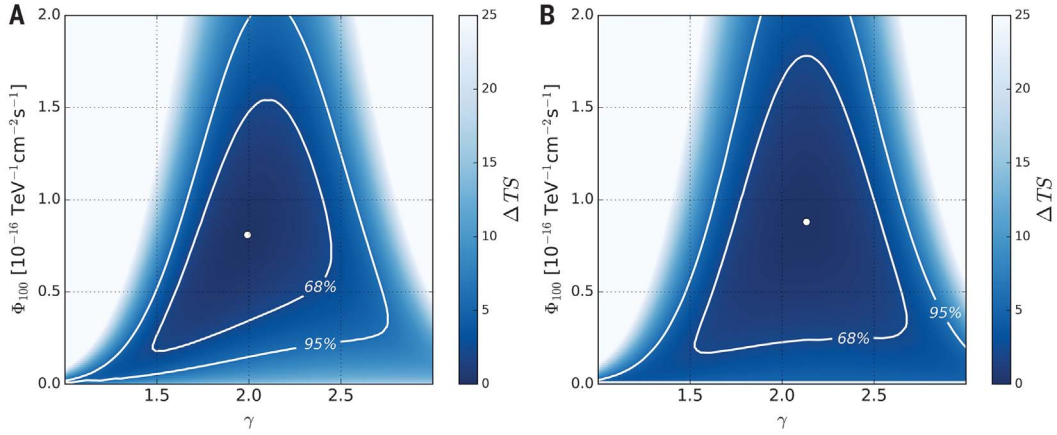


Fig. 4. Time-integrated analysis results. As in Fig. 3A, but for the time-integrated analysis of TXS 0506+056 using (**A**) the full 9.5-year sample (2008–2017), and (**B**) the 7-year sample (2008–2015).

be produced in the same processes that produce neutrinos—for example, when accelerated protons interact with ambient lower-energy photons near the source, producing both neutral pions (which decay to gamma rays) and charged pions (which decay to neutrinos and leptons). A higher luminosity in neutrinos than in gamma rays could imply that a substantial fraction of the gamma rays related to the neutrino production are either absorbed or arriving at energies above or below the Fermi-LAT energy band.

Although TXS 0506+056 is a bright object in gamma rays, it was not previously singled out as a predicted neutrino source. In the third catalog of active galactic nuclei detected by Fermi-LAT (*33*) listing 1773 objects (including those at low galactic latitudes), TXS 0506+056 is among the 50 brightest objects, with an average flux between 1 GeV and 100 GeV of 6.5 (±0.2) × 10^{-9} photons cm^{-2} s^{-1}. Its measured redshift now makes it one of the most luminous objects known out to the same distance, more than an order of magnitude more luminous than nearby blazars such as Markarian 421, Markarian 501, and 1ES 1959+650. With respect to these objects, an important observational distinction is the favorable declination of TXS 0506+056. As the neutrino-nucleon interaction cross section grows with energy, absorption in Earth becomes considerable for neutrinos above ~100 TeV. IceCube is most sensitive to high-energy neutrinos from sources at declinations near the equatorial plane, which is viewed along the horizon from the South Pole. The blazars mentioned above are at more northern declinations, and the likelihood that a neutrino with energy of ~300 TeV from one of these will be absorbed while traversing Earth is three to five times the likelihood that it will reach the detector. The explanation for why TXS 0506+056 is the first blazar associated with a significant neutrino excess may therefore depend on the combination of its intrinsic properties and the observational properties of the detector.

IceCube recently published (*34*) a search for neutrino emission from the blazars in the second catalog of active galactic nuclei detected by Fermi-LAT (*35*), constraining their contribution to the diffuse astrophysical neutrino flux under different model assumptions. An upper limit of 27% was found assuming the diffuse flux that is fit between 10 TeV and 100 TeV with a soft $E^{-2.5}$ spectrum (*16*). For an E^{-2} spectrum compatible with the diffuse flux fit above ~200 TeV (*17*), the upper limit is between 40% and 80%. The allowed contribution by blazars as a population is larger, because it would include the contribution of fainter and more distant blazars not yet resolved in the catalog. Averaged over 9.5 years, the neutrino flux of TXS 0506+056 by itself corresponds to 1% of the astrophysical diffuse flux and is fully compatible with the previous blazar catalog results.

The evidence presented above supports the hypothesis presented in (*20*) that the blazar TXS 0506+056 is a high-energy neutrino source. The 3.5σ evidence for neutrino emission during the 5-month period in 2014–2015 is statistically independent of the evidence presented in (*20*). The analysis of the IceCube-170922A event in (*20*) relies on correlation of a single neutrino with electromagnetic activity, whereas the analysis presented here relies only on self-correlation of multiple neutrinos. The coincidence of an IceCube alert with a flaring blazar, combined with a neutrino flare from the same object in archival IceCube data, pinpoints a likely source of high-energy cosmic rays.

REFERENCES AND NOTES

1. Pierre Auger Collaboration, *Science* **357**, 1266–1270 (2017).
2. F. Reines, *Annu. Rev. Nucl. Part. Sci.* **10**, 1–26 (1960).
3. C. M. Urry, P. Padovani, *Publ. Astron. Soc. Pac.* **107**, 803 (1995).
4. K. Mannheim, *Astropart. Phys.* **3**, 295–302 (1995).
5. F. Halzen, E. Zas, *Astrophys. J.* **488**, 669–674 (1997).
6. A. Mücke, R. J. Protheroe, R. Engel, J. P. Rachen, T. Stanev, *Astropart. Phys.* **18**, 593–613 (2003).
7. K. Murase, in *Neutrino Astronomy*, T. Gaisser, A. Karle, Eds. (World Scientific, 2017), pp. 15–31.
8. M. Petropoulou, C. D. Dermer, *Astrophys. J. Lett.* **825**, L11 (2017).
9. C. Guépin, K. Kotera, *Astron. Astrophys.* **603**, A76 (2017).
10. M.-H. Ulrich, L. Maraschi, C. M. Urry, *Annu. Rev. Astron. Astrophys.* **35**, 445–502 (1997).
11. M. G. Aartsen *et al.*, *J. Instrum.* **12**, P03012 (2017).
12. M. G. Aartsen *et al.*, *Astrophys. J.* **835**, 151 (2017).
13. M. G. Aartsen *et al.*, *Phys. Rev. D* **89**, 102004 (2014).
14. IceCube Collaboration, *Science* **342**, 1242856 (2013).
15. M. G. Aartsen *et al.*, *Phys. Rev. Lett.* **113**, 101101 (2014).
16. M. G. Aartsen *et al.*, *Astrophys. J.* **809**, 98 (2015).
17. M. G. Aartsen *et al.*, *Astrophys. J.* **833**, 3 (2016).
18. M. G. Aartsen *et al.*, *Astrophys. J.* **807**, 46 (2015).
19. E. Massaro *et al.*, *Astrophys. Space Sci.* **357**, 75 (2015).
20. IceCube, *Fermi*-LAT, MAGIC, *AGILE*, ASAS-SN, HAWC, H.E.S.S., *INTEGRAL*, Kanata, Kiso, Kapteyn, Liverpool Telescope, Subaru, *Swift/NuSTAR*, VERITAS, and VLA/17B-403 teams, *Science* **361**, eaat1378 (2018).
21. M. G. Aartsen *et al.*, *Astropart. Phys.* **92**, 30–41 (2017).
22. IceCube Collaboration, *GRB Coordinates Network, Circular Service 21916* (2017); https://gcn.gsfc.nasa.gov/gcn3/21916.gcn3.
23. Y. T. Tanaka, S. Buson, D. Kocevski, *Astronomer's Telegram 10791* (2017); www.astronomerstelegram.org/?read=10791.
24. W. B. Atwood *et al.*, *Astrophys. J.* **697**, 1071–1102 (2009).
25. J. Aleksić *et al.*, *Astropart. Phys.* **35**, 435–448 (2012).
26. R. Mirzoyan, *Astronomer's Telegram 10817* (2017); www.astronomerstelegram.org/?read=10817.
27. R. Abbasi *et al.*, *Astrophys. J.* **744**, 1 (2012).
28. J. Braun *et al.*, *Astropart. Phys.* **29**, 299–305 (2008).
29. J. Braun *et al.*, *Astropart. Phys.* **33**, 175–181 (2010).
30. See supplementary materials.
31. K. Meagher, C. Raab, IceCube Collaboration, in *35th International Cosmic Ray Conference, Proceedings of Science* (2017); https://pos.sissa.it/301/957/pdf.
32. S. Paiano, R. Falomo, A. Treves, R. Scarpa, *Astrophys. J.* **854**, L32 (2018).
33. M. Ackermann *et al.*, *Astrophys. J.* **810**, 14 (2015).
34. M. G. Aartsen *et al.*, *Astrophys. J.* **835**, 45 (2017).
35. M. Ackermann *et al.*, *Astrophys. J.* **743**, 171 (2011).
36. S. S. Wilks, *Ann. Math. Stat.* **9**, 60–62 (1938).

ACKNOWLEDGMENTS

Funding: We acknowledge support from the following agencies. *United States:* National Science Foundation (NSF) Office of Polar Programs, NSF Physics Division, Wisconsin Alumni Research Foundation, Center for High Throughput Computing (CHTC) at the University of Wisconsin–Madison, Open Science Grid (OSG), Extreme Science and Engineering Discovery Environment (XSEDE), U.S. Department of Energy National Energy Research Scientific Computing Center, particle astrophysics research computing center at the University of Maryland, Institute for Cyber-Enabled Research at Michigan State University, and the astroparticle physics computational facility at Marquette University. *Belgium:* Funds for Scientific Research (FRS-FNRS and FWO), FWO Odysseus and Big Science programmes, and Belgian Federal Science Policy Office (Belspo). *Germany:* Bundesministerium für Bildung und Forschung (BMBF), Deutsche Forschungsgemeinschaft (DFG), Helmholtz Alliance for Astroparticle Physics (HAP), Initiative and Networking Fund of the Helmholtz Association, Deutsches Elektronen Synchrotron (DESY), and High Performance Computing cluster of the RWTH Aachen. *Sweden:* Swedish Research Council, Swedish Polar Research Secretariat, Swedish National Infrastructure for Computing (SNIC), and Knut and Alice Wallenberg Foundation. *Australia:* Australian Research Council. *Canada:* Natural Sciences and Engineering Research Council of Canada, Calcul Québec, Compute Ontario, Canada Foundation for Innovation, WestGrid, and Compute Canada. *Denmark:* Villum Fonden, Danish National Research Foundation (DNRF). *New Zealand:* Marsden Fund. *Japan:* Japan Society for Promotion of Science (JSPS) and Institute for Global Prominent Research (IGPR) of Chiba University. *Korea:* National Research Foundation of

Korea (NRF). *Switzerland:* Swiss National Science Foundation (SNSF). **Author contributions:** The IceCube Collaboration designed, constructed, and now operates the IceCube Neutrino Observatory. Data processing and calibration, Monte Carlo simulations of the detector and of theoretical models, and data analyses were performed by a large number of collaboration members, who also discussed and approved the scientific results presented here. The manuscript was reviewed by the entire collaboration before publication, and all authors approved the final version. **Competing interests:** There are no competing interests to declare. **Data and materials availability:** Data and resources are available from the IceCube data archive at www.icecube.wisc.edu/science/data. For each sample listed in Table 1, these include the events within 3° of the TXS 0506+056 source coordinates, neutrino effective areas, background rates, and other supporting information in machine-readable formats.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/147/suppl/DC1
Materials and Methods
Figs. S1 to S6
IceCube Collaboration Author List

12 February 2018; accepted 15 June 2018
10.1126/science.aat2890

FERROELECTRICITY

Metal-free three-dimensional perovskite ferroelectrics

Heng-Yun Ye[1]*†, Yuan-Yuan Tang[1]*, Peng-Fei Li[1]*, Wei-Qiang Liao[1,2], Ji-Xing Gao[1], Xiu-Ni Hua[1], Hu Cai[2], Ping-Ping Shi[1,3], Yu-Meng You[1]‡, Ren-Gen Xiong[1,2]‡

Inorganic perovskite ferroelectrics are widely used in nonvolatile memory elements, capacitors, and sensors because of their excellent ferroelectric and other properties. Organic ferroelectrics are desirable for their mechanical flexibility, low weight, environmentally friendly processing, and low processing temperatures. Although almost a century has passed since the first ferroelectric, Rochelle salt, was discovered, examples of highly desirable organic perovskite ferroelectrics are lacking. We found a family of metal-free organic perovskite ferroelectrics with the characteristic three-dimensional structure, among which MDABCO (*N*-methyl-*N*′-diazabicyclo[2.2.2]octonium)–ammonium triiodide has a spontaneous polarization of 22 microcoulombs per square centimeter [close to that of barium titanate (BTO)], a high phase transition temperature of 448 kelvins (above that of BTO), and eight possible polarization directions. These attributes make it attractive for use in flexible devices, soft robotics, biomedical devices, and other applications.

The word "perovskite" is derived from the mineral $CaTiO_3$, which was first discovered by Gustav Rose in 1839 (*1*). Magnesium silicate perovskite and calcium silicate perovskite are also the most abundant minerals in Earth's interior (*2*). Typical perovskites have a distinct three-dimensional (3D) ABX_3 crystal structure, in which the A cations are located in the 3D corner-sharing framework of BX_6 octahedrons (B represents the other cation and X is an anion) (*3*). There are hundreds of known perovskites, constituting an essential class of functional materials in optoelectronics and microelectronics. Among their numerous attractive properties, ferroelectricity, which is the ability to switch spontaneous polarization (P_s) under an applied electric field, is of particular interest for theoretical studies and is important for a variety of applications (*4*–*7*).

Inorganic perovskites, such as $BaTiO_3$ (BTO) and $Pb(Zr,Ti)O_3$, dominate applications such as ferroelectric memories, piezoelectric sensors, actuators, capacitors, and nonlinear optical devices (*8*–*10*). However, the requirements for practical materials to be energy efficient, economically inexpensive, and environmentally friendly ("triple E") motivate the exploration for nontoxic, low-cost, and simple alternatives to inorganic perovskites (*11*). Organic compounds in general and hybrid organic-inorganic perovskites in particular are attractive alternatives. These materials are lightweight, low cost, mechanically flexible, structurally tunable, and easy to process. These properties make organics attractive for materials science and technology applications (*12*, *13*).

Several different hybrid halide organic-inorganic perovskites have notable properties. Hybrid halide perovskites such as $(CH_3NH_3)PbI_3$ and $(CH_3NH_3)PbBr_3$ have attracted much attention in the solar cell and light-emitting diode (LED) fields because of their outstanding photovoltaic and luminescent performance (*14*–*19*). A few organic-inorganic perovskite ferroelectrics containing halides, cyanides, and formates were developed as competitive candidates for microelectromechanical systems (MEMS) (*20*–*25*). We recently synthesized trimethylchloromethyl ammonium trichloromanganese(II), which has a large piezoelectric response and displays ferroelectric properties comparable to those of BTO (*13*).

An underexplored but intriguing area for developing ferroelectric materials is metal-free, or organic, perovskites. These materials have an organic cation occupying the B site of the crystal. The organic composition should allow for wide chemical diversity, structural flexibility, and additional functionalities that are difficult to achieve with inorganic perovskites. The concerns with organic perovskites are that the absence of the metal ion will affect the stability of the 3D perovskites and they may not be good ferroelectric materials. In 2002, a molecular perovskite hydrate, $(C_4H_{12}N_2)(NH_4Cl_3)\cdot H_2O$, featuring the piperazinium cation $(C_4H_{12}N_2)^{2+}$ and water molecules enclosing the 3D corner-sharing $[(NH_4)Cl_6]$ network, was described (*26*). The discovery marked the beginning for metal-free perovskites, although this perovskite has an ABX_3C rather than a real ABX_3 structure (*26*, *27*). Such a system is promising for constructing noncentrosymmetric or polar structures and homochiral analogs. However, 15 years after the initial discovery of metal-free perovskites, no observations of ferroelectricity or other functionalities have been reported.

We used a molecular design strategy (see supplementary materials) (*28*) and elaborate selection of the organic cations to develop a family of metal-free ABX_3-type 3D perovskite ferroelectrics, with the general formula of $A(NH_4)X_3$ (where A is a divalent organic cation and X is Cl, Br, or I, as listed in Fig. 1A). We synthesized a total of 23 different members of this family with a range of different structures (tables S1 to S9) and phase transition temperatures (table S10). Among the first $A(NH_4)X_3$ perovskites we discovered, MDABCO–NH_4I_3 (MDABCO is *N*-methyl-*N*′-diazabicyclo[2.2.2]octonium) has a high phase transition temperature (T_0) of 448 K and a large P_s of 22 $\mu C/cm^2$, which are comparable to those of BTO (table S11). Furthermore, using piezoresponse force microscopy (PFM), we discovered the coexistence of various ferroelectric domains with eight polarization directions and evidence for the flexible rotation of polarization directions by the application of an electric field. Our strategy demonstrates the viability of high-performance metal-free ferroelectric perovskites. Without the metal element, such organics are expected to be on par with their inorganic and organic-inorganic counterparts, enhanced by the high flexibility, tunable structure-property relationships, possible solution and vacuum processes, and biocompatibility of organic materials.

General structures

Our metal-free 3D perovskite ferroelectrics are all the classical ABX_3 type without water, with 3D networks of corner-shared $(NH_4)X_6$ (X is Cl, Br, or I) octahedra that enclose cavities occupied by the divalent organic ammonium cations, and all components are held together by ionic and hydrogen-bonding interactions (Fig. 1 and figs. S1 to S18). We collected x-ray diffraction data and determined the crystal data for the structures (tables S4 to S9). Most of these metal-free perovskite structures crystallize in the ferroelectric space group at room temperature. We summarize the space groups of their ferroelectric and paraelectric phases and the Curie temperatures in table S10.

We evaluated ionic size mismatches in this large class of metal-free perovskite compounds by calculating the geometrical parameter $\alpha = (R_A + R_X)/\sqrt{2}(R_B + R_X)$. This is the tolerance factor introduced by V. M. Goldschmidt in 1926 (*29*). According to the definition, R_i is the radii of the ions in the perovskite ABX_3, where i is A, B, and X, respectively. We obtained the radii of R_A from experimental single-crystal x-ray diffraction. All of the tolerance factors for the presented divalent cations in the NH_4X_3 cavities range between 0.87 and 1.00 (fig. S19 and table S12). For values of α in the range of 0.9 to 1.0, we found 3D perovskites with greater probability, consistent with the previous analysis (*30*). Values of 0.87 to 0.89 lead to 1D hexagonal structures. A stable 3D perovskite structure does not depend solely

[1]Jiangsu Key Laboratory for Science and Applications of Molecular Ferroelectrics, Southeast University, Nanjing 211189, P.R. China. [2]Ordered Matter Science Research Center, Nanchang University, Nanchang 330031, P.R. China. [3]Institute for Advanced Interdisciplinary Research, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P.R. China.
*These authors contributed equally to this work. †Present address: Chaotic Matter Science Research Center, Jiangxi University of Science and Technology, Ganzhou, 341000, P.R. China.
‡Corresponding author. Email: youyumeng@seu.edu.cn (Y.-M.Y.); xiongrg@seu.edu.cn (R.-G.X.)

on ionic size mismatching. It also requires intermolecular interactions, such as van der Waals forces and hydrogen bonding, to become predominant in the pure organic perovskite materials. The 3D perovskite structure of DABCO–NH_4Br_3 (DABCO is *N*-*N*′-diazabicyclo[2.2.2]octonium) with α = 0.88 is a good demonstration of the comprehensive and sophisticated factors involved in the stable formation of the 3D perovskite structure.

MDABCO–NH_4I_3

Ferroelectric phase

MDABCO–NH_4I_3 has a relatively simple 3D structure (a unit cell contains just one molecular unit) (fig. S1), and thus we take it here as an example to elucidate the ferroelectric mechanism. At room temperature, MDABCO–NH_4I_3 crystallizes in the polar space group *R*3, *Z* = 1 (where Z is the number of formula units), with unit cell dimensions $a = b = c = 7.259(4)$ Å and $\alpha = \beta = \gamma = 84.767(4)°$ (where numbers in parentheses indicate uncertainty). In the cage-like unit cell, the MDABCO cation locates at the center, NH_4^+ groups locate at vertexes, and I^- ions are at the centers of the edges. According to the Curie symmetry superposition principle, the symmetry group of the specific assembly mode of the MDABCO–NH_4I_3 crystal adopts the maximal common subgroup of its composition molecular symmetry groups. In the crystal of MDABCO–NH_4I_3, $MDABCO^{2+}$ and NH_4^+ are arranged alternately along the crystallographic [1 1 1] direction (Fig. 1B), and both the intramolecular C_3 axes of these two molecules are superimposed with the crystallographic C_3 axis. As the symmetry groups of the $MDABCO^{2+}$ and NH_4^+ molecules are C_3 and T_d, respectively, the maximal common subgroup will be the C_3 point group. During the molecular assembly, the symmetry elements of $\bar{4}$ and *m* from point group T_d disappear, and both symmetry elements of 3 from point groups T_d and C_3 are preserved. The resultant polar point group symmetry C_3 should lead to the generation of spontaneous polarization along the [1 1 1] direction, because the $MDABCO^{2+}$ molecule carries a dipole moment along the molecular threefold axis. The polar point group in such metal-free frameworks is not expected when we replace $MDABCO^{2+}$ with $DABCO^{2+}$, because of the D_{3h} symmetry of $DABCO^{2+}$. While in the paraelectric phase, the disordered $MDABCO^{2+}$ molecule can be regarded as a special spherical configuration with *O* point group symmetry composed of all eight orientations possible in the ferroelectric phase. Such dynamically disordered behavior of the $MDABCO^{2+}$ molecule is responsible for the 432 symmetry of its paraelectric phase.

Paraelectric phase

At T_0 = 448 K, MDABCO–NH_4I_3 transforms into the cubic paraelectric phase, which features the space group *P*432, *a* = 7.516(13) Å, *Z* =1. In the paraelectric phase, MDABCO becomes disordered and the $(NH_4)I_6$ octahedron becomes regular (fig. S1B). MDABCO is located on the special site of 432 symmetry. The high-symmetry site of 432 requires total disorder of the organic cation, and accordingly, we modeled MDABCO with a spherical structure discarding its chemical sense (Fig. 1C). This suggests that the MDABCO cations are nearly freely rotating in the high-temperature paraelectric phase. Dynamic disorder of organic molecules in crystals is usually observed for those molecules with spherical geometries, such as tetramethylammonium, adamantane, and DABCO. Compared with that of DABCO, the shape of MDABCO deviates just slightly from the spherical geometry (Fig. 1B, blue oval). Such a molecular geometry is the key element for the generation of ferroelectricity. On the one hand, the quasispherical geometry makes molecular reorientation easy, which is necessary for ferroelectric polarization reversal (*9*). On the other hand, the lowering of molecular symmetry through the attachment of a methyl group to DABCO leads to the generation of dipole moments, whose alignment results in the symmetry breaking and contributes to the polarization of the crystal. According to the symmetry change, MDABCO–NH_4I_3 belongs to the ferroelectric species of 432*F*3 (*31*).

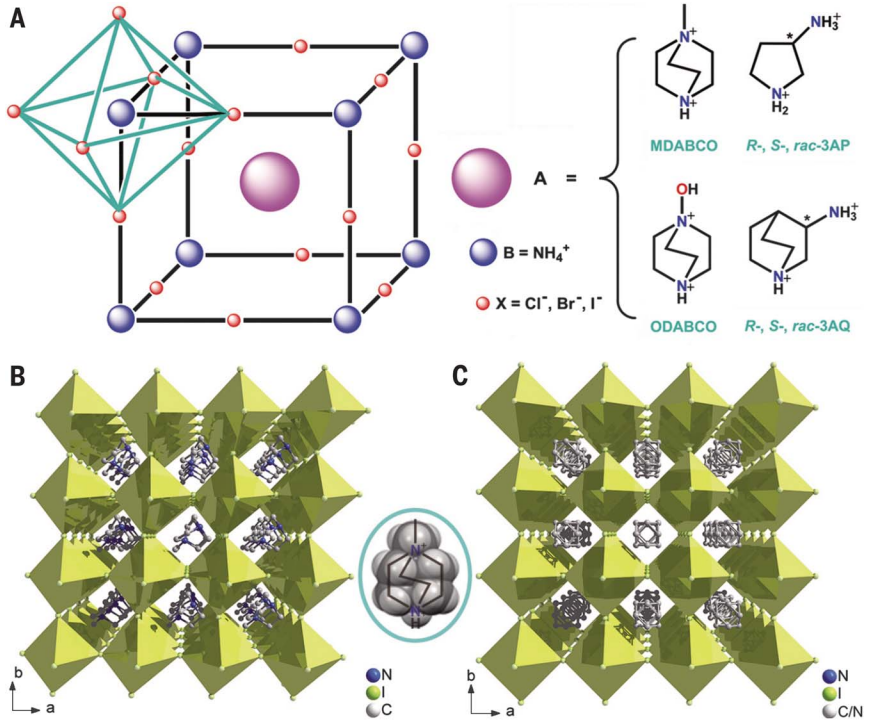


Fig. 1. Chemical and crystal structures of the metal-free $A(NH_4)X_3$ family. (**A**) Chemical structures of constituents of the metal-free 3D perovskite ferroelectrics. (**B**) The packing diagram of MDABCO–NH_4I_3 in the ferroelectric phase at 293 K. The oval to the right contains the space-fill diagram of the organic cation, showing the cationic geometry to be close to a ball. (**C**) The packing diagram of MDABCO–NH_4I_3 in the paraelectric phase at 463 K.

Because of the high symmetry of DABCO, the DABCO-containing molecular analogs, such as DABCO–$(NH_4)Br_3$ (fig. S13) and DABCO–$RbCl_3$ (*27*), do not have ferroelectricity. Additionally, the suitable size of the MDABCO ion allows it to template the 3D perovskite structure. Meanwhile, we also isolated hexagonal organic perovskites DABCO–$(NH_4)X_3$ (where X is Cl or I) and found that these materials do not show spontaneous polarization (figs. S14 and S15).

General characterizations

A pair of reversible thermal anomalies in the differential scanning calorimetry (DSC) curves reveal a phase transition around 448 K for MDABCO–NH_4I_3 (Fig. 2A). This T_0 is higher than those of the vast majority of high-temperature molecular ferroelectrics and BTO (393 K). We estimated the entropy change (ΔS) accompanying the phase transition to be ~50 J mol^{-1} K^{-1}. The main contribution of the high ΔS should be from the transition of the MDABCO cation from a freely rotating state to a static state. Meanwhile, we also confirmed the phase transition by the temperature-dependent second-harmonic generation (SHG) response (fig. S20). The detectable SHG intensity vanishes for temperatures above T_0, supporting the notion of a phase transition from the polar 3 point group into a nonpolar 432 one, which should have null SHG response under the restriction of Kleinman symmetry (*32*).

The real part (ε′) of the complex dielectric constant ($\varepsilon = \varepsilon' - i\varepsilon''$, where *i* is the imaginary unit and ε″ is the imaginary part of ε) as a function of temperature at various frequencies shows a prominent λ-shape peak anomaly around T_0 along the [1 1 1] direction (Fig. 2B), indicative of a proper ferroelectric transition. The peak value increases with decreasing frequency and reaches

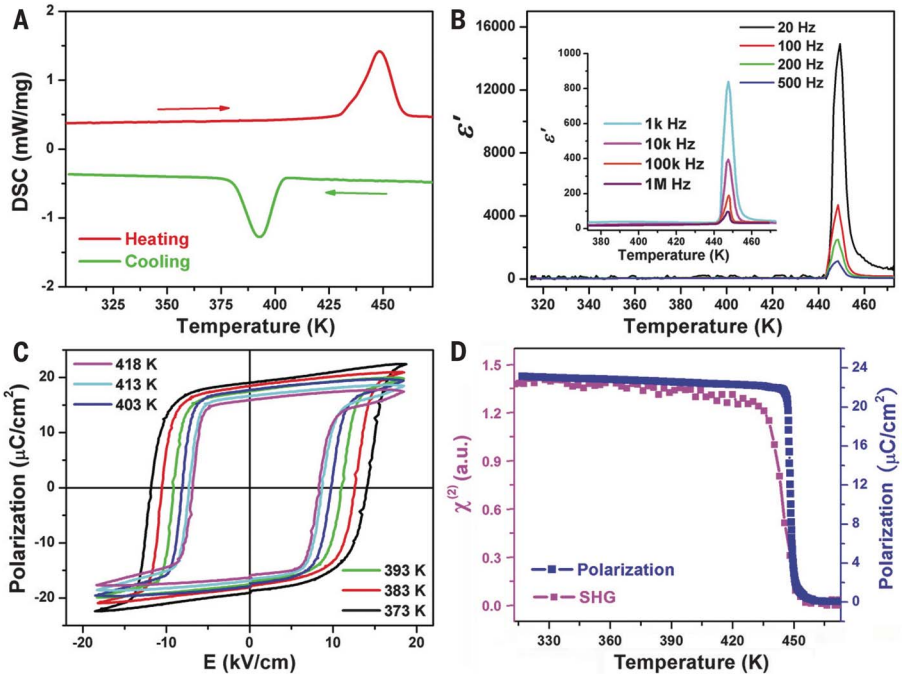


Fig. 2. Ferroelectric properties of MDABCO–NH_4I_3. (**A**) The results of DSC measurements, revealing a structural phase transition at around T_0 = 448 K. (**B**) The temperature dependence of the real part (ε′) of the complex dielectric constant at selected frequencies. k indicates thousands, and M represents a million. (**C**) *P-E* hysteresis loops measured at various temperatures. (**D**) Spontaneous polarization P_s and second-order nonlinear optical coefficient $\chi^{(2)}$ as a function of temperature. a.u., arbitrary units.

a large value of about 14,068 at 20 Hz. The dielectric response obeys the Curie-Weiss law well; that is, $\varepsilon = C/(T - T_c)$, where ε is the dielectric constant, C is the Curie constant, T ($T > T_c$) is the temperature, and T_c is the Curie-Weiss temperature. The reciprocal value of ε′ at 20 Hz as a function of temperature is linear, with the fitted Curie-Weiss constant C of 14,068 K, which is further evidence for a proper ferroelectric phase transition (fig. S21).

We used the Berlincourt method (*13*) to measure the piezoelectric coefficient d_{33} on the single crystal of MDABCO–NH_4I_3. At room temperature, we obtained a maximum d_{33} of ~14 pC/N along the [111] direction of the crystal (fig. S22). To investigate the optoelectronic potential, we also examined the optical absorption and emission properties by ultraviolet-visible absorption and photoluminescence spectroscopies, which showed that most of the perovskites tested are wide–band gap insulators and only those containing iodide can exhibit photoluminescence with the maximum emission wavelength of 561 nm (figs. S23 to 25).

Ferroelectricity

We further confirmed the ferroelectricity of MDABCO–NH_4I_3 by the measurement of polarization–electric field (*P-E*) hysteresis loops on the (1 1 1) plane recorded with the Sawyer-Tower circuit method (*33*) (Fig. 2C). We estimated the P_s value at 373 K as 19 μC/cm^2. The coercive field E_c is in the range of 6 to 12 kV/cm, comparable to that of BTO (10 kV/cm) and much smaller than that of polyvinylidene difluoride (PVDF) (500 kV/cm) (*34*). We also used pyroelectric measurements to evaluate the spontaneous polarization P_s. Spontaneous polarization appears below T_0 (Fig. 2D) and then gradually increases to a saturated value of 22 μC/cm^2, which is close to the values of 23 μC/cm^2 for diisopropylammonium bromide and 26 μC/cm^2 for BTO and greater than the 8 μC/cm^2 for PVDF and those for other organic-inorganic perovskite molecular ferroelectrics (*35*). The variation of P_s has the same trend as the effective second-order nonlinear optic susceptibility $\chi^{(2)}$, in good agreement with the Landau relationship $\chi^{(2)} = 6\varepsilon_0\beta P_s$, where β represents a coefficient almost independent of the temperature (*36*). The variations of both P_s and $\chi^{(2)}$ in the vicinity of T_0 are abrupt, revealing a typically first-order phase transition.

Spontaneous polarization

The mechanism of ferroelectric spontaneous polarization has different characteristics. Because of the ABX_3-type perovskite structure and metal-free composition, the origin of polarization in MDABCO–NH_4I_3 is more complicated than that in conventional inorganic perovskites, such as BTO, and organic-inorganic hybrid perovskites, such as $(CH_3NH_3)PbI_3$.

In the classic ABO_3-type inorganic perovskite, such as BTO, spontaneous polarization is induced by the off-centering displacement (OCD) of the small cation at the B site. Such phase transition is mainly displacive (*37*), with some contributions from order-disorder characteristics (*38*). Similar to the B-site cation in BTO, the B-site NH_4^+ ion in MDABCO–NH_4I_3 undergoes 0.2734-Å displacement from the center of the six surrounding I^- ions during the phase transition because of the octahedral tilting near T_0. The OCD of the B-site NH_4^+ ion is determined along the [111] direction. Besides OCD of the B-site NH_4^+ ion, the charge transfer through the N_{NH4}–H···I hydrogen bonds (where N_{NH4} indicates the N atom of NH_4^+ and ··· indicates interaction between H and I) should also contribute to the polarization. Those hydrogen-bonding interactions [$d_{H\cdots N}$ ($H_{NH4}\cdots N_{NH4}$ distance), 0.9 Å; $d_{H\cdots I}$ ($H_{NH4}\cdots I$ distance), 2.78 Å; ∠NHI, 159.7°; and $d_{N\cdots I}$, 3.637(9) Å] (fig. S26) are formed in the ferroelectric phase with the ordering of hydrogen bonds within the $(NH_4)I_6$ octahedron. Thus, for the B site, both charge transfer and OCD contribute to the ferroelectric polarization.

On the other hand, because of the metal-free composition and nature of molecular ferroelectrics, different from those of BTO, the A-site MDABCO also plays an important role in the generation of ferroelectricity. During the paraelectric-ferroelectric phase transition, A-site MDABCO changes from a nearly free rotation state to a static state. Such freezing of the dynamic motion tends to be discontinuous (*22*, *39*–*41*), and the ordering of the MDABCO cation leads not only to the alignment of the molecular dipole moment but also to substantial displacement. For conventional order-disorder–type inorganic compounds or organic salts, such as $NaNO_2$, KH_2PO_4, and triglycine sulfate (*9*, *10*), the order-disorder transitions are usually continuous and displacements of those dipoles are small and contribute a little spontaneous polarization accordingly. In MDABCO–NH_4I_3, the polarization direction induced by the displacement is the same as that of the molecular dipole moment of MDABCO (fig. S27). The displacement of ~0.65 Å is induced by the shift of the positive charge carried by MDABCO, which is accompanied by the elongation of the cubic cell along the body diagonal. In this sense, the ferroelectric contribution of the A site is essentially the displacive type.

The new class of metal-free perovskite ferroelectrics has a distinct polarization origin, different not only from that of conventional inorganic perovskites but also from that of organic-inorganic hybrid perovskites, such as $(CH_3NH_3)PbI_3$. The order-disorder mechanism has been found to dominate the phase transitions of those hybrid perovskites, such as $(CH_3NH_3)PbI_3$ (*42*, *43*) and $[(CH_3)_2NH_2]M(HCOO)_3$ (M is Zn, Mn, Fe, Co, or Ni) (*23*, *24*). In those organic-inorganic hybrid perovskites, the hydrogen-bonding interactions between the A-site organic ammoniums and the X-site anions play an important role. The order-disorder transitions are accompanied by the formation and breaking of hydrogen bonds. On the contrary, in MDABCO–NH_4I_3, the possible hydrogen bonding between the A-site MDABCO and the X-site I^- ion is through the N_{MDABCO}–H···I interactions, which are not suitable to form a

classic hydrogen bond because the distance of $d_{H\cdots A}$ is too large [$d_{H\cdots I}$ = 3.2098(10) Å] (*44*). Nevertheless, the weak N_{MDABCO}–H···I interactions also contribute to the spontaneous polarization (fig. S27). Because other atoms are farther away from I^- ions, the A-site MDABCO should be stabilized mostly by the ionic interactions, as supported by the density functional theory (DFT) calculations (figs. S28 to S30). The relatively low coercive electric field (E_c) may be related to the lack of strong hydrogen-bonding interactions, along with the quasispherical molecular geometry and the relatively large cavity enclosed by the octahedra.

At this stage, one can conclude that the ferroelectricity in MDABCO–NH_4I_3 is mainly ionic and influenced by hydrogen-bonding effects and that the large spontaneous polarization probably results from the parallel contributions of the displacement of charges, the alignment of the molecular dipole moments, hydrogen bond effects, and electronic contributions.

PFM characterization

Finally, we used PFM to observe the polarization ordering and switching dynamics of MDABCO–NH_4I_3 (*45*–*48*). Initial testing of single crystals (along the [1 1 1] direction) (figs. S31 and S32) indicates that the organic perovskite is dominated mainly by a large-area single-domain state with out-of-plane polarization (fig. S33). We were able to generate box-in-box bipolar domain patterns after applying a tip bias of +100 V in a 10 μm by 10 μm area and then −100 V in a centered 5 μm by 5 μm square region, demonstrating that the polarization of the ferroelectric domain can be switched back and forth in the MDABCO–NH_4I_3 crystal (Fig. 3A). In the microcrystalline state of the material, because of the existence of stress imposed by the matrix, non-180° domains with different polarization directions can appear. These twin domains can form various patterns, including stripes, which is typical for materials such as $BiFeO_3$ (Fig. 3B and fig. S34) and occurs sometimes when multiple domains intersect one another (Fig. 3C). To identify different polarization directions and analyze the angles between them, we performed comprehensive PFM studies for patterns with multiple domain intersections (figs. S35 to S38), the results of which clearly indicate the existence of eight polarization directions, corresponding well with the four polar axes deduced from the 432*F*3 transition. Most notably, we also successfully realized the rotation of the polarization direction in the MDABCO–NH_4I_3 microcrystal by using an external electric field, as the out-of-plane amplitude intensities of the central zone change from weak to strong in Fig. 3D. The emerging domain shape tends toward a hexagon, in good agreement with the growth habit of the trigonal crystal. Furthermore, a local PFM study also showed the capability for non-180° polarization rotation (fig. S39). This means that a larger spontaneous polarization can be achieved in the corresponding polycrystalline sample after electric poling, facilitating the sample's practical applications in the thin-film form.

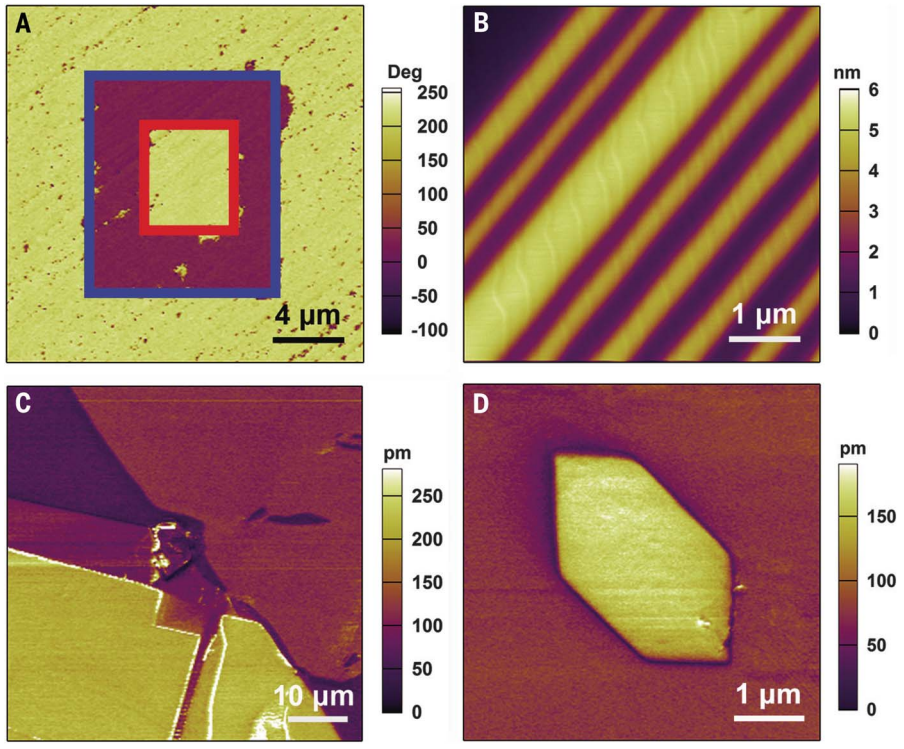


Fig. 3. PFM of MDABCO–NH_4I_3. (**A**) Out-of-plane PFM phase image recorded after writing an area of 10 μm by 10 μm with +100 V and then the central 5 μm by 5 μm square with −100 V by using a biased conductive tip in the bulk crystal. (**B** and **C**) PFM amplitude images in the as-grown microcrystals. (**D**) Out-of-plane PFM amplitude image after point polarization switching at the center point with positive bias at +40 V for 2 s in the microcrystal.

Other metal-free perovskite ferroelectrics

In addition to MDABCO–NH_4I_3, by changing the A cation and X anion according to the Goldschmidt tolerance factor, we have developed a wide variety of metal-free ABX_3-type 3D perovskite ferroelectrics (Fig. 4), such as ODABCO–NH_4Br_3 (ODABCO is *N*-hydroxy-*N*′-diazabicyclo[2.2.2]octonium) with an Aizu notation of 432*F*3, ODABCO–NH_4Cl_3 with $\bar{4}mFmm2$, *rac*-3AP–NH_4Br_3 (*rac*-3AP is *rac*-3-ammoniopyrrolidinium) with $m\bar{3}mFm$, *S*-3AQ–NH_4Br_3 (*S*-3AQ is *S*-3-ammonioquinuclidinium) with 432*F*2, and others (table S10). Chemical modifications of the constituent ions may promote the performance of the resulting ferroelectrics. For example, the spontaneous polarization will be enhanced by the introduction of ionic molecules with larger dipole moments. Moreover, using chiral template organic cations, we have successfully created optically active noncentrosymmetric organic perovskites. For example, molecular cationic isomers $(R\text{-3AP})^{2+}$ and $(S\text{-3AP})^{2+}$ and $(R\text{-3AQ})^{2+}$ and $(S\text{-3AQ})^{2+}$ can be easily embedded in ammonium halide cages because of their suitable geometric sizes and exact charge balancing. At room temperature, all these enantiomorphic perovskites crystallize in polar and chiral space group $P2_1$. The corresponding enantiomorphism relationships are verified both from single-crystal x-ray diffraction and vibrational circular dichroism (VCD) spectra (Fig. 4, C and D, and figs. S40 to S43). As expected, the infrared spectra of both enantiomorphic crystals are almost the same, whereas the corresponding VCD spectra are nearly mirror images. Hence, the introduction of organic components can lead to chemical diversity and chiral centers that may be impossible to achieve with inorganic perovskites. Homochiral molecules will form the enantiomorphic crystals of corresponding handedness that are easy to crystallize in the five enantiomorphic polar point groups (C_1, C_2, C_3, C_4, and C_6). Combining the high symmetry of 3D networks, this system can be regarded as an ideal target for the development of new multiaxial ferroelectrics, which hold great potential for use in thin-film devices because of their minimum requirements for crystallinity and specific substrates. Meanwhile, the optically active chiral compounds have a wide range of industrial and commercial applications, from uses in pharmaceutical and biological industries to applications as electro-optical elements (*49*, *50*), and ferroelectric optical activity may have a variety of potential applications.

Summary

We successfully designed a family of metal-free ABX_3-type 3D perovskites with high phase transition temperatures that are promising targets for new ferroelectric materials. In particular, MDABCO–NH_4I_3 has a large spontaneous polarization (P_s = 22 μC/cm^2) close to that of the

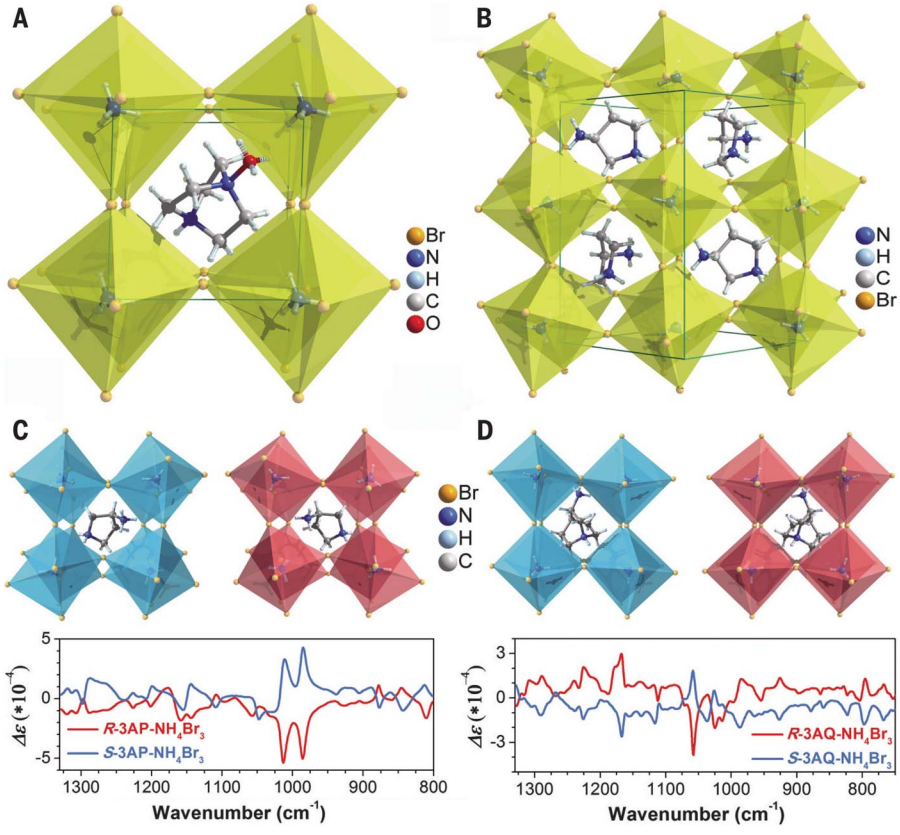


Fig. 4. Crystal structures and related chirality of the metal-free 3D perovskite ferroelectric family. Packing diagrams of (**A**) ODABCO–NH_4Br_3 and (**B**) *rac*-3AP–NH_4Br_3 in the ferroelectric phases. Crystal structures and measured VCD spectra of (**C**) *S*-3AP–NH_4Br_3 and *R*-3AP–NH_4Br_3 and (**D**) *S*-3AQ–NH_4Br_3 and *R*-3AQ–NH_4Br_3 in the ferroelectric phases. (Crystal structures of the analogs with Cl or I anions are shown in figs. S5 to S7.)

inorganic perovskite ferroelectric BTO (P_s = 26 μC/cm^2) and a high phase transition temperature (T_0 = 448 K) beyond that of BTO (390 K). The room-temperature processing not only reduces cost but also broadens the application potential by avoiding fracture during high-temperature annealing. Using PFM, we observed the coexistence of various ferroelectric domains with eight polarization directions, and we provided direct experimental proof that a flexible rotation of the polarization direction can occur in MDABCO–NH_4I_3 through the application of an electric field. The family of organic perovskites resembles inorganic perovskite ferroelectrics $PbTiO_3$, $BaTiO_3$, and $BiFeO_3$, as they can be optimized by using the Curie symmetry superposition principle and crystal engineering strategy. We believe that this class of materials, combining the outstanding ferroelectric properties and application advantages of molecular materials, holds great potential for the next generation of MEMS, flexible devices, wearable devices, and bionics.

REFERENCES AND NOTES

1. G. Rose, *Ann. Phys.* **124**, 551–573 (1839).
2. K. Hirose, R. Sinmyo, J. Hernlund, *Science* **358**, 734–738 (2017).
3. H. R. Wenk, A. Bulakh, *Minerals: Their Constitution and Origin* (Cambridge Univ. Press, 2016).
4. J. F. Scott, *Science* **315**, 954–959 (2007).
5. L. G. Tejuca, J. Fierro, *Properties and Applications of Perovskite-Type Oxides* (CRC Press, 2000).
6. L. Hu, S. Dalgleish, M. M. Matsushita, H. Yoshikawa, K. Awaga, *Nat. Commun.* **5**, 3279 (2014).
7. T. Akutagawa *et al.*, *Nat. Mater.* **8**, 342–347 (2009).
8. J. F. Scott, C. A. Paz de Araujo, *Science* **246**, 1400–1405 (1989).
9. M. E. Lines, A. M. Glass, *Principles and Applications of Ferroelectrics and Related Materials* (Clarendon Press, ed. 1, 1977).
10. F. Jona, G. Shirane, *Ferroelectric Crystals* (Pergamon, 1962), vol. 1.
11. S. Brittman, G. W. Adhyaksa, E. C. Garnett, *MRS Commun.* **5**, 7–26 (2015).
12. J. P. Correa-Baena *et al.*, *Science* **358**, 739–744 (2017).
13. Y. M. You *et al.*, *Science* **357**, 306–309 (2017).
14. N. Arora *et al.*, *Science* **358**, 768–771 (2017).
15. A. Kojima, K. Teshima, Y. Shirai, T. Miyasaka, *J. Am. Chem. Soc.* **131**, 6050–6051 (2009).
16. B. Saparov, D. B. Mitzi, *Chem. Rev.* **116**, 4558–4596 (2016).
17. J. Berry *et al.*, *Adv. Mater.* **27**, 5102–5112 (2015).
18. H. Cho *et al.*, *Science* **350**, 1222–1225 (2015).
19. Y.-H. Kim, H. Cho, T.-W. Lee, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11694–11702 (2016).
20. W. Li *et al.*, *Nat. Rev. Mater.* **2**, 16099 (2017).
21. W. J. Xu *et al.*, *J. Am. Chem. Soc.* **139**, 6369–6375 (2017).
22. Q. Pan *et al.*, *J. Am. Chem. Soc.* **139**, 3954–3957 (2017).
23. P. Jain, N. S. Dalal, B. H. Toby, H. W. Kroto, A. K. Cheetham, *J. Am. Chem. Soc.* **130**, 10450–10451 (2008).
24. P. Jain *et al.*, *J. Am. Chem. Soc.* **131**, 13625–13627 (2009).
25. D. W. Fu *et al.*, *Angew. Chem. Int. Ed.* **50**, 11947–11951 (2011).
26. C. A. Bremner, M. Simpson, W. T. A. Harrison, *J. Am. Chem. Soc.* **124**, 10960–10961 (2002).
27. L. A. Paton, W. T. Harrison, *Angew. Chem. Int. Ed.* **49**, 7684–7687 (2010).
28. Materials and methods are available as supplementary materials.
29. V. M. Goldschmidt, *Naturwissenschaften* **14**, 477–485 (1926).
30. G. Kieslich, S. Sun, A. K. Cheetham, *Chem. Sci.* **5**, 4712–4715 (2014).
31. K. Aizu, *J. Phys. Soc. Jpn.* **27**, 387–396 (1969).
32. D. Kleinman, *Phys. Rev.* **126**, 1977–1979 (1962).
33. C. B. Sawyer, C. Tower, *Phys. Rev.* **35**, 269–273 (1930).
34. S. Horiuchi, Y. Tokura, *Nat. Mater.* **7**, 357–366 (2008).
35. D. W. Fu *et al.*, *Science* **339**, 425–428 (2013).
36. H. L. Cai *et al.*, *Phys. Rev. Lett.* **107**, 147601 (2011).
37. R. E. Cohen, *Nature* **358**, 136–138 (1992).
38. R. Comes, M. Lambert, A. Guinier, *Solid State Commun.* **6**, 715–719 (1968).
39. W. Zhang, Y. Cai, R. G. Xiong, H. Yoshikawa, K. Awaga, *Angew. Chem. Int. Ed.* **49**, 6608–6610 (2010).
40. H. Y. Ye *et al.*, *J. Am. Chem. Soc.* **138**, 13175–13178 (2016).
41. P. F. Li *et al.*, *Nat. Commun.* **7**, 13635 (2016).
42. J. H. Lee, N. C. Bristowe, P. D. Bristowe, A. K. Cheetham, *Chem. Commun.* **51**, 6434–6437 (2015).
43. J. H. Lee, J. H. Lee, E. H. Kong, H. M. Jang, *Sci. Rep.* **6**, 21687 (2016).
44. G. A. Jeffrey, W. Saenger, *Hydrogen Bonding in Biological Structures* (Springer-Verlag, 1991).
45. D. Lee *et al.*, *Science* **349**, 1314–1317 (2015).
46. S. V. Kalinin *et al.*, *Annu. Rev. Mater. Res.* **37**, 189–238 (2007).
47. D. A. Bonnell, S. V. Kalinin, A. Kholkin, A. Gruverman, *MRS Bull.* **34**, 648–657 (2009).
48. V. Garcia *et al.*, *Nature* **460**, 81–84 (2009).
49. P. L. Polavarapu, *Chirality* **14**, 768–781 (2002).
50. M.-R. Goldsmith, N. Jayasuriya, D. N. Beratan, P. Wipf, *J. Am. Chem. Soc.* **125**, 15696–15697 (2003).

ACKNOWLEDGMENTS

We thank Y. Yan (University of Toledo) for help on DFT calculations. **Funding:** This work was supported by the 973 project (2014CB932103), the National Natural Science Foundation of China (91422301, 91622113, 21573041, 21574094, 21427801), and the Natural Science Foundation of Jiangsu Province (BK20150596 and BK20160029). **Author contributions:** H.-Y.Y. performed crystallographic analysis. Y.-Y.T. carried out the PFM study. P.-F.L., H.C., and P.-P.S. synthesized the sample. W.-Q.L. carried out ferroelectric characterization. J.-X.G. performed theoretical calculation. X.-N.H. carried out general characterizations. Y.-M.Y. and R.-G.X. conceived of the study and wrote the manuscript with input from other authors. **Competing interests:** All authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/151/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S43
Tables S1 to S12
References (*51–109*)

12 January 2018; accepted 6 June 2018
10.1126/science.aas9330

SYNTHETIC BIOLOGY

Programming self-organizing multicellular structures with synthetic cell-cell signaling

Satoshi Toda[1], Lucas R. Blauch[2], Sindy K. Y. Tang[2], Leonardo Morsut[1]*†, Wendell A. Lim[1]†

A common theme in the self-organization of multicellular tissues is the use of cell-cell signaling networks to induce morphological changes. We used the modular synNotch juxtacrine signaling platform to engineer artificial genetic programs in which specific cell-cell contacts induced changes in cadherin cell adhesion. Despite their simplicity, these minimal intercellular programs were sufficient to yield assemblies with hallmarks of natural developmental systems: robust self-organization into multidomain structures, well-choreographed sequential assembly, cell type divergence, symmetry breaking, and the capacity for regeneration upon injury. The ability of these networks to drive complex structure formation illustrates the power of interlinking cell signaling with cell sorting: Signal-induced spatial reorganization alters the local signals received by each cell, resulting in iterative cycles of cell fate branching. These results provide insights into the evolution of multicellularity and demonstrate the potential to engineer customized self-organizing tissues or materials.

During the development of multicellular organisms, tissues self-organize into the complex architectures essential for proper function. Even with minimal external instructions, cells proliferate, diverge into distinct cell types, and spatially self-organize into complex structures and patterns. Such self-organized structures are radically different from most human-made structures, because they are not assembled from preexisting parts that are physically linked according to a defined Cartesian blueprint. Rather, these structures emerge through a series of genetically programmed sequential events. To test and better develop our understanding of the principles governing multicellular self-organization, it would be powerful to design synthetic genetic programs that could direct the formation of custom multicellular structures (*1*–*7*).

Extensive studies of natural developmental programs have implicated many genes that control cell-cell signaling and cell morphology. Despite their molecular diversity, a common theme in these developmental systems is the use of cell-cell signaling interactions to conditionally induce morphological responses (*8*, *9*). Thus, we explored whether simple synthetic circuits in which morphological changes are driven by cell-cell signaling interactions could suffice to generate self-organizing multicellular structures.

A simple toolkit for engineering morphological programs

As a modular platform for engineering new, orthogonal cell-cell signaling networks, we focused on using the synthetic notch (synNotch) receptor system (Fig. 1A). SynNotch receptors contain the core regulatory domain of the juxtacrine signaling receptor Notch, linked to a chimeric extracellular recognition domain (e.g., single-chain antibody) and a chimeric intracellular transcriptional domain (*10*). When it recognizes its cognate ligand on a neighboring cell, the synNotch receptor undergoes cleavage of the transmembrane region, releasing the intracellular transcriptional domain to enter the nucleus and drive the expression of user-specified target genes. Thus, we can design synthetic cell-cell communication programs using synNotch circuits. SynNotch receptor-ligand pairs do not cross-talk with native signaling pathways such as Notch-Delta, or with one another, as long as they have different recognition and transcriptional domains. Here, we used two synNotch receptor-ligand pairs—an anti-CD19 single-chain antibody (scFv) receptor paired with CD19 ligand, and an anti–green fluorescent protein (GFP) nanobody receptor paired with surface GFP ligand—as orthogonal cell-cell communication channels.

We created potential developmental programs by linking synNotch signaling to two possible transcriptional outputs: (i) expression of specific cadherin molecules (E-, N-, and P-cadherins), which lead to homotypic cell-cell adhesion and differential sorting of cells expressing different classes of adhesion molecules (*11*–*13*); and (ii) expression of new synNotch ligands (Fig. 1A). Morphological sorting driven by cadherin expression can change what cells are next to each other, thus altering what synNotch signals will or will not be transmitted. Similarly, expression of new synNotch ligands can also create a subsequent stage of new cell-cell signals. Consequently, both of these outputs can propagate regulatory cascades by generating new signaling interactions between cells in the collective assembly.

We also constructed the synNotch circuits so that they drive expression of different fluorescent proteins, allowing color to indicate "differentiation" into new cell types (Fig. 1B). We expressed these synNotch circuits in mouse L929 fibroblasts, placed the cells in a low-adhesion U-bottom well (*14*), and followed their organization over time by fluorescence microscopy. L929 cells do not self-organize; normally, they only form a loose and randomly organized multicellular aggregate. We then tested whether any of the synthetic circuits we constructed from this small set of components could drive higher-order self-organization.

Engineering interacting cells that self-organize into a two-layer structure

We first focused on engineering two cell types that, when mixed, might communicate with and activate one another to induce the formation of a self-organized structure. We engineered a sender cell that expresses the synNotch ligand CD19 and blue fluorescent protein (BFP) (cell A) and a receiver cell that expresses the cognate anti-CD19 synNotch receptor and its response element (cell B). To induce cell sorting as an output of synNotch signaling, we placed the E-cadherin (Ecad) and GFP genes under the control of the synNotch-responsive promoter in the receiver cells (cell B in Fig. 2A). The circuit is represented by the following scheme:

[**cell A:** CD19] → [**cell B:** αCD19 synNotch → $Ecad^{hi}$ + GFP]

As predicted, when cocultured with A-type sender cells, B-type receiver cells were activated to express Ecad and GFP (C-type cell phenotype). Subsequently, the green (GFP) C-type cells self-sorted to form a tight inner core, resulting in a well-defined two-layer structure (Fig. 2, B and C). Without induction of Ecad, the A- and B-type cells remained well-mixed (fig. S1A). When the synNotch signaling was inhibited by the γ-secretase inhibitor (2*S*)-*N*-[(3,5-difluorophenyl)acetyl]-L-alanyl-2-phenylglycine 1,1-dimethylethyl ester (DAPT), which blocks synNotch cleavage and signaling, sorting into two layers did not occur, as the inhibitor blocked the Ecad induction response in B-type cells (fig. S1B).

Using a bidirectional signaling cascade to engineer a self-organizing three-layer structure

To create more a complex structure, we added an additional layer of reciprocal cell-cell signaling to the above two-layer circuit (Fig. 2D). We engineered

[1]Department of Cellular and Molecular Pharmacology, Howard Hughes Medical Institute, and Center for Systems and Synthetic Biology, University of California, San Francisco, CA 94158, USA. [2]Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, USA.
*Present address: Eli and Edythe Broad CIRM Center for Regenerative Medicine and Stem Cell Research, University of Southern California, Los Angeles, CA 90033, USA.
†Corresponding author. Email: leonardo.morsut@med.usc.edu (L.M.); wendell.lim@ucsf.edu (W.A.L.)

the receiver (B-type) cell so that in addition to inducibly expressing Ecad, it also inducibly produced surface-tethered GFP as a synNotch ligand (GFP_{lig}). This modified form of GFP is constructed by fusion with the platelet-derived growth factor receptor transmembrane (PDGFR TM) domain (*10*). Surface-tethered GFP served as both a fluorescent reporter of the new cell type and a ligand for a secondary synNotch receptor with the cognate anti-GFP nanobody binding domain. In the sender cells, which constitutively express BFP and CD19 ligand, we also expressed the anti-GFP synNotch receptor, which when induced would drive expression of a low amount of E-cadherin ($Ecad^{lo}$) fused with an mCherry reporter for visualization. Thus, the interaction between this pair of cell types can in principle yield a two-step cascade of reciprocal signaling: In the first step, CD19 on cell A activates anti-CD19 synNotch in cell B to induce expression of a high level of E-cadherin ($Ecad^{hi}$) and the GFP_{lig}. In the second step, the GFP_{lig} on cell B can reciprocally activate the anti-GFP synNotch receptors in neighboring A-type cells to induce a low level of E-cadherin alongside the mCherry reporter. In this case, the A-type cell starts out as a sender cell but later becomes a receiver cell. The circuit is represented by the following scheme:

[**cell A:** CD19; αGFP synNotch]
→ [**cell B:** αCD19 synNotch → $Ecad^{hi}$ + GFP_{lig}]
→ [**cell A:** αGFP synNotch → $Ecad^{lo}$ + mCherry]

This circuit was predicted to form a three-layer structure: a green internal core ($Ecad^{hi}$ + GFP) with the highest homotypic adhesion, an outer layer of blue cells (no Ecad), and a new population of red ($Ecad^{lo}$ + mCherry) cells at a middle interface layer (Fig. 2E). We first engineered and established cell A and cell B lines from single-cell clones, and then confirmed that they showed synNotch-driven expression of high or low amounts of Ecad along with the appropriate marker fluorescent proteins (fig. S2A).

When we cocultured 200 A-type cells and 40 B-type cells, a three-layer structure was robustly formed, with a development process that required ~20 hours to fully unfold (Fig. 2F, fig. S2B, and movie S1). The structure emerged in a stereotypical stepwise fashion: induction of the green cells, sorting to form an inner core, and then the formation of a red middle layer. Here, the cascade of cell sorting and reciprocal signaling from the green core cells drives fate branching of the original A-type sender cells into two distinct fates (red and blue). Thus, this program has substantial ordering power: (i) The program generates an increased number of cell types (two cell genotypes become three phenotypic cell types), and (ii) the program leads to spatial sorting into three distinct compartments. This change represents a decrease in entropy relative to the starting point of a random mixture of two cell types, as shown in the cell lineage map (Fig. 2E). Many of these features of increased self-ordering observed in this engineered assembly mimic the behavior of natural developmental systems, such as the simple formation of distinct progenitor cell types in early embryogenesis (*15*, *16*).

The observed self-organization could be blocked by disrupting either synNotch signaling or cadherin expression. When we blocked cell-cell signaling with an inhibitor of synNotch signaling (DAPT), we observed no increase in cell types and no cell sorting into distinct layers (Fig. 2G and fig. S3B). When we removed E-cadherin expression from the system (fig. S3A), the assembled cells induced expression of the GFP and mCherry markers, but the different cell types remained randomly mixed (Fig. 2G). Thus, the interlinking of signaling and cell sorting is required for cell fate divergence and spatial ordering.

Synthetic assembly is robust, reversible, and self-repairing

To see how reproducibly the synthetic cell-cell signaling program could drive three-layer formation, we followed 28 independent replicate cocultures starting with 200 A-type cells and 40 B-type cells (Fig. 3A). In most wells (57%), cells formed a single three-layer spheroid. In other wells, we observed "twin" multicore three-layer spheroids (21%) or multiple (separate) three-layer spheroids in the same well (11%).

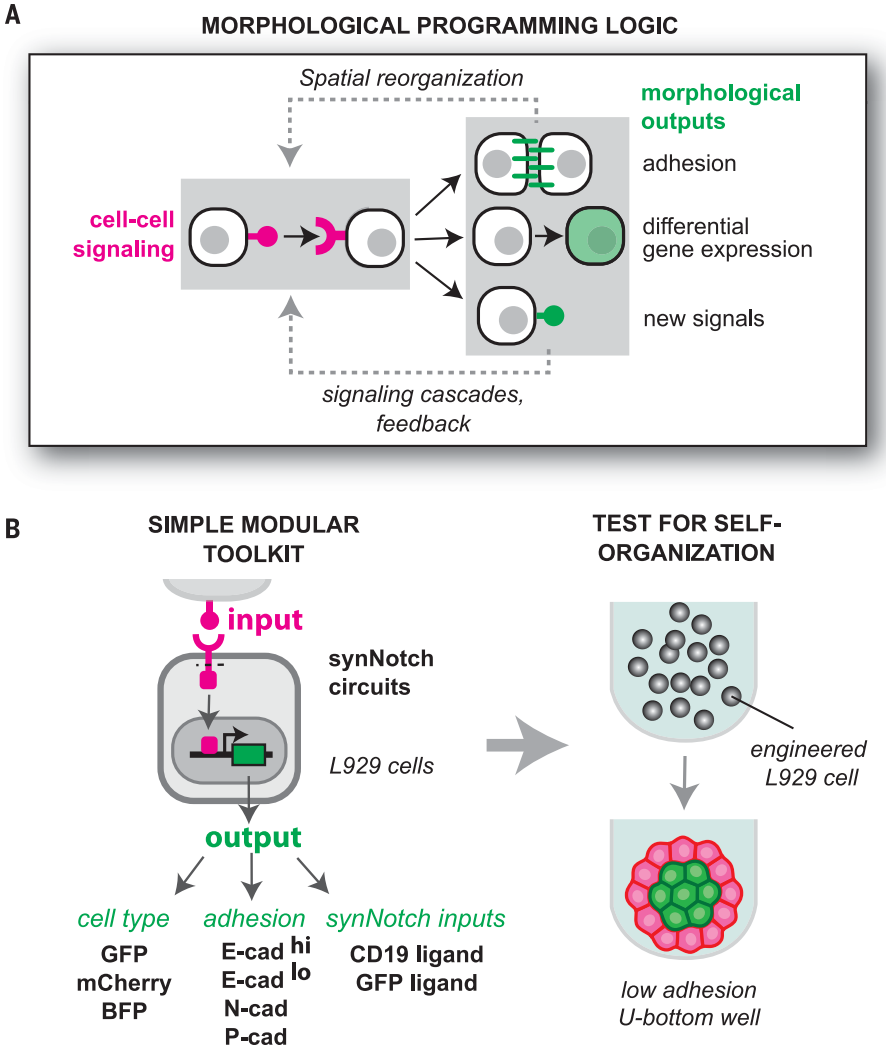


Fig. 1. Engineering cell-cell communication networks to program synthetic morphogenesis. (**A**) Design logic underlying our synthetic morphogenesis circuits. Engineered cell-cell signaling is used to drive changes in cell adhesion, differentiation, and production of new cell-cell signals. These outputs can subsequently be propagated to generate new cell-cell signaling relationships. (**B**) Molecular components used for assembly of simple morphological circuits. We used two synNotch ligand-receptor pairs (surface ligands CD19 and GFP) for cell signaling, three fluorescent proteins as markers of "differentiation," and several cadherin molecules (expressed at different levels) as morphological outputs. Engineered circuits are transduced into L929 fibroblast cells, placed in defined numbers in low-adhesion U-bottom wells, and screened by microscopy for spatial self-organization.

Thus, the overall three-layer architecture of green, red, and blue cells was robustly generated in ~90% of the cultures. A 3D reconstruction image of three-layer structure is shown in Fig. 3B and movie S1. Three-layer formation was robust to variation in the initial number or ratio of starting cells (fig. S2C). Only when we used a low number of starting A-type cells did we begin to see formation of two-layer structures (green and red only), because all the A-type cells were converted to Ecad^{lo} cells (i.e., the number of A cells was limiting).

In many cases, natural self-organized tissues have an ability to regenerate after injury (*17*). To test how this three-layer structure would respond to injury, we cut the structure into two fragments with a microfluidic guillotine system (*18*) (movie S2). Immediately after cleavage, the GFP-positive core cells were exposed to the surface, but within 24 hours, the green core cells were re-enveloped by the red layer, regenerating the spherical three-layer structure (Fig. 3C). To further test the reversibility of the self-assembled three-layer structure, we added the synNotch inhibitor DAPT to preformed structures. The layered structure and distinct cell types were totally disrupted within 3 days of treatment; hence, this dynamically maintained structure can be disassembled by turning off cell-cell signaling (Fig. 3D).

A single-genotype circuit that induces cell fate bifurcation and spatial ordering into a two-layer structure

We also wanted to explore whether we could program self-organizing structures that could start from a single cell type. Alternative bistable

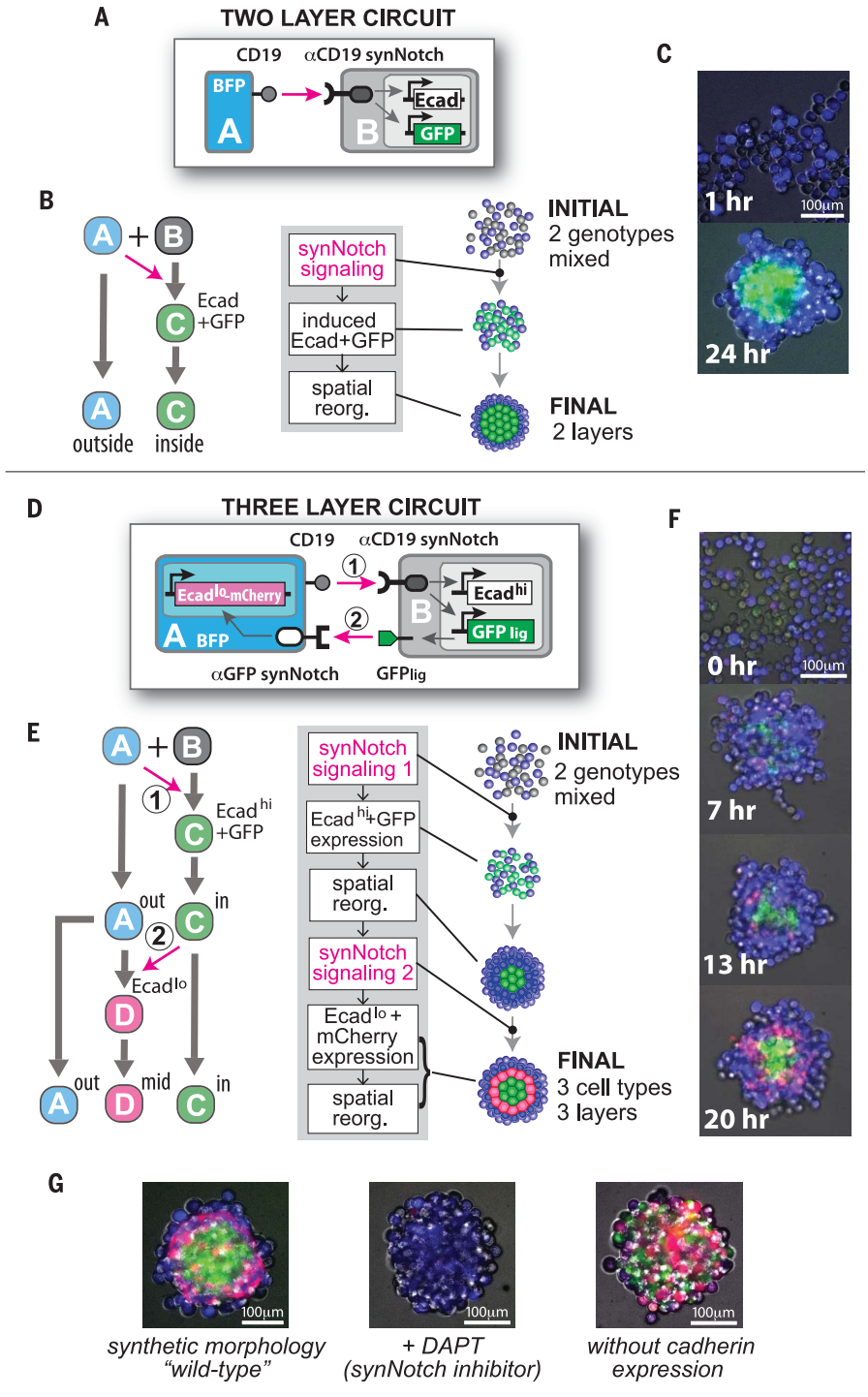


Fig. 2. Engineering self-organizing multilayered spheroids. (**A** to **C**) Two-layer circuit. (A) An A-type sender cell expressing CD19 ligand induces a B-type receiver cell to express E-cadherin and GFP. (B) SynNotch cell-cell signals drive receiver cells to express E-cadherin (Ecad), which leads to their segregation into a central layer. The system starts with two disordered cell genotypes but organizes to form a structure with two distinct spatial compartments. (C) Images of the spheroid at 1 and 24 hours. See fig. S1 for other data. (**D** to **G**) Three-layer circuit. (D) An A-type cell can send signals to a B-type cell using CD19 ligand, which induces expression of E-cadherin (high expression) and GFP$_{lig}$ (surface-expressed GFP). The induced B-type cell can then send reciprocal signals to the A-type cell; GFP$_{lig}$ serves as ligand to stimulate anti-GFP synNotch receptor expressed in the A-type cell. This reciprocal interaction is programmed to drive a low level of E-cadherin and mCherry. (E) Cell fate diagram showing how this program drives a two-step differentiation process in which the A→B synNotch signal first drives conversion of B-type cells to C-type cells that self-adhere and sort to the center of the structure. The sorted C-type cells then present the C→A synNotch signal (driven by GFP$_{lig}$) to convert spatially adjacent A-type cells into the middle-layer D-type cell (mCherry and low-level E-cadherin expression). A-type cells bifurcate into two phenotypes, depending on their spatial proximity to the C-type cells in the core of the structure. Here, the system starts with two disordered cell genotypes but self-organizes into three distinct cell phenotypes organized into three spatially distinct compartments. (F) Images from the development of the three-layer system from 0 to 20 hours. See fig. S2 and movie S1 for other data and time-lapse videos. (G) Formation of the three-layer structure is disrupted if synNotch signaling is inhibited (using DAPT, a γ-secretase inhibitor) or if cadherins are not driven as outputs. See fig. S3 for more information.

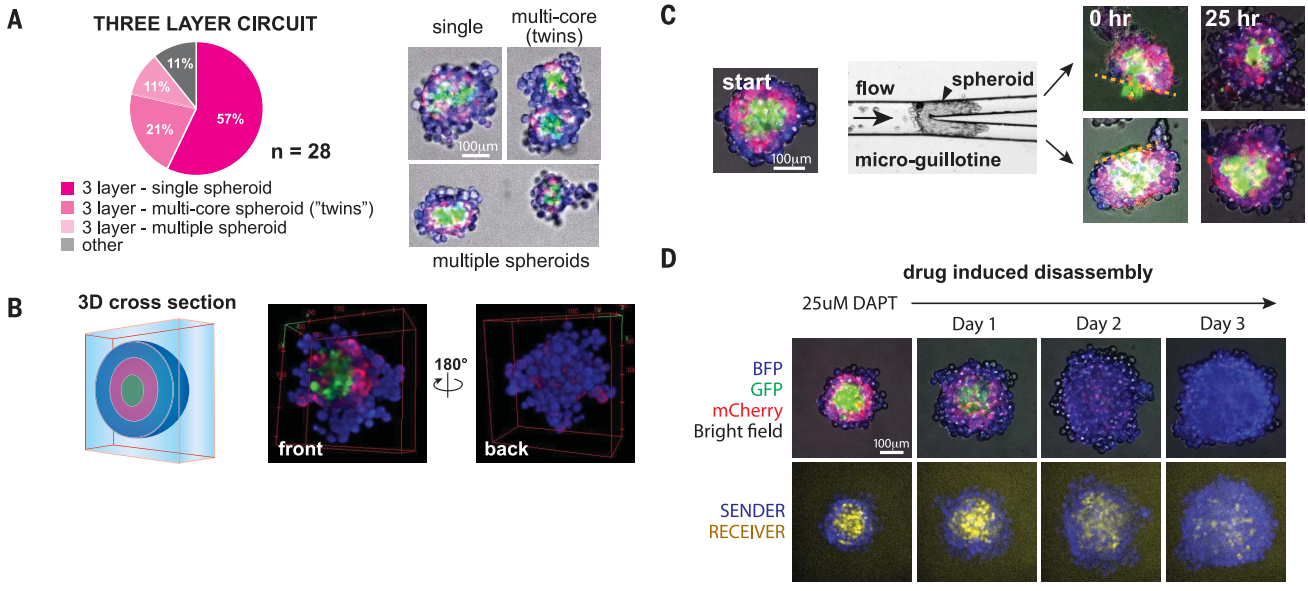


Fig. 3. Three-layer self-organized structure is robust, reversible, and self-repairing. (**A**) Distribution of structures generated in 28 independent wells (starting with 200 A-type cells and 40 B-type cells). About 90% of the wells showed formation of three-layer structures; the majority of these showed one spheroid per well, with the remainder showing either twinned spheroids or multiple independent three-layer spheroids. Example images of these structural subtypes are shown at the right. (**B**) Three-dimensional confocal reconstruction of a three-layer structure cross section, shown from two views. See movie S1 for full rotational view of the 3D structure. (**C**) Self-repair of a cleaved three-layer structure. The preformed spheroid was cleaved using a microfluidic guillotine, and the two resulting fragments were observed for 25 hours. The frames at 0 hours show the two fragments, with a dotted line indicating the cleavage plane that exposes the internal core of the spheroid. Images at 25 hours show self-repair of the spherical three-layer structure. (**D**) The structure is reversible if treated with the synNotch inhibitor DAPT. Within 3 days, the differentiation and spatial organization of cells disappeared. Original A- and B-type cells became randomly organized.

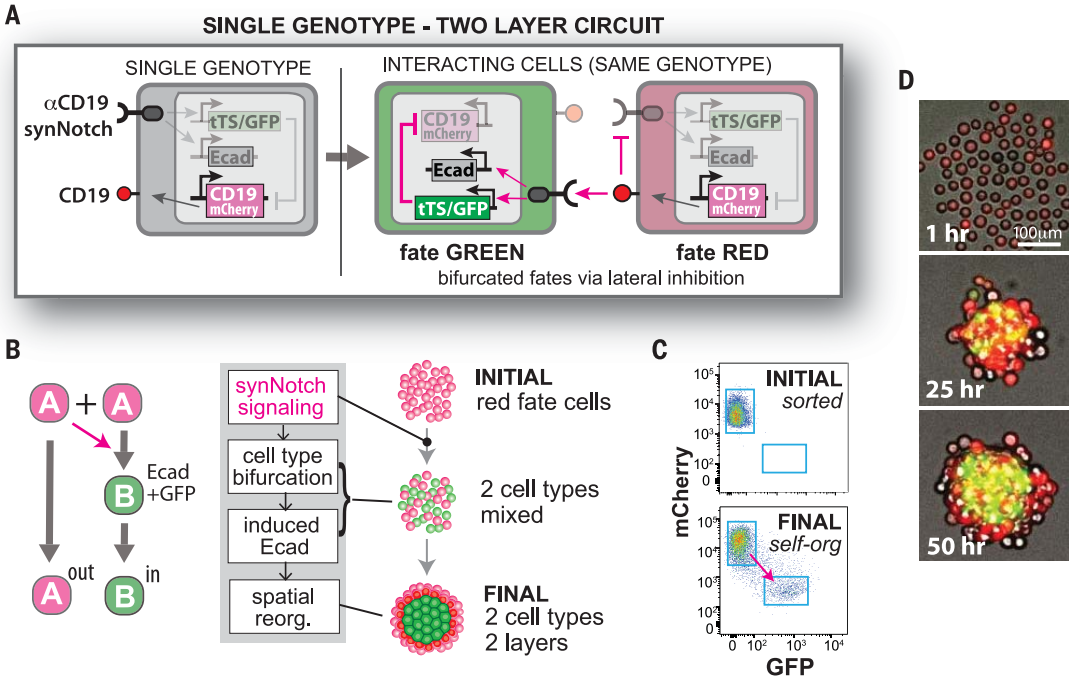


Fig. 4. Single-genotype circuit that induces fate bifurcation and spatial ordering into a two-layer structure. (**A**) Design of single-genotype circuit with lateral inhibition between sender (CD19^{+}) and receiver (antiCD19-synNotch–activated) states. The cell encodes both CD19 and anti-CD19 synNotch, but activated synNotch receptor drives expression of tet repressor (tTS), which inhibits CD19 expression. Thus, neighboring cells will drive each other into opposite states indicated by red and green fluorescent markers (fate RED and GREEN). (**B**) E-cadherin expression driven from the synNotch-activated promoter. An initially homogeneous population of red cells undergoes bifurcation into RED fate and Ecad-positive GREEN fate by lateral inhibition, and GREEN-fate cells are finally sorted inside to form an inner core. The system starts with a single-genotype population but is expected to organize into a two-layer structure. (**C**) Purification of a homogeneous population by sorting for mCherry^{high}/GFP^{low} cells. When allowed to communicate through lateral inhibition, the cells rebifurcate into two distinct fluorescently labeled populations (bottom). See fig. S4 and supplementary materials for more information on how the lateral inhibition circuit was constructed and executed. (**D**) Development of the single-genotype two-layer structure. Time frames are shown at 1, 25, and 50 hours, showing initial cell fate bifurcation followed by formation of a stable two-layer structure. See fig. S5 for more information and movie S3 for time-lapse video.

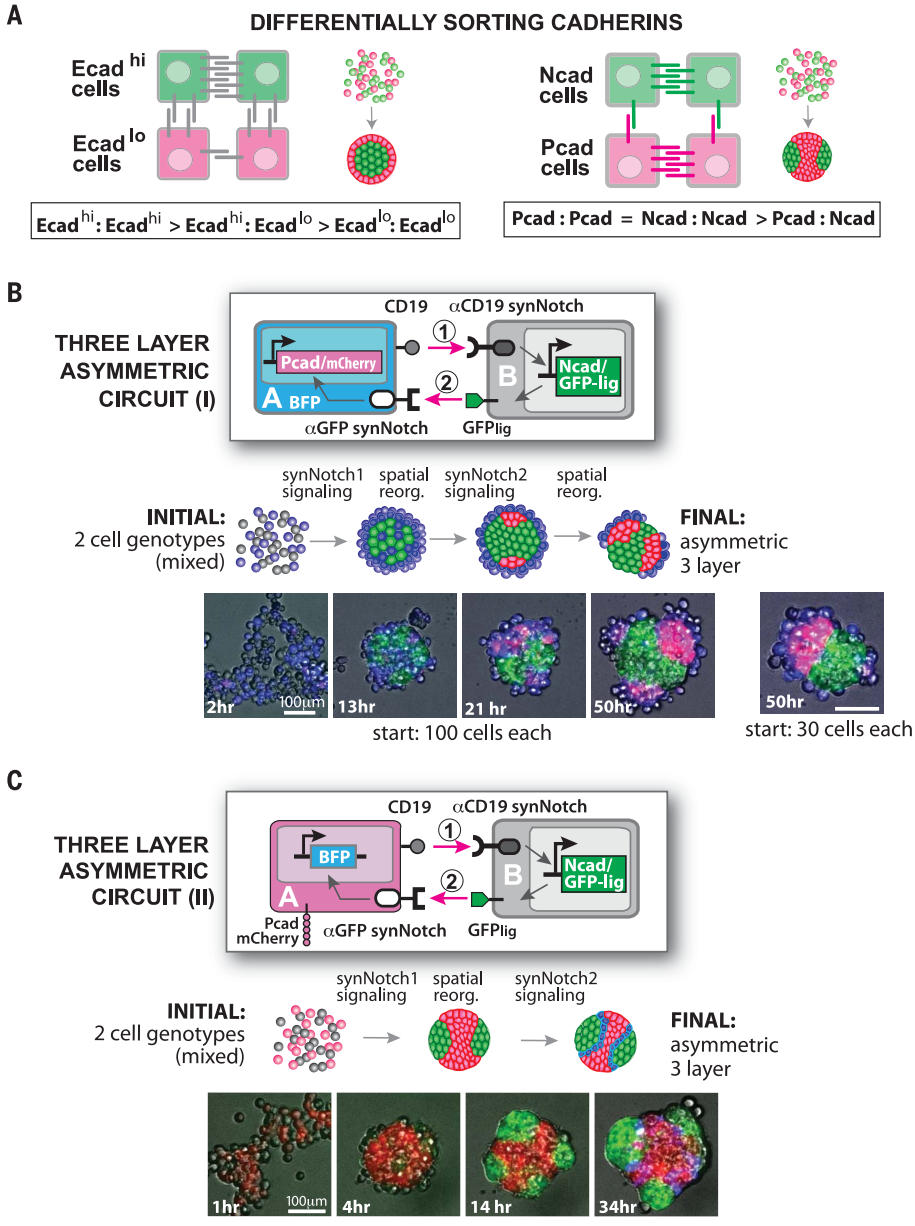


Fig. 5. Programming spherically asymmetric structures by inducing differentially sorting adhesion molecules. (**A**) Logic of deploying alternative adhesion outputs to generate different spatial structures. In the spherically symmetric structures of Figs. 2 to 4, we used high and low levels of Ecad expression to define different populations of cells. High- and low-Ecad populations lead to sorting into concentric shells, because $Ecad^{lo}$ cells still prefer to bind $Ecad^{hi}$ cells. In contrast, two cell populations that express either Ncad or Pcad will sort into distinct compartments (nonconcentric) because each of these cadherins prefers homotypic self-association to heterotypic cross-association. (**B**) Three-layer asymmetric circuit I, with the same architecture as that shown in Fig. 2, except that B-type cells are induced to express Ncad and A-type cells are induced to express Pcad. In phase II of the development (reciprocal B→A signaling), the A-type cells become red and self-sort to form one to three external poles (with unactivated A-type cells associated at their periphery). The starting population included 100 cells of each type. When we started with only 30 cells of each type (right image), we reproducibly generated single-pole structures. See fig. S7 and movie S4 for more information, time-lapse videos, and 3D structure. (**C**) Three-layer asymmetric circuit II. An A-type cell constitutively expresses Pcad and mCherry as well as CD19 ligand. B-type cells recognize CD19 with anti-CD19 synNotch receptor, which drives expression of Ncad and GFP_{lig}. In reciprocal signaling, GFP_{lig} drives induction of a BFP marker in A-type cells. Here, the red A-type cells first form a central core and the induced green B-type cells form polar protrusions. A third cell type (blue) forms at the boundary between the red core and the green protrusions. See fig. S8 and movie S5 for more information, time-lapse video, and 3D structure. Information on other structures using different cadherin pairs is shown in figs. S9 and S10 and movies S6 and S7.

cell fates can be generated from a single starting cell genotype through a mechanism known as lateral inhibition (*19*). For example, cross-repression between Notch receptor and its ligand in neighboring cells can result in a bistable, checkerboard fate pattern, where individual cells bifurcate into either $Notch^{active}$-$ligand^{low}$ or $Notch^{inactive}$-$ligand^{high}$ states (*20*, *21*). We built an analogous lateral inhibition circuit using synNotch cross-repression in L929 cells (fig. S4B). Each cell encoded both CD19 (ligand) and the anti-CD19 synNotch receptor, but these are antagonistic to each other because the synNotch receptor induces expression of the Tet repressor (tTS), which can repress CD19 expression (controlled by a TetO promoter). Thus, if synNotch is stimulated by a neighboring cell with high CD19 expression, it will repress CD19 ligand expression, thereby forcing cells to choose between either a sender or receiver fate. CD19 and tTS expression were monitored by mCherry and GFP, respectively (expressed in linked transcriptional cassettes through a ribosomal skipping porcine teschovirus-1 2A sequence). We established multiple clones that bifurcated spontaneously into two populations of mCherry or GFP-positive cells (fig. S4B; see supplementary materials for details of how we established lateral inhibition lines). These cell lines consistently reestablished the two phenotypic states, even when starting with a pure sorted population of either the red or green state (Fig. 4C and fig. S4B).

To produce a spatially ordered structure from a single cell type, we then functionally combined two different organizational circuit modules: this bifurcating cell fate circuit and the self-organized E-cadherin–driven two-layer circuit (Fig. 2A). To construct such a composite circuit, we expressed E-cadherin from the synNotch-driven promoter (in addition to inducing expression of the tet repressor) (Fig. 4A and fig. S5A). The objective was to start with a single cell type and observe self-driven fate bifurcation followed by self-driven sorting into two layers.

To track how the system developed from a single cellular phenotype, we sorted red-fate cells ($CD19_{high}$), placed 100 cells in each well, and followed the development of the spheroid by time-lapse microscopy. These cells developed into a spheroid in which the cells first underwent bifurcation into a red-green checkerboard pattern and then, over the course of hours, formed a two-layer structure with green cells inside and red cells outside (fig. S5, B and C). These two-layer structures were stable for 100 hours. Addition of the Notch signaling inhibitor DAPT prevented fate bifurcation (fig. S5C). But after removal of the drug and re-sorting, the cells remained bipotent; they could still bifurcate and

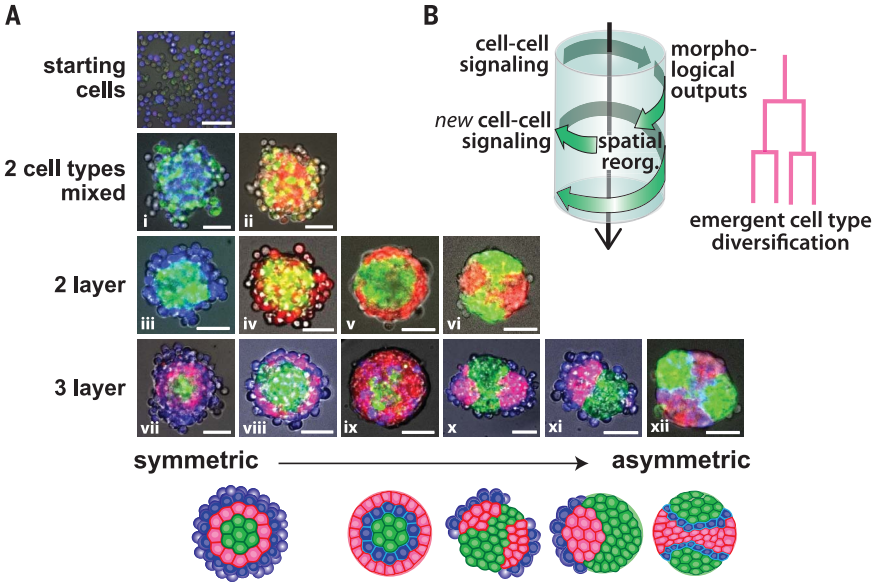


Fig. 6. Gallery of different self-organizing multicellular structures that can be programmed using the simple synNotch→adhesion toolkit. (**A**) Gallery of spatially organized behaviors generated in this work, organized by resulting number of cell types and spatially distinct compartments as well as by increasing asymmetry. See table S1 for details of the construction of these 12 structures. Diagrams of several of the different three-layer structures are shown schematically below. (**B**) These synthetic developmental systems share the common principles in which cascades of cell-cell signaling, linked by morphological responses, lead to increasing diversification of cell types. As signaling drives morphological changes and reorganization, new cell-cell interactions arise, resulting in increasingly distinct positional information encountered by each cell in the structure.

reform the two-layer structure (Fig. 4D, fig. S5C, and movie S3). Thus, we can engineer synthetic programs in which a single cell genotype bifurcates and spatially self-organizes into multiple layers.

Programming spherically asymmetric structures by inducing differentially sorting cell adhesion molecules

Another key feature of natural morphogenesis is symmetry breaking, used repeatedly during development to generate body axes and elaborate an initially uniform ball of cells (*22*, *23*). The structures described above are all spherically symmetric, but we could program asymmetric structure formation with the same signaling cascade circuit by simply changing the adhesion molecules that were expressed.

To build the spherically symmetric three-layer structure described above (Fig. 2D), we programmed different subsets of cells to express different amounts of the same adhesion molecule (E-cadherin), which generates spherically symmetric concentric layers (because $Ecad^{lo}$ cells still prefer to interact with $Ecad^{hi}$ cells; see relative interaction energies in Fig. 5A). However, if cells express different cadherins that have high homotypic affinity but low heterotypic affinity, they phase-separate into two spatially distinct populations (Fig. 5A). N-cadherin (Ncad) and P-cadherin (Pcad) have high homotypic affinity (Ncad-Ncad and Pcad-Pcad) but low heterotypic affinity (Ncad-Pcad) (*24*), so we used the combination of Ncad and Pcad expression to try to drive asymmetric sorting and structure formation (fig. S6).

We introduced Ncad and Pcad as morphological outputs in the basic three-layer circuit. First, CD19 synNotch signaling from cell A induced expression of Ncad and GFP_{lig} in cell B; second, the induced GFP_{lig} on cell B reciprocally activated anti-GFP synNotch in the adjacent subpopulation of A cells, driving Pcad expression (Fig. 5B, fig. S7, and movie S4). When we cultured 100 cells each of type A and B together, we observed a stereotypical developmental sequence: By 13 hours, B-type cells expressed both Ncad and GFP_{lig}, and by 21 hours, A-type cells adjacent to B-type cells began to express Pcad and mCherry. Because of the resulting self-segregation of the Ncad- and Pcad-expressing cells, the ensemble self-organized into a nonspherically symmetric three-layer structure (green, red, blue) with one to three distinct poles of mCherry (Pcad) cells. A-type cells (blue) not activated through their anti-GFP synNotch receptors were associated with the outer surface of these poles.

When we initiated cultures with a smaller number of starting cells (30 cells each of type A and B), the ensemble reproducibly formed a single-pole asymmetric structure (a single cluster of red cells instead of multiple clusters), consistent with many examples of polarized organization in which a smaller starting size minimizes the chance of initiation of multiple independent poles (Fig. 5B, fig. S7, and movie S4) (*25*). Thus, we could reliably program systems that would form three-layer asymmetric or polarized structures.

We designed other circuits that induced alternative types of asymmetric structures with the same Ncad-Pcad output combination but were regulated in different sequential programs. In the circuit shown in Fig. 5C, cell A was similar to the above example (it expressed CD19 ligand and anti-GFP synNotch receptor driving expression of BFP), except that it also constitutively expressed Pcad [connected with mCherry via an internal ribosome entry site (IRES) sequence]. Cell B was the same as in Fig. 5B (it expressed anti-CD19 synNotch receptor that induced Ncad and GFP_{lig} expression). When cultured together, the Pcad-expressing A cells (red) immediately formed an adherent aggregate (4 hours); then, after 14 hours, Ncad and GFP_{lig} were expressed in B-type cells, leading to the formation of polarized B-type protrusions (green) segregated from the A-type cells (red). Finally, at 34 hours, A-type cells at the interface with B-type cells were activated by GFP-synNotch signaling to turn on BFP, resulting in a thin boundary layer of blue cells between the polarized red and green regions (time-lapse and 3D reconstruction image shown in fig. S8 and movie S5). Additional types of combinatorial circuits using different cadherin pairs are shown in figs. S9 and S10 and in movies S6 and S7.

These results confirmed that we can build various self-organizing structures that break spherical symmetry by inducing distinct self-segregating adhesion molecules in different subpopulations of cells. Initial conditions with small cell numbers can reproducibly yield structures with a single polar axis. Moreover, we can generate many different three-layer morphological structures by altering the combinations of adhesion molecules used and by altering at what stage in the circuit they are expressed (Figs. 5 and 6).

Minimal intercellular communication programs can drive synthetic self-organizing cellular structures

Figure 6A and table S1 summarize the various self-organizing synthetic structures we programed with our minimal logic of controlling cell adhesion (cadherin expression) through cell-cell communication (synNotch signaling). The diversity and complexity of these structures, and the robustness with which they are formed, illustrate the ordering power of even these highly simplified cell-cell signaling programs. In all of these systems, we observed a cyclic sequence of events in which initial cell signaling interactions induced morphological rearrangements, which in turn generated new cell-cell interactions and new morphological refinements (Fig. 6B). Complex structures emerge because these cell-cell signaling cascades drive increasing cell type diversification.

These diverse emergent structures can form even in the absence of many of the molecular components normally used in natural developmental systems. For example, these circuits do not incorporate diffusible morphogens for cell-cell communication, irreversible cell fate commitment, or direct regulation of cell proliferation, death, or motility (*8*, *26*–*29*). It is likely that the

synthetic platforms used here could be extended to include many of these additional elements to generate even more sophisticated engineered self-organizing multicellular structures (*30*–*35*).

The observation that even minimal circuits that link cell-cell signaling to adhesion can lead to the formation of defined self-organizing structures may help to explain the general principles by which multicellular organisms could have evolved. Choanoflagellates, the closest single-cell relatives of metazoans, have both primitive cadherin and notch genes (*36*). The cadherin genes are thought to have originally functioned to trap prey bacteria in the environment and may have later been co-opted for cell-cell adhesion (*37*, *38*). In some choanoflagellate species, environmental signals from prey bacteria can induce the formation of multicellular assemblies (*39*, *40*). It seems plausible that cell-to-bacteria adhesion transitioned to cell-cell adhesion, and that bacteria-to-cell signaling transitioned to cell-cell signaling. During the course of evolution, these systems may have begun to regulate one another, providing a starting point for circuits capable of driving formation of complex multicellular structures.

More generally, these findings suggest that it may be possible to program the formation of synthetic tissues, organs, and other non-native types of dynamic, multicellular materials. We may be able to apply tools like synNotch, perhaps enhanced by an even larger toolkit of modular developmental signals, to construct customized self-assembling tissue-like biomaterials of diverse types. These tools and approaches also provide powerful tools to systematically probe and better understand the principles governing self-organization and development.

REFERENCES AND NOTES

1. J. Davies, *Development* **144**, 1146–1158 (2017).
2. M. Elowitz, W. A. Lim, *Nature* **468**, 889–890 (2010).
3. B. P. Teague, P. Guye, R. Weiss, *Cold Spring Harb. Perspect. Biol.* **8**, a023929 (2016).
4. A. Kicheva, N. C. Rivron, *Development* **144**, 733–736 (2017).
5. S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, R. Weiss, *Nature* **434**, 1130–1134 (2005).
6. S. S. Jang, K. T. Oishi, R. G. Egbert, E. Klavins, *ACS Synth. Biol.* **1**, 365–374 (2012).
7. M. Rubenstein, A. Cornejo, R. Nagpal, *Science* **345**, 795–799 (2014).
8. L. Wolpert, *J. Theor. Biol.* **25**, 1–47 (1969).
9. D. Gilmour, M. Rembold, M. Leptin, *Nature* **541**, 311–320 (2017).
10. L. Morsut *et al.*, *Cell* **164**, 780–791 (2016).
11. A. Nose, A. Nagafuchi, M. Takeichi, *Cell* **54**, 993–1001 (1988).
12. D. Duguay, R. A. Foty, M. S. Steinberg, *Dev. Biol.* **253**, 309–323 (2003).
13. R. A. Foty, M. S. Steinberg, *Dev. Biol.* **278**, 255–263 (2005).
14. M. Vinci *et al.*, *BMC Biol.* **10**, 29 (2012).
15. C. Chazaud, Y. Yamanaka, *Development* **143**, 1063–1074 (2016).
16. S. E. Harrison, B. Sozen, N. Christodoulou, C. Kyprianou, M. Zernicka-Goetz, *Science* **356**, eaal1810 (2017).
17. A. Sánchez Alvarado, P. A. Tsonis, *Nat. Rev. Genet.* **7**, 873–884 (2006).
18. L. R. Blauch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7283–7288 (2017).
19. J. R. Collier, N. A. M. Monk, P. K. Maini, J. H. Lewis, *J. Theor. Biol.* **183**, 429–446 (1996).
20. S. J. Bray, *Nat. Rev. Mol. Cell Biol.* **7**, 678–689 (2006).
21. M. Matsuda, M. Koga, K. Woltjen, E. Nishida, M. Ebisuya, *Nat. Commun.* **6**, 6195 (2015).
22. S. Wennekamp, S. Mesecke, F. Nédélec, T. Hiiragi, *Nat. Rev. Mol. Cell Biol.* **14**, 452–459 (2013).
23. S. C. van den Brink *et al.*, *Development* **141**, 4231–4242 (2014).
24. J. Vendome *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E4175–E4184 (2014).
25. E. Cachat *et al.*, *Sci. Rep.* **6**, 20664 (2016).
26. A. M. Turing, *Philos. Trans. R. Soc. London Ser. B* **237**, 37–72 (1952).
27. M. D. Jacobson, M. Weil, M. C. Raff, *Cell* **88**, 347–354 (1997).
28. E. Scarpa, R. Mayor, *J. Cell Biol.* **212**, 143–155 (2016).
29. E. Li, *Nat. Rev. Genet.* **3**, 662–673 (2002).
30. J. A. Davies, *J. Anat.* **212**, 707–719 (2008).
31. S. A. Newman, R. Bhat, *Int. J. Dev. Biol.* **53**, 693–705 (2009).
32. Y. E. Antebi, N. Nandagopal, M. B. Elowitz, *Curr. Opin. Syst. Biol.* **1**, 16–24 (2017).
33. J. B. A. Green, J. Sharpe, *Development* **142**, 1203–1211 (2015).
34. Y. Hart *et al.*, *Cell* **158**, 1022–1032 (2014).
35. Y. Okabe, R. Medzhitov, *Nat. Immunol.* **17**, 9–17 (2016).
36. N. King *et al.*, *Nature* **451**, 783–788 (2008).
37. M. Abedin, N. King, *Science* **319**, 946–948 (2008).
38. S. A. Nichols, B. W. Roberts, D. J. Richter, S. R. Fairclough, N. King, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13046–13051 (2012).
39. R. A. Alegado *et al.*, *eLife* **1**, e00013 (2012).
40. A. Woznica *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7894–7899 (2016).

ACKNOWLEDGMENTS

We thank B. Honig and L. Shapiro for discussions regarding cadherin sorting and for providing cadherin plasmids, O. Klein for comments, members of the Lim Lab and the Morsut lab for discussion and assistance, and K. Roybal, M. Thomson, Z. Gartner, W. Marshall, J. Fung, and members of the UCSF Center for Systems and Synthetic Biology and the NSF Center for Cellular Construction. **Funding:** Supported by Japan Society for the Promotion of Science (JSPS) Overseas Research Fellowships (S.T.); a Human Frontiers of Science Program (HFSP) (S.T. and L.M.); a European Molecular Biology Organization (EMBO) Postdoctoral Fellowship (L.M.); NIH grants K99 1K99EB021030 (L.M.), 5P50GM081879 (W.A.L.), and T32GM008412 (L.R.B.); NSF DBI-1548297 Center for Cellular Construction (W.A.L. and S.K.Y.T.); the DARPA Engineered Living Materials program; and the Howard Hughes Medical Institute (W.A.L.). **Author contributions:** S.T. developed and planned research, carried out the design, construction, and testing of self-organizing circuits, and wrote the manuscript; L.R.B. designed and built the microfluidic guillotine and executed experiments for testing regeneration; S.K.Y.T. designed and conceptualized the microfluidic guillotine; L.M. developed and planned research and wrote and edited the manuscript; and W.A.L. developed, planned, and oversaw research and wrote and edited the manuscript. **Competing interests:** The synNotch platform has been patented (U.S. Patent 9,670,281) by the Regents of the University of California and is licensed by Gilead Biosciences. W.A.L. has a financial interest in Gilead Biosciences. **Data and materials availability:** All constructs will be made available on Addgene, and all relevant data are included within the main paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/156/suppl/DC1
Materials and Methods
Figs. S1 to S10
Table S1
Movies S1 to S7
References (*41*, *42*)

17 January 2018; accepted 22 May 2018
Published online 31 May 2018
10.1126/science.aat0271

REPORT

PHYSICS

Phase transitions in a programmable quantum spin glass simulator

R. Harris[1]*, Y. Sato[1], A. J. Berkley[1], M. Reis[1], F. Altomare[1], M. H. Amin[1,2], K. Boothby[1], P. Bunyk[1], C. Deng[1], C. Enderud[1], S. Huang[1], E. Hoskinson[1], M. W. Johnson[1], E. Ladizinsky[1], N. Ladizinsky[1], T. Lanting[1], R. Li[1], T. Medina[1], R. Molavi[1,3], R. Neufeld[1], T. Oh[1], I. Pavlov[1], I. Perminov[1], G. Poulin-Lamarre[1], C. Rich[1], A. Smirnov[1], L. Swenson[1], N. Tsai[1], M. Volkmann[1], J. Whittaker[1], J. Yao[1]

Understanding magnetic phases in quantum mechanical systems is one of the essential goals in condensed matter physics, and the advent of prototype quantum simulation hardware has provided new tools for experimentally probing such systems. We report on the experimental realization of a quantum simulation of interacting Ising spins on three-dimensional cubic lattices up to dimensions 8 × 8 × 8 on a D-Wave processor (D-Wave Systems, Burnaby, Canada). The ability to control and read out the state of individual spins provides direct access to several order parameters, which we used to determine the lattice's magnetic phases as well as critical disorder and one of its universal exponents. By tuning the degree of disorder and effective transverse magnetic field, we observed phase transitions between a paramagnetic, an antiferromagnetic, and a spin-glass phase.

Can quantum physics be efficiently simulated on a computer? What kind of computer would one need to do so? Feynman posed these questions to motivate the concept of a probabilistic quantum mechanical computer that could be used to simulate nature (*1*). Whereas the implementation of such computers seemed daunting in Feynman's time, subsequent progress in the laboratory has made small-scale quantum simulation possible today (*2*–*9*). Although the field of quantum simulation is still in its infancy, the scale and the control of engineered quantum systems has reached the point where they can begin to address fundamental problems in physics.

In this Report, we describe the use of an engineered quantum system to investigate the properties of a particular quantum magnetism problem known as the transverse field Ising model (TFIM). The TFIM is often viewed as the archetypal model for order-disorder phase transitions in quantum systems (*10*). A diverse collection of condensed matter systems, such as ferroelectrics (*11*) and magnetic spin glasses (*12*), can be described by the TFIM. Although this model can be analytically solved in one dimension, it is generally only amenable to numerical study via computationally intensive means, such as quantum Monte Carlo (*13*), in higher dimensions. We have simulated the TFIM on three-dimensional (3D) simple cubic (SC) lattices up to dimensions 8 × 8 × 8 by embedding the model on a D-Wave quantum processing unit (QPU) (D-Wave Systems, Burnaby, Canada) (*14*). By tuning the amount of disorder within the lattice and varying the effective transverse magnetic field, we demonstrate phase transitions between a paramagnetic (PM), an ordered antiferromagnetic (AFM), and a spin-glass (SG) phase. The experimental results compare well with theory for this particular SG problem, thus validating the use of a probabilistic quantum computer to simulate materials physics. This represents an important step forward in the realization of integrated quantum circuits at a scale that is relevant for condensed matter research.

The system studied here is a SC lattice of Ising spin-1/2 particles subject to a global transverse field splitting (*10*). The single-spin energy splitting induced by the transverse field is denoted by Γ, and the magnitude of the nearest-neighbor interaction energy is denoted by $\mathcal{J}$. To induce glassy behavior, let the sign of the interaction vary randomly, with p and $1 - p$ denoting the probability that a particular interaction J_{ij} be –1 (ferromagnetic) and +1 (AFM), respectively. The Hamiltonian for this system can be written as

$$\hat{\mathcal{H}}_{\mathrm{SG}} = \mathcal{J}\sum_{\langle ij\rangle} J_{ij}\hat{\sigma}_i^z\hat{\sigma}_j^z - \Gamma\sum_i \hat{\sigma}_i^x \qquad (1)$$

where $\hat{\sigma}_i^z$ and $\hat{\sigma}_i^x$ are the Pauli z and x operators, respectively, acting on spin i, and the notation $\langle ij\rangle$ indicates summing over all nearest neighbor pairs. A schematic phase diagram for this system is depicted in Fig. 1A. For $\Gamma = 0$, this is a well-studied classical system, with phase transitions induced by thermal fluctuations. The critical disorder $p_c = 0.222$ (*15*) marks the value of p below which one observes an ordered AFM phase, whereas above this value one observes a SG phase. Critical temperatures $k_B T^c_{\mathrm{AFM}}/\mathcal{J} = 4.5$ at $p = 0$ (*16*) and $k_B T^c_{\mathrm{SG}}/\mathcal{J} = 1.1$ at $p = 0.5$ (*17*), where k_B is the Boltzmann constant, indicate the locations of PM-AFM and PM-SG thermal phase transitions, respectively, at $\Gamma = 0$. As Γ is increased, phase boundaries evolve because of the increased role that quantum fluctuations play over thermal fluctuations. The critical transverse field energies $\Gamma^c_{\mathrm{AFM}}/\mathcal{J} \approx 5.2$ at $p = 0$ (*18*) and $\Gamma^c_{\mathrm{SG}}/\mathcal{J} \approx 3.1$ at $p = 0.5$ (*19*) indicate locations of quantum phase transitions between the PM-AFM and PM-SG phases, respectively, at $T = 0$.

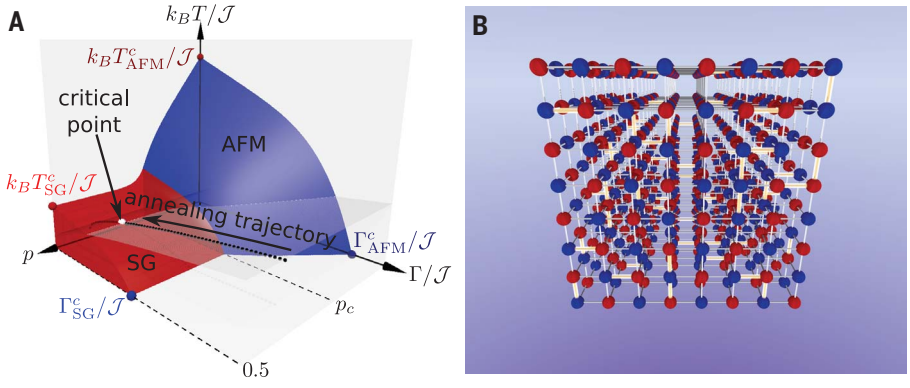


Fig. 1. The bond-disordered TFIM on a simple cubic lattice. (**A**) Schematic phase diagram. Axes correspond to scaled transverse magnetic field splitting $\Gamma/\mathcal{J}$, disorder p, and scaled temperature $k_BT/\mathcal{J}$, where $\mathcal{J}$ is the Ising interaction strength. The experiments reported on here probed this space on a curved manifold (translucent surface). Any single experiment for fixed p involves annealing along a trajectory on this manifold (black dotted curve) that starts in the PM phase, and a critical point is encountered where the trajectory penetrates a phase boundary (silver point). (**B**) Illustration of a ground state at $\Gamma = 0$ for an 8 × 8 × 8 lattice instance for $p = 0.1 < p_c$. Ferromagnetic ($J_{ij} = -1$) and AFM ($J_{ij} = \pm 1$) bonds are indicated by gold and silver cylinders, respectively. Ising spin orientations ↑ and ↓ are represented by red and blue spheres, respectively.

[1]D-Wave Systems, 3033 Beta Avenue, Burnaby, BC V5G 4M9, Canada. [2]Department of Physics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. [3]Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.
*Corresponding author. E-mail: rharris@dwavesys.com

The SC lattice described above was embedded into the native graph of a prototype QPU (*14*). The QPU and the embedding are described in the supplementary materials. Lattices up to size $8 \times 8 \times 8$ fit into the QPU, an example of which is shown in Fig. 1B. Within this experimental platform, the coupling energy $\mathcal{J}$ and transverse field splitting Γ are monotonic functions of a dimensionless control parameter $0 \le s \le 1$ so that $\Gamma(s=0) \gg \mathcal{J}(s=0)$ and $\Gamma(s=1) \ll \mathcal{J}(s=1)$. For fixed temperature $T \approx 12$ mK ($k_B T/h \approx 250$ MHz), the QPU can be used to explore the phase space on a curved manifold in $(p, \Gamma/\mathcal{J}, k_B T/\mathcal{J})$–space, as indicated in Fig. 1A. Details concerning the functional form of this manifold can be found in the supplementary materials.

To experimentally determine the critical disorder p_c, we first worked in a classical regime by noting that the final Hamiltonian at $s = 1$ is given by the classical form of Hamiltonian (Eq. 1) with $\Gamma(s=1) \to 0$. Ensembles of up to 1000 random instances, with each instance corresponding to a specific set of J_{ij}, were generated for a series of values of p and lattice sizes $L \times L \times L$. Each instance was programmed into the QPU, and low-energy states of the final classical Hamiltonian were obtained by slowly evolving from $s = 0$ to $s = 1$ in a process known as quantum annealing (QA) (*20*–*22*). Experimental details are available in the supplementary materials. QA was performed 1000 times per instance, and the final states of every spin were recorded. In order to approximately estimate ground-state properties, the results were then filtered in order to identify the lowest-energy state returned by the QPU for each instance. In cases in which the lowest observed energy was degenerate, all such states were kept. The result was a set of low-energy states $|\vec{s}\rangle = |s_1 s_2 \ldots s_N\rangle$, where $N = L^3$ is the number of lattice spins, for each value of p that could then be used to determine disorder-averaged quantities, as described below.

To detect the AFM-SG phase transition as a function of p, we calculated the absolute value of the AFM magnetization (also known as staggered magnetization)

$$|m_{\mathrm{AFM}}| = \left| \frac{1}{N} \sum_{i=1}^{N} (-1)^{a_i + b_i + c_i} s_i \right| \qquad (2)$$

for each state $|\vec{s}\rangle$, where the vector of integers (a_i, b_i, c_i) corresponds to the position of site i in the SC lattice (relative to an arbitrary origin), and s_i is the state of the spin at site i. A summary plot of $\langle |m_{\mathrm{AFM}}| \rangle$ versus p, where $\langle \ldots \rangle$ denotes averaging over disorder, is shown in Fig. 2A. Data are shown for three system sizes $L \times L \times L$, for $L \in [4, 6, 8]$. For all system sizes, $\langle |m_{\mathrm{AFM}}| \rangle$ monotonically decreases with increasing p, and the slope near $p = p_c$ becomes increasingly steep with larger L. However, this quantity does not strictly go to zero for $p > p_c$ because of finite size effects (*23*). To determine the experimentally observed critical point p_c^{exp}, we calculated the Binder cumulant (*24*)

$$g = 1 - \frac{\langle m_{\mathrm{AFM}}^4 \rangle}{3(\langle m_{\mathrm{AFM}}^2 \rangle)^2} \qquad (3)$$

versus p for $L \in [4, 6, 8]$, because this quantity should be scale invariant at a critical point. The results of this analysis are shown in Fig. 2B, with independent fits of the data near p_c to second-order polynomials for each value of L. From these fits, we estimate the mutual crossing point for all three system sizes to be $p_c^{\mathrm{exp}} = 0.22 \pm 0.01$. A collapse of the data are shown in Fig. 2C for a posited scaling function $g(p, L) \equiv \mathrm{Fcn}[L^{1/\nu_{\mathrm{exp}}}(p - p_c^{\mathrm{exp}})]$, from which we obtain a critical exponent $1/\nu_{\mathrm{exp}} = 0.85 \pm 0.05$. These experimental results are in good agreement with the numerical results of $p_c = 0.222$ and $1/\nu = 0.9$ reported in literature (*15*). This observation confirms that the embedded classical problem is in the same universality class as the 3D cubic lattice problem in the limit of vanishing transverse field (*25*).

We next turned to the identification of phase transitions between the PM, AFM, and SG phases as a function of $\Gamma/\mathcal{J}$ and $k_B T/\mathcal{J}$. Unlike the above section, we worked in the quantum regime, where Γ is nonzero. The PM-to-AFM phase transition is known to be of second order (*18*), thus motivating the use of a linear susceptibility measurement that would diverge at a critical point in the thermodynamic limit $L \to \infty$. Because the ordered phase at $p = 0$ is AFM, we applied a global longitudinal magnetic field that was uniform in magnitude but alternating in sign between nearest-neighbor sites ($\pm h$). We refer to the quantity $\chi_{\mathrm{AFM}} \equiv \delta m_{\mathrm{AFM}}/\delta h|_{h \to 0}$ as the AFM or staggered linear susceptibility. χ_{AFM} was

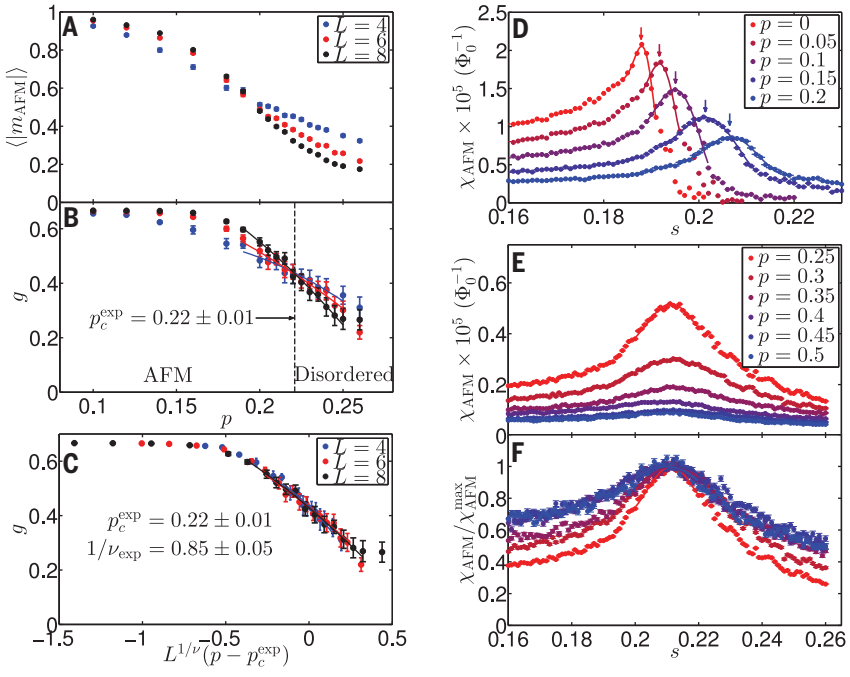


Fig. 2. Determination of the critical disorder p_c and AFM susceptibility χ_{AFM} versus quantum annealing parameter s. (**A**) Disorder-averaged AFM magnetization $\langle |m_{\mathrm{AFM}}| \rangle$ versus p for three different lattice sizes $L \times L \times L$. (**B**) Binder cumulant g versus p. For each L, the results near p_c have been independently fit to second-order polynomials (solid curves). The critical disorder is given by the mutual crossing point, which is estimated to be $p_c^{\mathrm{exp}} = 0.22 \pm 0.01$ (dashed vertical line). (**C**) Finite-size scaling collapse of the experimental results. Assuming that g is a function of scaled disorder $L^{1/\nu_{\mathrm{exp}}}(p - p_c^{\mathrm{exp}})$, the best collapse was obtained for $1/\nu_{\mathrm{exp}} = 0.85 \pm 0.05$. (**D**) χ_{AFM} versus s for $p < p_c$. (**E**) χ_{AFM} versus s for $p > p_c$. (**F**) Normalized χ_{AFM} versus s for $p > p_c$.

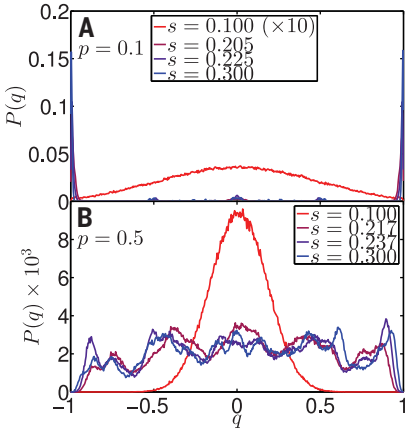


Fig. 3. Example spin-spin overlap probability distributions $P(q)$. (**A**) Example PM-to-AFM phase transition for a $p = 0.1$ instance at four representative values of s. The results for $s = 0.1$ have been magnified ×10. (**B**) Example PM-to-SG phase transition for a $p = 0.5$ instance at four representative values of s.

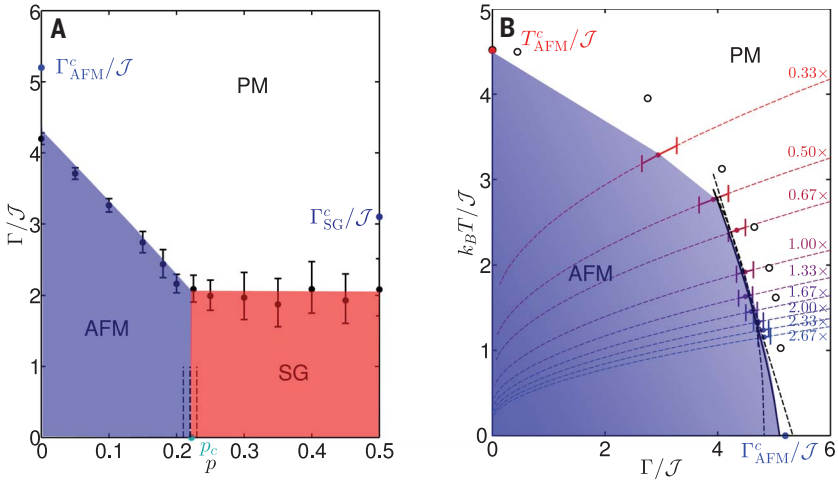


Fig. 4. $p - \Gamma/\mathcal{J}$ and $k_BT/\mathcal{J} - \Gamma/\mathcal{J}$ phase diagrams. (**A**) $p - \Gamma/\mathcal{J}$ phase diagram computed from χ_{AFM} peak positions (black points with error bars). All measurements performed at T = 12.0 mK mK. Thick dashed black line indicates experimentally determined critical disorder p_c^{exp}, and the thin dashed black lines indicate error estimates for that quantity. Critical transverse fields $\Gamma^c_{AFM}/\mathcal{J}$ and $\Gamma^c_{SG}/\mathcal{J}$ for an infinite system at T = 0 and critical disorder p_c were taken from the literature. (**B**) χ_{AFM} peak positions for varying relative coupling strength $\mathcal{J}$ translated into the $(\Gamma/\mathcal{J}, k_BT/\mathcal{J})$–plane for p = 0. $\chi_{AFM}(s)$ data are provided in fig. S13. Annealing trajectories for varying relative coupling strength are indicated by dashed curves (supplementary materials), and the corresponding transition points are indicated by color-coordinated points, with error estimate along each trajectory. Trajectory and critical point denoted as 1.00× correspond to the experimental manifold depicted in Fig. 1A and the p = 0 point in (A), respectively. Trajectories and points labeled 0.33×, 0.50×, and 0.67× indicate lower $\mathcal{J}$, whereas those labeled 1.33×, 1.67×, 2.00×, 2.33×, and 2.67× indicate higher $\mathcal{J}$. Best fit of phase boundary and extrapolation to T = 0 is indicated by solid black curve. Dashed black curves indicate error bounds on the best fit model. Open circles represent the theoretical results (*30*) for an infinite system.

measured as a function of QA control parameter $0 < s < 1$ for small sets of randomly chosen instances for a series of values of p. Disorder-averaged traces of χ_{AFM} versus s for 8 × 8 × 8 lattices are presented in Fig. 2, D to F. For $p < p_c$ = 0.22, this quantity yields obvious peaks that broaden and shift to increasing s with increasing p in Fig. 2D. In what follows, we refer to the peak positions as s_c. For $p > p_c$ = 0.22, s_c remains fixed to within experimental resolution, but the peak amplitude continues to decrease with increasing p, as shown in Fig. 2E. Furthermore, the peak broadens until p = 0.4, beyond which the breadth remains fixed. This latter observation is made more clear in the scaled plot of $\chi_{AFM}(s)/\chi_{AFM}(s_c)$ versus s in Fig. 2F.

The peaks observed for $p < p_c$ suggest an incipient divergence, which reliably indicates a second-order phase transition between the PM and AFM phases. For $p > p_c$, the absence of such a divergence is in agreement with the results of numerical simulations of the PM-to-SG phase transition in a related system (*26*). We consider this to be only part of the evidence needed to confirm that a PM-to-SG phase transition has been observed.

The AFM magnetization $|m_{AFM}|$, as given in Eq. 2, is an order parameter that one can use to easily identify the AFM phase. However, $\langle|m_{AFM}|\rangle = 0$ in both the PM and SG phases. The latter two phases may be distinguished by using an order parameter proposed by Edwards and Anderson (*27*), hereafter referred to as the spin-spin overlap q, which is defined as

$$q \equiv \frac{1}{N}\sum_{i=1}^{N} s_i^A s_i^B \qquad (4)$$

where A and B are independently determined states for a given instance of disorder. Given a statistically large number of such states, one can construct the probability distribution $P(q)$, whose characteristics dramatically differ between the PM, AFM, and SG phases. Examples of experimentally determined $P(q)$ for an AFM (p = 0.1) and a SG (p = 0.5) instance of size 8 × 8 × 8 at four values of s are shown in Fig. 3, A and B, respectively. In both cases, the results reveal a Gaussian-shaped $P(q)$ centered at q = 0 for $s < s_c$, which is consistent with the PM phase in a finite-sized system. For the AFM example in Fig. 3A, $P(q)$ bifurcates for $s > s_c$, and all of the weight becomes consolidated in the two sharp peaks at $q = \pm 1$ above the transition. This is the signature of an ordered phase, in which there is a single thermodynamic state up to Z_2 symmetry (*28*). For the SG example in Fig. 3B, $P(q)$ broadens in the vicinity of s_c and then develops multiple peaks that are approximately symmetric around q = 0 and continuous throughout. This is consistent with the existence of many competing thermodynamic states, as typically characterizes a disordered SG phase (*28*).

The final analysis presented here is the phase diagram projected on to the $(p, \Gamma/\mathcal{J})$–plane for the manifold depicted in Fig. 1A. To convert s_c inferred from linear susceptibility peak positions to $\Gamma/\mathcal{J}$, we used the lattice spin energy scale versus s mapping given in the supplementary materials. The results of this mapping are shown in Fig. 4A along with the phases identified based on the order parameter analysis presented above. For reference, we have also included p_c from the literature (*15*), as well as values of $\Gamma^c_{AFM}/\mathcal{J}$ (*18*) and $\Gamma^c_{SG}/\mathcal{J}$ (*19*) for an infinitely large system at T = 0. The experimentally determined phase diagram projection yields a value of p_c at Γ = 0 that agrees with numerical results. As of the time of writing, we had not established an experimental method to resolve any possible reentrant behavior of the AFM-SG phase boundary in the (p, Γ)–plane, as is suspected to exist in the (p, T)–plane (*29*). Critical transverse fields are systematically lower than the $L \to \infty$ results at T = 0 because we are constrained to measure on a particular manifold, as depicted in Fig. 1A. The intersection between this experimentally accessible manifold and the phase boundaries can be varied by altering the annealing trajectory, which can be achieved by adjusting the magnitude of the interaction energy $\mathcal{J}$ by an overall scale factor (with further details given in the supplementary materials). For fixed p, one can measure the χ_{AFM} peak position, similar to that for p = 0 depicted in Fig. 2D, as a function of relative scaling of $\mathcal{J}$. Doing so samples the PM-AFM phase boundary in the $(\Gamma/\mathcal{J}, k_BT/\mathcal{J})$–plane. Results for p = 0 are shown in Fig. 4B. Decreasing or increasing $\mathcal{J}$ moves the phase boundary toward the classical transition point $k_BT^c_{AFM}/\mathcal{J}$ (*16*) or quantum critical point $\Gamma^c_{AFM}/\mathcal{J}$ (*18*), respectively. Fitting the rightmost edge of the phase boundary to a power law $k_BT/\mathcal{J} \propto |\Gamma/\mathcal{J} - \Gamma^c_{exp}/\mathcal{J}|^{a_{exp}}$ yielded an experimental quantum critical point $\Gamma^c_{exp}/\mathcal{J} = 5.1 \pm 0.3$ and exponent $a_{exp} = 0.7 \pm 0.3$. We have also included the theoretical results of Elliott and Wood (*30*) for an infinite-sized cubic lattice in Fig. 4B.

The experiments reported on here demonstrate how an in situ programmable TFIM in an integrated circuit can be used as a materials physics simulator. The ability to manipulate individual spins and bonds allows one to explore order parameters in ways that are not possible in bulk condensed matter systems. This work opens many potential avenues of research within the field of quantum magnetism, with the possibility of studying lattices with built-in geometrical frustration (*31*, *32*), lattice gauge theories (*33*), and defect/domain formation.

REFERENCES AND NOTES

1. R. P. Feynman, *Int. J. Theor. Phys.* **21**, 467–488 (1982).
2. C. Gross, I. Bloch, *Science* **357**, 995–1001 (2017).
3. H. Weimer, M. Müller, I. Lesanovsky, P. Zoller, H. P. Büchler, *Nat. Phys.* **6**, 382–388 (2010).
4. H. Bernien *et al.*, *Nature* **551**, 579–584 (2017).
5. J. Zhang *et al.*, *Nature* **551**, 601–604 (2017).

6. A. Friedenauer, H. Schmitz, J. T. Glueckert, D. Porras, T. Schaetz, *Nat. Phys.* **4**, 757–761 (2008).
7. K. Kim *et al*., *Nature* **465**, 590–593 (2010).
8. X. Ma, B. Dakic, W. Naylor, A. Zeilinger, P. Walther, *Nat. Phys.* **7**, 399–405 (2011).
9. J. Zhang, M.-H. Yung, R. Laflamme, A. Aspuru-Guzik, J. Baugh, *Nat. Commun.* **3**, 880 (2012).
10. S. Suzuki, J.-I. Inoue, B. K. Chakrabarti, *Quantum Ising Phases and Transitions in Transverse Ising Models* (Springer, 2012).
11. R. Blinc, B. Žekš, J. F. Sampaio, A. S. T. Pires, F. C. S. Barreto, *Phys. Rev. B* **20**, 1991–2001 (1979).
12. W. Wu, B. Ellman, T. F. Rosenbaum, G. Aeppli, D. H. Reich, *Phys. Rev. Lett.* **67**, 2076–2079 (1991).
13. M. Suzuki, ed., *Quantum Monte Carlo Methods* (Springer, 1986).
14. R. Harris *et al*., *Phys. Rev. B* **82**, 024511 (2010).
15. A. K. Hartmann, *Phys. Rev. B* **59**, 3617–3623 (1999).
16. C. Domb, M. Green, eds., *Phase Transitions and Critical Phenomena* (Academic, 1974).
17. R. N. Bhatt, A. P. Young, *Phys. Rev. Lett.* **54**, 924–927 (1985).
18. R. M. Stratt, *Phys. Rev. B Condens. Matter* **33**, 1921–1930 (1986).
19. R. R. P. Singh, A. P. Young, *Phys. Rev. B* **96**, 214210 (2017).
20. T. Kadowaki, H. Nishimori, *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **58**, 5355–5363 (1998).
21. E. Farhi *et al*., *Science* **292**, 472–475 (2001).
22. J. Brooke, D. Bitko, T. F. Rosenbaum, G. Aeppli, *Science* **284**, 779–781 (1999).
23. V. Privman, ed., *Finite Size Scaling and Numerical Simulations of Statistical Systems* (World Scientific, 1990).
24. K. Binder, *Z. Phys. B Condens. Matter* **43**, 119–140 (1981).
25. H. G. Katzgraber, F. Hamze, R. S. Andrist, *Phys. Rev. X* **4**, 021008 (2014).
26. M. Guo, R. N. Bhatt, D. A. Huse, *Phys. Rev. Lett.* **72**, 4137–4140 (1994).
27. S. F. Edwards, P. W. Anderson, *J. Phys. F Met. Phys.* **5**, 965–974 (1975).
28. H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford Science Publications, 2001).
29. M. Hasenbusch, A. Pelissetto, E. Vicari, *Phys. Rev. B* **78**, 214205 (2008).
30. R. J. Elliott, C. Wood, *J. Phys. C Solid State Phys.* **4**, 2359–2369 (1971).
31. Y. L. Loh, D. X. Yao, E. W. Carlson, *Phys. Rev. B* **77**, 134402 (2008).
32. R. Moessner, S. L. Sondhi, P. Chandra, *Phys. Rev. Lett.* **84**, 4457–4460 (2000).
33. G. Misguich, D. Serban, V. Pasquier, *Phys. Rev. Lett.* **89**, 137202 (2002).

ACKNOWLEDGMENTS

We thank the technical staff of D-Wave Systems for making these experiments possible; W. Bernoudy for illustrations and animations of the embedding; and A. King, J. Raymond, A. P. Young, and H. G. Katzgraber for useful discussions and for reviewing this manuscript. **Funding:** This work was wholly funded by D-Wave Systems. **Authors contributions:** All authors contributed to the conceptualization of the quantum processing unit used in this study. K.B., P.B., C.E., R.M., N.T., and M.V. provided resources by generating the detailed design of the circuit. S.H., E.L., N.L., T.O., and J.Y. provided resources by directly participating in and directing the fabrication of the circuit. R.L., T.M., and R.N. provided resources by preparing and qualifying the circuit. R.H., Y.S., A.J.B., M.R., F.A., C.D., E.H., M.W.J., T.L., I.Pa., I.Pe., G.P.-L., C.R., L.S., and J.W. provided resources by calibrating and characterizing the circuit. R.H. and Y.S. performed the investigation, validation, and data curation. R.H., Y.S., M.H.A., and A.S. contributed to the formal analysis. R.H. and Y.S. wrote the original draft of the manuscript. R.H., Y.S., A.J.B., F.A., M.H.A., E.H., E.L., T.L., and A.S. contributed to the revised manuscript. **Competing interests:** All authors are employees and holders of stock options in D-Wave Systems. R.H. and K.B. are inventors of U.S. patent application no. 15/881260. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/162/suppl/DC1
Materials and Methods
Figs. S1 to S15
Table S1
References (*34*–*44*)
Movie S1
Data Files

2 February 2018; accepted 15 May 2018
10.1126/science.aat2025

GEODETIC MONITORING

The rise, collapse, and compaction of Mt. Mantap from the 3 September 2017 North Korean nuclear test

Teng Wang[1]*, Qibin Shi[1,2], Mehdi Nikkhoo[3], Shengji Wei[1,2]*, Sylvain Barbot[1,2], Douglas Dreger[4]*, Roland Bürgmann[4], Mahdi Motagh[3,5], Qi-Fu Chen[6]

Surveillance of clandestine nuclear tests relies on a global seismic network, but the potential of spaceborne monitoring has been underexploited. We used satellite radar imagery to determine the complete surface displacement field of up to 3.5 meters of divergent horizontal motion with 0.5 meters of subsidence associated with North Korea's largest underground nuclear test. Combining insight from geodetic and seismological remote sensing, we found that the aftermath of the initial explosive deformation involved subsidence associated with subsurface collapse and aseismic compaction of the damaged rocks of the test site. The explosive yield from the nuclear detonation with best-fitting source parameters for 450-meter depth was 191 kilotonnes of TNT equivalent. Our results demonstrate the capability of spaceborne remote sensing to help characterize large underground nuclear tests.

World peace benefits from adherence to internationally negotiated nuclear test ban treaties whose signatories strive to promote the nonproliferation of nuclear weapons. In 2003, the Democratic People's Republic of Korea (North Korea) became the first country to withdraw from the 1968 Treaty on the Non-Proliferation of Nuclear Weapons. North Korea has been conducting underground nuclear weapon tests with increasing intensity since 2006. On 3 September 2017, two seismic events separated by ~8.5 min were detected at North Korea's Punggye-ri nuclear test site. Soon thereafter, North Korea's state media reported the successful firing of a two-stage thermonuclear bomb test. The U.S. Geological Survey and the China Earthquake Networks Center determined a body-wave magnitude (m_b) of 6.3 for the first event (NKNT 6), much larger than any of the five nuclear tests since 2006 (NKNT 1–5). Shortly thereafter, the scientific community started to determine the location, focal mechanism, and yield of the explosion by means of seismic waveforms and satellite optical imagery (*1*). Preliminary analysis revealed a predominantly isotropic explosive source located beneath Mount Mantap (*1*–*3*), which also hosted NKNT 2–5 (Fig. 1).

The source properties of previous North Korean underground nuclear tests have been extensively studied using seismic waveforms (*4*–*12*), but surface displacements associated with these explosions are rarely reported. Remote sensing with synthetic aperture radar (SAR) is a powerful technique for monitoring deformation of Earth's surface (*13*, *14*) but its contribution to characterizing nuclear tests has been limited. NKNT 4, conducted on 6 January 2016, has been studied using SAR interferometry, but the interpretation of interferometric phase is difficult because of the single imaging geometry (*15*). Tracking the amplitude features of the SAR images (so-called pixel-offset tracking) is better suited when the interferometric phase is decorrelated (*16*). Moreover, pixel offsets can map displacement along the radar line-of-sight and satellite-flying (azimuth) directions. In contrast to offset tracking of optical images, the SAR range offset is sensitive to the vertical displacement because of the slant-range imaging geometry, allowing for derivation of three-dimensional (3D) displacements (*17*–*20*). Here, we rely on detailed 3D displacements derived from submeter-resolution SAR images together with seismic waveform data to reveal the complex processes that took place during and in the immediate aftermath of NKNT 6.

We measured the surface displacements caused by NKNT 6 by cross-correlating high-resolution spotlight radar images acquired by the German TerraSAR-X satellite, with an azimuth resolution of 1.1 m and a slant-range resolution as fine as 0.45 m (fig. S1 and table S1). The accuracy of the offset measurement is about one-tenth of the imaging resolution (*21*). We combined the azimuth and range offsets from two ascending and two descending tracks to calculate the total 3D surface displacements produced during and in the immediate aftermath of the explosion on a 300 m × 300 m grid (Fig. 1 and figs. S3 to S5) (*22*). The horizontal motions of up to 3.5 m show a divergent pattern at the top of Mount Mantap with a central zone of subsidence of ~0.5 m. We decomposed the 3D displacements into vertical and horizontal directions along two topographic profiles across the top of Mount Mantap (Fig. 1). The along-profile displacements show that the horizontal displacement is generally larger where the topography is steeper (the west and south flanks). However, the direction of motion does not follow the slope of the terrain but is nearly horizontal. This indicates that although there is strong topographic control on the surface displacement field caused by the buried explosion, it does not resemble the slope-parallel motions expected from triggered landslides. Although optical imagery suggests isolated landslide deposits at the 10- to 100-m scale (*23*), these appear to be debris flows localized in preexisting channels that could not produce the large-scale horizontal motions we observed.

To resolve the horizontal location and depth of the detonation chamber, we set up numerical models that predict the surface displacements attributable to the expansion and subsequent collapse of an underground cavity embedded in a uniform elastic crust below realistic surface topography (*24*, *25*) (Fig. 2). We constrained the 3D location of the source by minimizing the misfit between predicted and observed surface displacements (*26*, *27*). The explosion and immediate collapse of a spherical cavity of 300 m radius that includes the detonation chamber and the surrounding damaged material reproduced the horizontal displacement well, but this was not sufficient to explain the small vertical motion around Mount Mantap. A third, mostly aseismic process involving the compaction of a larger volume is invoked to explain the low uplift (Fig. 2A and fig. S6). A similar compaction process has been inferred from the seismic analysis of other explosions (*28*) and was observed in the weeks to months following underground nuclear tests conducted in Nevada (*29*). As we do not have any constraints on the geometry of the compaction zone, we assumed a generalized ellipsoidal geometry for it and inferred its dimensions by using the geodetic observations. We estimated the explosive source to be located at 129.078°E, 41.300°N ± 50 m, 1750 ± 100 m above mean sea level (i.e., 450 ± 100 m below the top of Mount Mantap). Incorporating the large-scale compaction source into the model does not influence our inferred epicenter of the explosion/collapse source much (*22*).

Assuming that the hypocenter of the first event coincides with the center of the spherical cavity, we refined the relative location for the second seismic event using local seismic waveform records from the NorthEast China Seismic Array to Investigate Deep Subduction (NECsaids) (*30*) and regional data from South Korean sites archived at the Incorporated Research Institutions for Seismology (IRIS) (Fig. 3A) (*22*). With the calibration from the first event and careful *P*-wave arrival picks of the second event (figs. S7 and S8), our grid search showed that the second event occurred 8 min 31.79 s after the first event and was located about 700 m to the south.

[1]Earth Observatory of Singapore, Nanyang Technological University, Singapore. [2]Asian School of Environment, Nanyang Technological University, Singapore. [3]GFZ, German Research Centre for Geosciences, Potsdam, Germany. [4]Department of Earth and Planetary Science, University of California, Berkeley, CA, USA. [5]Institute for Photogrammetry and GeoInformation, Leibniz Universität Hannover, Hannover, Germany. [6]Key Laboratory of Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China.
*Corresponding author. Email: wang.teng@ntu.edu.sg (T.W.); shjwei@ntu.edu.sg (S.W.); dreger@seismo.berkeley.edu (D.D.)

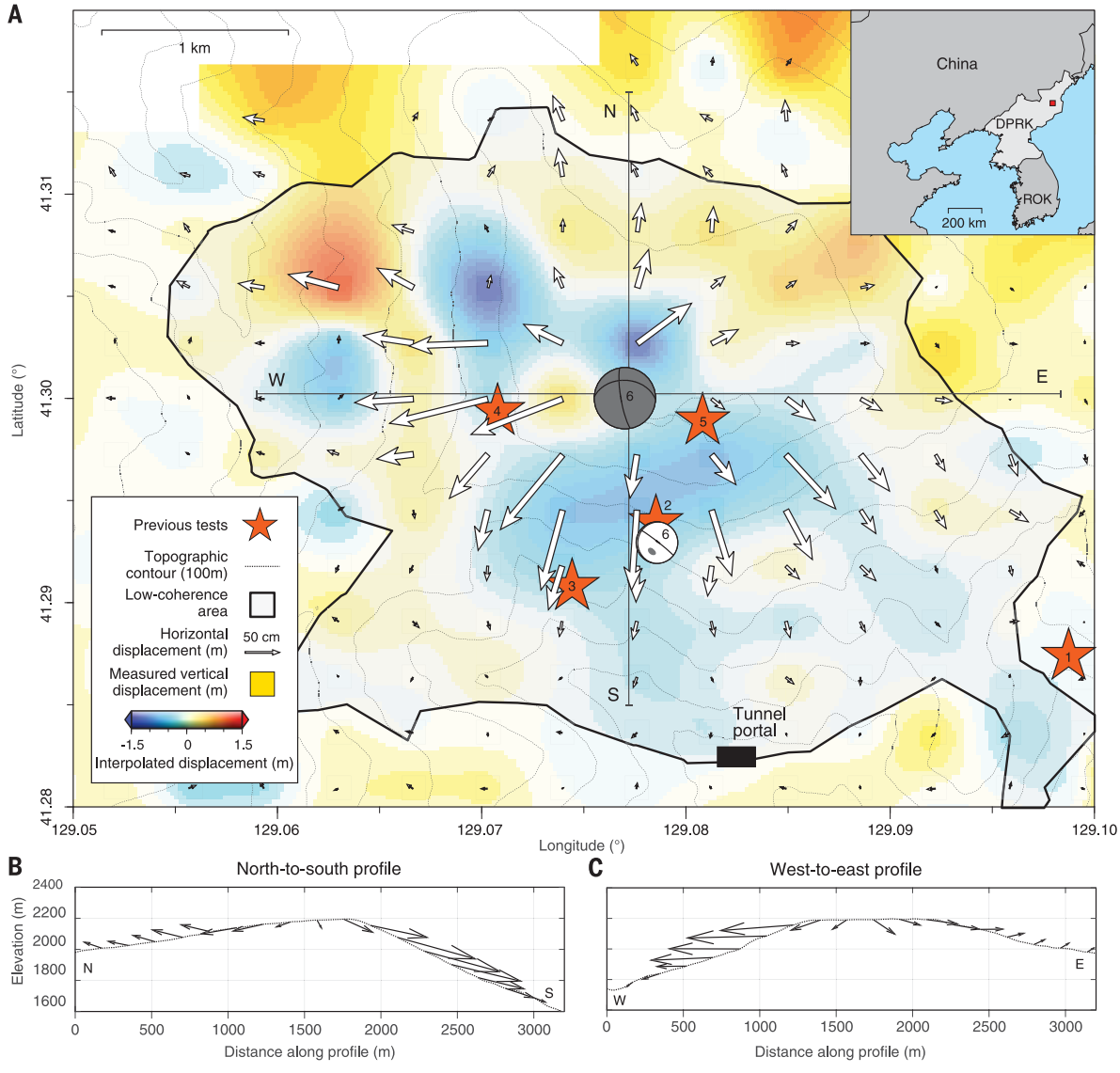


Fig. 1. Three-dimensional displacement associated with the 3 September 2017 North Korea nuclear test (NKNT 6). (**A**) 3D displacements derived from radar imagery. Arrows indicate horizontal displacement; color indicates vertical motions spanning the explosion and ~1 week of additional deformation. The uncertainties are shown in fig. S4 and provided in data S1 with the displacements. The black outline derived from ALOS-2 coherence loss indicates the substantial surface disturbance and large displacement gradients caused by the explosion over an area of ~9 km^2 (figs. S1 and S2). Thin gray lines are topographic contours at 100-m intervals. The red square in the upper right inset shows the location of Mount Mantap (DPRK, Democratic People's Republic of Korea; ROK, Republic of Korea). Red stars indicate the locations of NKNT 1–5 (*1*, *6*, *9*, *15*, *37*), among which NKNT 2–5 were all located within the NKNT 6 low-coherence region; NKNT 1 on 9 October 2006 was in a different location (*5*). Beach balls show locations and focal mechanisms of the M_w 5.24 and M_w 4.47 events on 3 September 2017. (**B** and **C**) 2D (horizontal along the profile and vertical) displacements along two profiles across the top of Mount Mantap from north to south and from west to east, respectively. No vertical exaggeration in (B) or (C).

Because of the azimuthal gap in the station coverage, the east-west location (±700 m) is less well constrained relative to the north-south separation (±200 m with 96% confidence) (Fig. 3C and fig. S9). The refined location of the second implosive event is beneath the area of large subsidence and southward horizontal motion under the south flank of Mount Mantap, between the initial explosion and the south portal of the tunnel system (Fig. 1).

We applied the generalized cut-and-paste (gCAP) method (*31*) to the regional and local waveform data to invert for the full moment tensor solutions of the two seismic events, including an isotropic component (i.e., explosive or implosive volume source), a compensated linear vector dipole (CLVD) component (i.e., ring faulting along a certain axis, such as a collapse), and double-couple component (i.e., shear dislocation on a planar fault) (*22*). Our preferred solution of the first event indicates a moment of 9.5 × 10^{16} N·m (moment magnitude M_w = 5.24), a 50 to 90% positive isotropic component, and relatively small CLVD or double-couple contributions (figs. S10 to S13). The second seismic event (M_w = 4.5) has a large negative isotropic component (~50 to 70% of the total moment) (figs. S14 and S15). Although we obtained a high waveform cross-correlation coefficient between the data and synthetics for most of the waveform components of the first event (e.g., Fig. 3B), the noise level for the second event is larger, resulting in a much smaller variance reduction of the observations (fig. S16). To overcome this limitation of the data, we sought more information about the moment tensor of the second event by directly comparing the waveforms with those of the

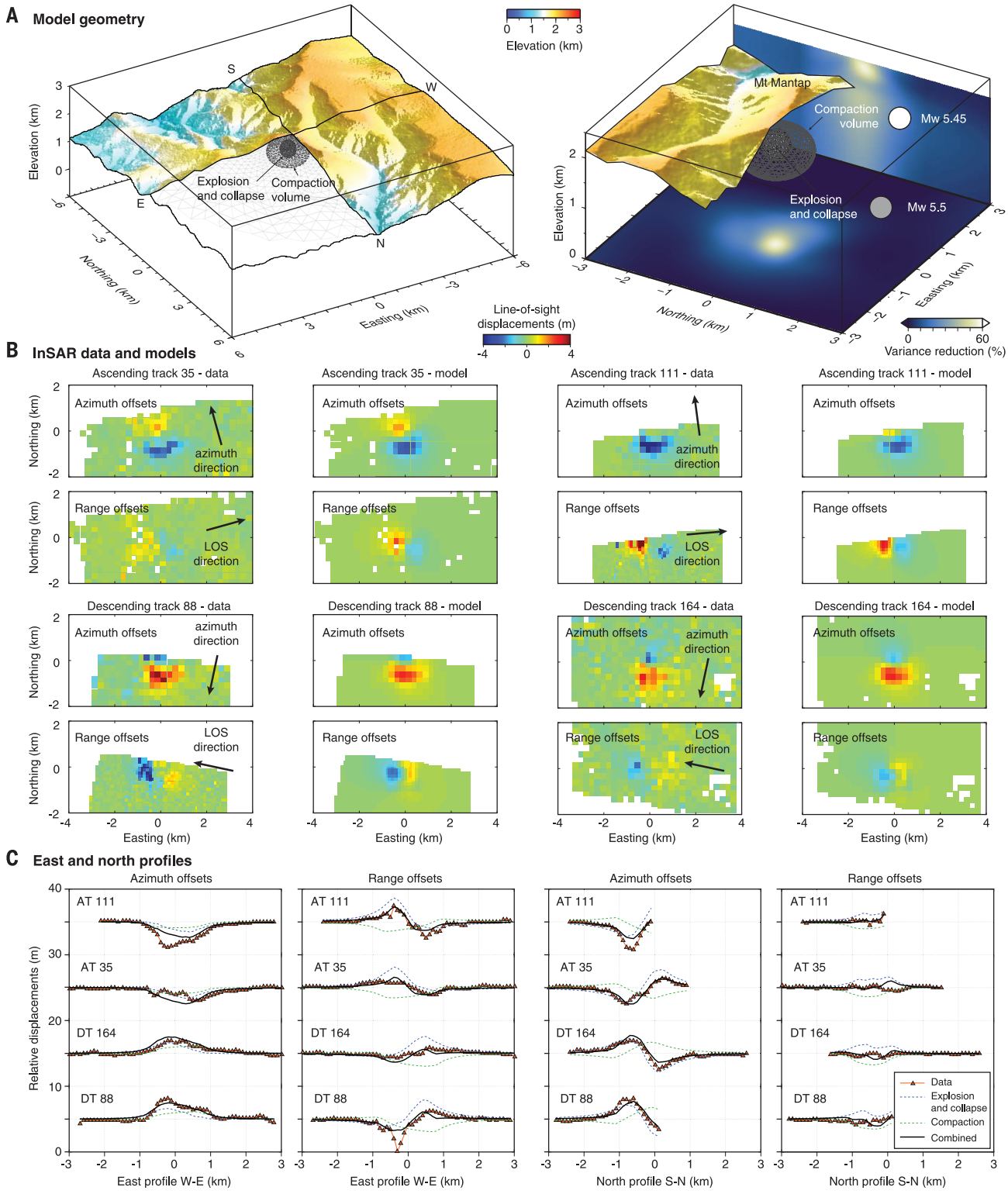


Fig. 2. Model geometry and fit to the observed surface displacements. (**A**) Perspective view of the model with topography and variance reductions as a function of centroid position (both cross sections are centered on the best-fit location). We represent the first event, combining the explosion and immediate collapse, using a sphere of 300 m radius with a centroid located at a depth about 450 m below Mount Mantap. We model the aseismic subsidence detected with geodetic data about a week after the seismic events 1 and 2 with an ellipsoid of dimension 800 m × 800 m × 470 m (semi-axes), centered at 100 m deeper than the explosive source. The isotropic components of the moment tensors are represented as beach balls. (**B**) The interferometric SAR (InSAR) observed and simulated surface displacements. (**C**) The west-east and south-north profiles of the surface displacements from the SAR observations and the best-fitting models. The dashed profiles represent the contributions of the explosion/collapse (event 1) and the subsequent aseismic compaction on the surface displacement. We ignore the deformation caused by event 2.

first event. We multiplied the amplitude of the vertical-component waveforms of the second event by a factor of −60 and compared them with the waveforms of the first event at higher frequencies (~0.2 to 0.9 Hz). The result (Fig. 3D and fig. S16) shows very high waveform cross-correlation coefficients, even for some coda waves, supporting the close locations but opposite isotropic polarities of the two events (*2*).

Combining the depth constraints from geodesy and energy constraints from seismology, we can refine the explosive yield of the nuclear explosion (*32*–*34*). We assumed the seismic velocity model MDJ2 (*4*) for the elastic Earth structure. We based the overburden pressure on the best-fitting centroid source depth of 450 ± 100 m from the geodetic modeling. The medium in which the device was detonated was likely the granodiorite that lies beneath the stratified volcanic rocks that make up the high elevations of Mount Mantap (*8*). We assumed a gas porosity of 1% for granitic rocks (*35*). Considering an isotropic seismic M_w = 5.05 (the mean value of solutions fitting within 95% of the maximum fit for a source depth of 450 m) and the possible range of source depths of 350 to 550 m, the yield estimates range between 171 and 209 kt of TNT equivalent, with 191 kt corresponding to the best-fitting source parameters from geodetic and seismic data (Fig. 3E). Doubling the gas porosity results in an 8% increase in the magnitude of the estimated yield.

The source characteristics we derived from surface displacement and seismic waveforms are in remarkable agreement. The divergent horizontal motions and the moment tensor of the first event consistently suggest a predominant isotropic explosive source buried at shallow depth. The moment of the geodesy-derived models, assuming an empirical rigidity of 5.7 GPa (*36*), is M_w = 5.5, larger than the one inferred seismically (M_w = 5.24), because it includes slow deformation that did not generate seismic waves, with a total volume change of 0.01 km^3. The seismic analysis of the second event reveals an implosive seismic source that occurred south of the first event with a dominant negative isotropic component, suggesting an inverse process of the main explosion. This may reflect the combination of negative isotropic compaction of the overpressured cavity and/or vertical collapse of the explosion chimney and nearby tunnel segments due to gravity, contributing to the subsidence on the south flank of Mount Mantap (Fig. 4). The larger-scale compaction source in the geodetic model is independent of the first and second seismic events, and the post-explosion compaction of surrounding rocks may continue aseismically for an extended period, as seen in the Nevada underground nuclear test site with initial subsidence rates of ~1 to 7 cm/year (*29*).

Our 3D surface displacement measurements and elastic modeling incorporating realistic topography allow for locating the main explosive event within ±50 m, assuming a uniform elastic medium, although a nonuniform structure and small-scale surficial processes (e.g., landslides) may bias our results. Because the largest deformation occurred above the explosive source as a result of the chimneying and spalling effect (*28*), the centroid of the modeled geodetic source may be located above the actual detonation point. After the explosion, water was observed to be flowing from the tunnel portal (*23*). Assuming a slope of 2° to 4° to provide drainage, the depth

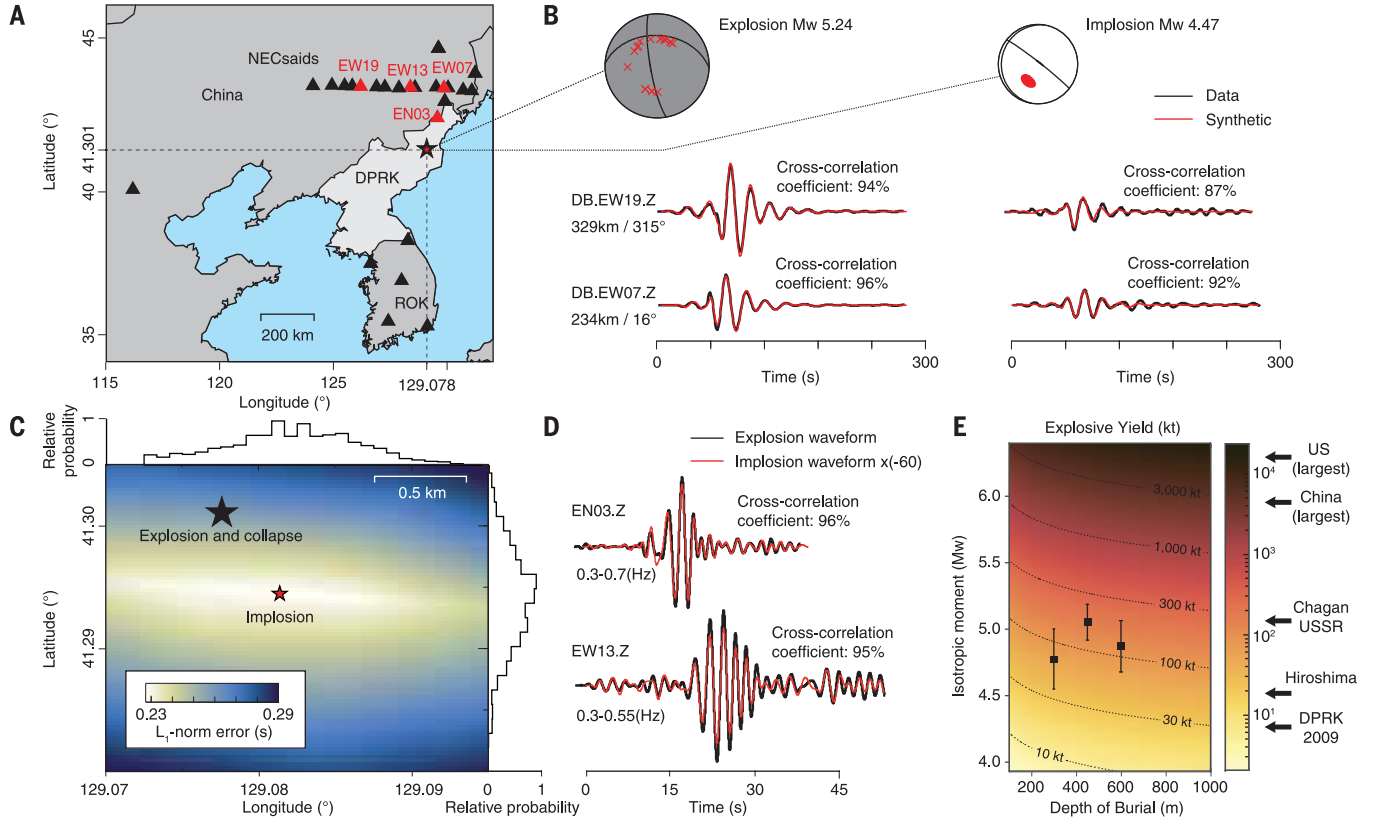


Fig. 3. Analysis of seismic waves. (**A**) Station map of broadband seismometers with four stations in (B) and (D) highlighted in red. The black and red stars are the epicenter locations of the first and second event, respectively. (**B**) Moment tensor solutions for the first explosive event (left) and the second implosive event (right), with the vertical component of two representative stations shown at the bottom. Both data and synthetics are filtered between 0.02 and 0.045 Hz. Station names are shown at the beginning of waveform pairs; distance (in kilometers) and azimuth (in degrees) are indicated below. (**C**) Grid search result (under L1 norm) for relocating the second event relative to the first event (black star). Marginal distributions for the epicentral position are plotted along the northing and easting axes. (**D**) Vertical component waveform comparison between the first (black) and the second (red) event at two representative stations, with the second event waveforms multiplied by −60. (**E**) Explosive yield in the context of historical nuclear tests. The black dots and error bars show yields estimated according to the mean and standard deviation of tabulated moment within 95% of the best-fitting solutions with depths of 300 m, 450 m, and 600 m, respectively.

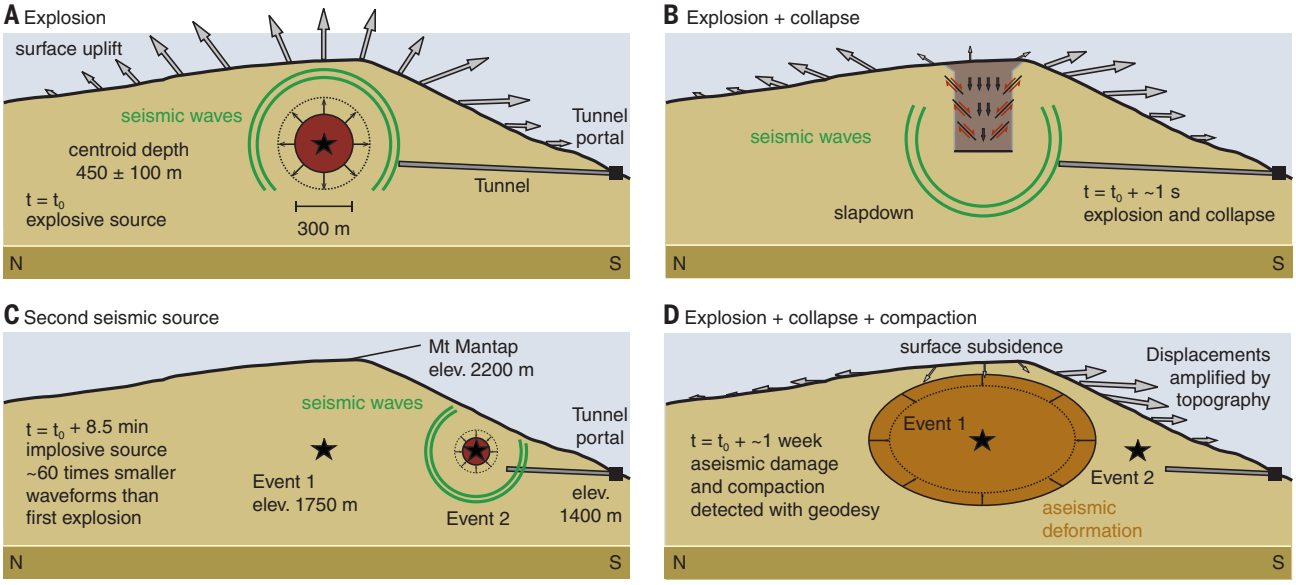


Fig. 4. Summary deformation scenario for the 3 September 2017 North Korea nuclear test. (**A**, **B**, and **D**) The unfolding of events includes the succession of explosive (A), collapse (B), and compaction (D) processes, with different associated surface displacements. (**C**) The implosive source may be shallow and may only contribute localized surface displacements. The radar imagery reveals the deformation [arrows in (D)] resulting from the three processes.

implied from the elevation of the tunnel entrance is about 600 to 700 m below the surface, consistent with a detonation point about 150 m deeper than the centroid of the geodetic model.

Combining the available space-borne geodetic and seismic records provided new insights into the mechanics of deformation surrounding North Korea's sixth underground nuclear test, revealing the explosion, collapse, and subsequent compaction sequence (Fig. 4). The modeling of the geodetic observations reduces the epicentral and depth uncertainties that otherwise hinder the analysis of seismic waveforms. The derived horizontal location of the first event is important to relatively relocate the second event, which likely indicates the collapse of the tunnel system of the test site. The inclusion of geodetic data also helps to resolve the aseismic deformation processes that may follow nuclear tests. Finally, our findings demonstrate the capability of monitoring shallow underground nuclear tests by means of remote-sensing observations and seismic sensors.

REFERENCES AND NOTES

1. Incorporated Research Institutions for Seismology, "Special event: 2017 North Korean nuclear test"; http://ds.iris.edu/ds/nodes/dmc/specialevents/2017/09/03/2017-north-korean-nuclear-test/.
2. J. Liu *et al.*, *Geophys. Res. Lett.* 10.1002/2018GL077095 (2018).
3. D. Tian, J. Yao, L. Wen, *Geophys. Res. Lett.* 10.1029/2018GL077649 (2018).
4. S. Ford, D. Dreger, W. Walter, *Geophys. Res. Lett.* **36**, L21304 (2009).
5. N. Selby, *Bull. Seismol. Soc. Am.* **100**, 1779–1784 (2010).
6. L. Wen, H. Long, *Seismol. Res. Lett.* **81**, 26–29 (2010).
7. K.-Y. Chun, Y. Wu, G. Henderson, *Bull. Seismol. Soc. Am.* **101**, 1315–1329 (2011).
8. E. Rougier, H. J. Patton, E. E. Knight, C. R. Bradley, *Geophys. Res. Lett.* **38**, L16316 (2011).
9. L. F. Zhao, X. B. Xie, W. M. Wang, Z. X. Yao, *Bull. Seismol. Soc. Am.* **102**, 467–478 (2012).
10. M. Zhang, L. Wen, *Geophys. Res. Lett.* **40**, 2941–2946 (2013).
11. L.-F. Zhao, X.-B. Xie, W.-M. Wang, Z.-X. Yao, *Seismol. Res. Lett.* **85**, 130–134 (2014).
12. S. Cesca, S. Heimann, M. Kriegerowski, J. Saul, T. Dahm, *Seismol. Res. Lett.* **88**, 300–310 (2017).
13. J. Hu *et al.*, *Earth Sci. Rev.* **133**, 1–17 (2014).
14. J. R. Elliott, R. J. Walters, T. J. Wright, *Nat. Commun.* **7**, 13844 (2016).
15. M. Wei, *Geophys. J. Int.* **209**, 762–769 (2017).
16. R. Michel, J.-P. Avouac, J. Taboury, *Geophys. Res. Lett.* **26**, 875–878 (1999).
17. E. Pathier *et al.*, *Geophys. Res. Lett.* **33**, L20310 (2006).
18. T. Wang, S. Jonsson, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **8**, 3271–3278 (2015).
19. J. Ruch, T. Wang, W. Xu, M. Hensch, S. Jónsson, *Nat. Commun.* **7**, 12352 (2016).
20. L. N. Schaefer *et al.*, *Geophys. Res. Lett.* **44**, 135–142 (2017).
21. R. Bamler, M. Eineder, *IEEE Geosci. Remote Sens. Lett.* **2**, 151–155 (2005).
22. See supplementary materials.
23. F. V. Pabian, J. S. Bermudez Jr., J. Liu, North Korea's Punggye-ri Nuclear Test Site: Satellite Imagery Shows Post-Test Effects and New Activity in Alternate Tunnel Portal Areas (2017); www.38north.org/2017/09/punggye091217/.
24. M. Nikkhoo, T. R. Walter, *Geophys. J. Int.* **201**, 1117–1139 (2015).
25. M. Nikkhoo, T. R. Walter, P. R. Lundgren, P. Prats-Iraola, *Geophys. J. Int.* **208**, 877–894 (2017).
26. J. H. Holland, *Adaptation in Natural and Artificial Systems* (Univ. of Michigan Press, 1975).
27. R. L. Haupt, S. E. Haupt, *Practical Genetic Algorithms* (Wiley, ed. 2, 2004).
28. H. J. Patton, S. R. Taylor, *J. Geophys. Res. Solid Earth* **116**, B03310 (2011).
29. P. Vincent *et al.*, *Geophys. Res. Lett.* **30**, 1–5 (2003).
30. X. Wang, Q.-F. Chen, J. Li, S. Wei, *Seismol. Res. Lett.* **87**, 901–911 (2016).
31. L. Zhu, Y. Ben-Zion, *Geophys. J. Int.* **194**, 839–843 (2013).
32. E. Teller, W. K. Talley, G. H. Higgins, G. W. Johnson, *The Constructive Uses of Nuclear Explosives* (McGraw-Hill, 1968).
33. M. D. Denny, L. R. Johnson, in *Explosion Source Phenomenology*, S. R. Taylor, H. J. Patton, P. G. Richards, Eds. (American Geophysical Union, 1991), pp. 1–24.
34. S. R. Ford, W. R. Walter, *Bull. Seismol. Soc. Am.* **103**, 2937–2945 (2013).
35. D. L. Springer, *Bull. Seismol. Soc. Am.* **56**, 861–876 (1966).
36. D. L. Springer, G. A. Pawloski, J. L. Ricca, R. F. Rohrer, D. K. Smith, *Bull. Seismol. Soc. Am.* **92**, 1806–1840 (2002).
37. S. J. Gibbons, F. Pabian, S. P. Näsholm, T. Kværna, S. Mykkeltveit, *Geophys. J. Int.* **208**, 101–117 (2016).

ACKNOWLEDGMENTS

We thank three anonymous reviewers for their comments. **Funding:** T.W., Q.S., S.W., and S.B. at Earth Observatory of Singapore (EOS) are supported by the Singapore Ministry of Education under the Research Centres of Excellence initiative, the National Research Foundation (NRF) of Singapore under the NRF Fellowship scheme (National Research Fellow Award NRF-NRFF2013-04), EOS startup fund M4430240.B50, and Ministry of Education Singapore Academic Research Fund Tier 1 RG181/16. M.N. is supported by VOLCAPSE, a research project funded by the European Research Council under the European Union's H2020 Programme/ERC consolidator grant ERC-CoG 646858. D.D. is supported by Air Force Research Laboratory contract FA9453-16-C-0024. Q.-F.C. is supported by National Science Foundation of China grant 41474041. This work constitutes Earth Observatory of Singapore contribution no. 185. **Author contributions:** T.W. analyzed SAR images provided by M.M.; Q.S. and S.W. conducted the seismic analysis using the data collected by Q.-F.C.; M.N. and S.B. conducted the geodetic modeling; D.D. estimated the yield; T.W., M.N., S.B., R.B., and M.M. contributed to the interpretation of SAR offsets and geodetic models; Q.S., S.W., D.D., and Q.-F.C. contributed to the interpretation of seismic waves and moment tensors; and T.W. drafted the manuscript with inputs from all co-authors. **Competing interests:** All authors declare no conflicts of interest. **Data and materials availability:** The TerraSAR-X images are copyright of German Aerospace Agency (DLR) and provided under proposal motagh_GGEO1217. The Advanced Land Observing Satellite-2 (ALOS-2) images are provided by JAXA under the ALOS-2 RA4 Research Project 1413 and 1161. SAR offsets, 3D displacement data points and regional seismic waveform data are provided in the supplementary materials. Teleseismic data are archived at the Incorporated Research Institutions for Seismology (IRIS).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/166/suppl/DC1
Materials and Methods
Figs. S1 to S16
Tables S1 to S3
References (*38*–*46*)
Data S1 and S2

22 December 2017; accepted 1 May 2018
Published online 10 May 2018
10.1126/science.aar7230

ORGANIC CHEMISTRY

Deconstructive fluorination of cyclic amines by carbon-carbon cleavage

Jose B. Roque*, Yusuke Kuroda*, Lucas T. Göttemann, Richmond Sarpong†

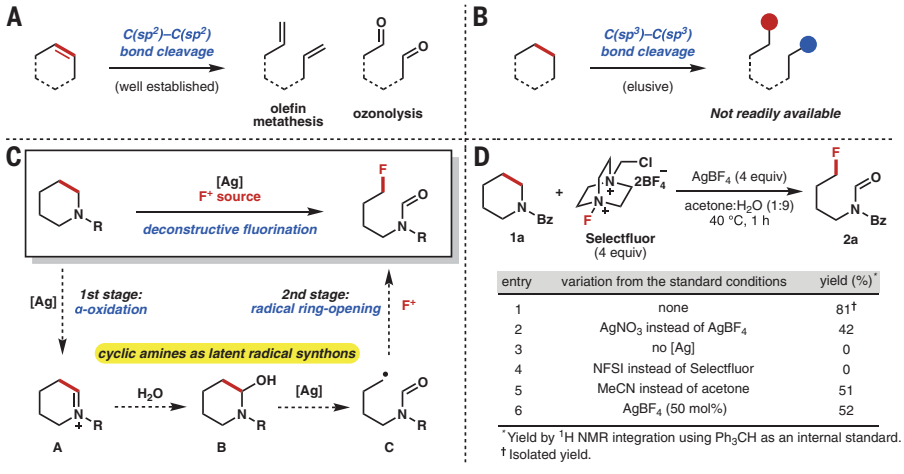
Deconstructive functionalizations involving scission of carbon-carbon double bonds are well established. In contrast, unstrained C(sp^3)–C(sp^3) bond cleavage and functionalization have less precedent. Here we report the use of deconstructive fluorination to access mono- and difluorinated amine derivatives by C(sp^3)–C(sp^3) bond cleavage in saturated nitrogen heterocycles such as piperidines and pyrrolidines. Silver-mediated ring-opening fluorination using Selectfluor highlights a strategy for cyclic amine functionalization and late-stage skeletal diversification, establishing cyclic amines as synthons for amino alkyl radicals and providing synthetic routes to valuable building blocks.

The ubiquity of carbon-carbon (C–C) bonds in organic compounds places a premium on methods that construct such bonds. In general, these bond constructions lead to an increase in structural complexity. However, in certain cases, the cleavage of C–C bonds may lead to more synthetically complex products that cannot be prepared efficiently in any other way. Historically, the full benefit of C–C bond cleavage (deconstructive strategies) has often been realized by coupling this process with value-added bond constructions such as C–C bond formation (e.g., in olefin metathesis processes) (*1*, *2*) or C–O bond formation (e.g., in ozonolysis) (*3*) (Fig. 1A). Whereas the benefits of these C(sp^2)=C(sp^2) double-bond cleavage and functionalization processes are well established, the value of deconstructive processes becomes even more apparent when C(sp^3)–C(sp^3) single-bond cleavage and functionalization are considered, especially in the context of late-stage skeletal diversification to access unexplored chemical space (*4*) (Fig. 1B). The development of deconstructive functionalizations of cyclic amines would be particularly useful, given their ubiquity in pharmaceuticals and agrochemicals (*5*, *6*). However, methods for the ring-opening of cyclic amines remain extremely limited and are dominated by C–O bond formation through α-oxidation followed by heterolytic C–N bond cleavage via the well-established equilibrium with the hemiaminal that forms (*7*–*12*). Although variants of the von Braun reaction of nucleophilic tertiary *N*-alkyl–substituted cyclic amines lead to C–Cl bond formation via heterolytic ring-opening, this strategy is limited to cyclic amines with small ring sizes owing to competing N-dealkylation (*13*). With these limitations in mind, we sought a mechanistically distinct strategy that would provide a general entry to the deconstructive functionalization of cyclic amines.

In this context, reactions that form C(sp^3)–F bonds are among the most valued bond constructions because of the influence of fluorine substitution on the properties of pharmaceuticals, agrochemicals, and organic materials (*14*–*17*). For example, installation of fluorine may lead to increased metabolic stability, altered physicochemical properties such as increased lipophilicity, reduced basicity of nearby nitrogen atoms, and conformational tuning. As a consequence, substantial progress has been made on site-selective reactions that form C(sp^3)–F bonds (*18*, *19*). Nevertheless, the development of methods for C(sp^3)–F bond formation that facilitate the preparation of a variety of fluorine-containing building blocks from easily available starting materials remains a prominent goal. Here we report a deconstructive strategy to transform cyclic amine derivatives into versatile fluorine-containing acyclic amine derivatives, using commercially available reagents, through C(sp^3)–C(sp^3) single-bond cleavage followed by C(sp^3)–F bond formation (Fig. 1C).

Our strategy for deconstructive fluorination of cyclic amine derivatives is based on two discrete stages, each mediated by a silver salt (Fig. 1C). In the first stage, we envisioned that under an appropriate set of oxidative conditions, a saturated cyclic amine would be oxidized to the corresponding iminium ion **A**, which would be trapped by H_2O to form hemiaminal **B**. In the second stage, the resulting hemiaminal **B** could undergo homolytic ring-opening upon engaging the silver salt to yield primary radical **C** (*20*). A subsequent fluorine atom transfer would deliver the desired fluorinated product. On this basis, cyclic amines could be viewed as synthons for amino alkyl radicals, which have conventionally only been shown to arise from the corresponding halide, alcohol, or carboxylic acid derivatives (*21*). Although this strategy is conceptually simple, several challenges were inherent in putting it into practice. First, only a few methods exist for oxidation of amines to the corresponding hemiaminals owing to the competing over-oxidation to amides (*22*). Furthermore, no reports exist of α-oxidation of cyclic amine derivatives using silver salts. However, given the oxidation potential of N-protected cyclic amines such as **1a** [anodic peak potential = +1.13 V versus saturated calomel electrode (SCE)] (fig. S1), we theorized that Ag(II) salts could be sufficiently oxidizing [standard reduction potential (Ag^{2+}/Ag^{+}) = +1.98 V versus SCE] (*23*) to effect single-electron transfer. Second, most of the reports of successful ring-opening fluorination are limited to strained tertiary cycloalkanols such as cyclobutanols (*24*–*27*); only limited examples that feature relatively unstrained cycloalkanols such as cyclopentanols and cyclohexanols are known, and these cases resulted in low yields (*25*). The challenge in achieving our envisioned transformation rested on identifying a silver salt and fluorinating reagent combination that would act in synergy to selectively cleave and functionalize the desired C–C bond.

We began our investigation by establishing the conditions for the overall transformation using *N*-Bz piperidine **1a** as the substrate (Bz, benzoyl)



entry	variation from the standard conditions	yield (%)*
1	none	81†
2	$AgNO_3$ instead of $AgBF_4$	42
3	no [Ag]	0
4	NFSI instead of Selectfluor	0
5	MeCN instead of acetone	51
6	$AgBF_4$ (50 mol%)	52

*Yield by ^{1}H NMR integration using Ph_3CH as an internal standard.
† Isolated yield.

Fig. 1. Development of a deconstructive fluorination of cyclic amines. (**A**) Well-established deconstructive functionalization. (**B**) An elusive deconstructive functionalization. (**C**) A blueprint for deconstructive fluorination of cyclic amines. (**D**) Optimization of silver-mediated deconstructive fluorination of *N*-Bz piperidine **1a**. R, any functional group; Me, methyl; Ph, phenyl; equiv, equivalents; h, hour.

Department of Chemistry, University of California, Berkeley, CA 94720, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: rsarpong@berkeley.edu

(Fig. 1D). After extensive screening of various conditions, we identified the optimized conditions shown in entry 1, which use cheap and commercially available $AgBF_4$ in a 9:1 (v/v) mixture of H_2O/acetone at 40°C. Other silver sources led to lower yields, with $AgNO_3$ providing the highest yield among them (entry 2). A control experiment established that a silver salt is essential to obtaining the desired fluorinated product (entry 3). Other fluorinating reagents such as *N*-fluorobenzenesulfonimide (NFSI) led to no reaction (entry 4). A 9:1 (v/v) mixture of H_2O/MeCN gave a diminished yield (51%) of **2a** (entry 5), pointing to the superiority of acetone as the cosolvent. The overall transformation can be conducted with substoichiometric amounts of $AgBF_4$ to provide **2a**, albeit in modest yield (entry 6).

With the optimized conditions established, we proceeded to investigate the scope of the deconstructive fluorination process. As shown in Fig. 2, several structurally and electronically distinct N-substituted piperidine derivatives were fluorinated effectively, including those bearing acetyl (Ac, **1b**), *tert*-butoxycarbonyl (Boc, **1c**), and pivaloyl (Piv, **1d**) groups. The deconstructive fluorination method is not limited to piperidine derivatives; a range of N-benzoylated saturated azacycles including azetidine **1e**, pyrrolidine **1f**, azepane **1g**, and azocane **1h** were all viable in the deconstructive fluorination reaction. Fluorinated products **2e** and **2f** were obtained in the deformylated form (vide infra). The variation of the cyclic amine substrate ring size led to fluoroamine derivatives bearing carbon chains of varying lengths. A variety of substitution patterns on the piperidine ring were also well tolerated, and the corresponding acyclic fluorinated amines were obtained in moderate to good yields (50 to 85%). Some of the fluorinated alkyl amine products have not been previously reported and would not be readily accessible by conventional deoxyfluorination strategies owing to the limited availability of the corresponding substituted linear amino alcohols. For example, 2-substituted piperidines **1j** and **1k** afforded the corresponding fluoroamines **2j** and **2k**, respectively, with complete positional selectivity. The observed selectivity for cleavage away from the substituents may be attributed to the steric hindrance imparted by these groups at the α-position of the cyclic amines. 3-Substituted piperidines were also good substrates, as evidenced by **1l** undergoing ring-opening and fluorination to provide **2l** in 50% yield, demonstrating that secondary alkyl fluorides can be accessed by this method. Fused piperidines such as **1m** underwent deconstructive fluorination to provide **2m** in 43% yield. This example demonstrates that polycyclic molecules can be functionalized as well, paving the way for late-stage skeletal diversification of complex molecules. Moreover, L-pipecolic acid derivative **1n** gave 5-fluoro- L-norvaline derivative **2n** in 68% yield (three steps from L-pipecolic acid), considerably shortening the synthesis of 5-fluoro-L-norvaline (previously prepared in seven steps from L-glutamic acid) (*28*). L-proline methyl ester derivative **1o** was converted to pyrrolidinone derivative **2o**, presumably by over-oxidation of an intermediate 5-hydroxyproline derivative. Similarly, *N*-methyl-2-piperidinone (**1p**) and *N*-methyl-2-pyrrolidinone (**1q**) were oxygenated under the reaction conditions to give *N*-methyl imides **2p** and **2q**, respectively. These oxygenation reactions are important in their own right because methods for the direct α-oxygenation of cyclic amides are dominated by strongly oxidizing RuO_4 (*22*). *N*-Piv-2-pyrrolidinone **1r** afforded fluorinated product **2r** under the same conditions. Moreover, piperidines containing carboxylic acid groups underwent dual functionalization to provide difluorinated amines through decarboxylative (*29*) and deconstructive fluorination. For example, *N*-Bz piperidine **1s** and **1t** underwent dual fluorination to provide 3,5-difluorinated amine **2s** (28%) and 4,5-difluorinated amine **2t** (22%), along with *N*-Bz 3-fluoropiperidine (28%) and *N*-Bz 4-fluoropiperidine (20%), respectively.

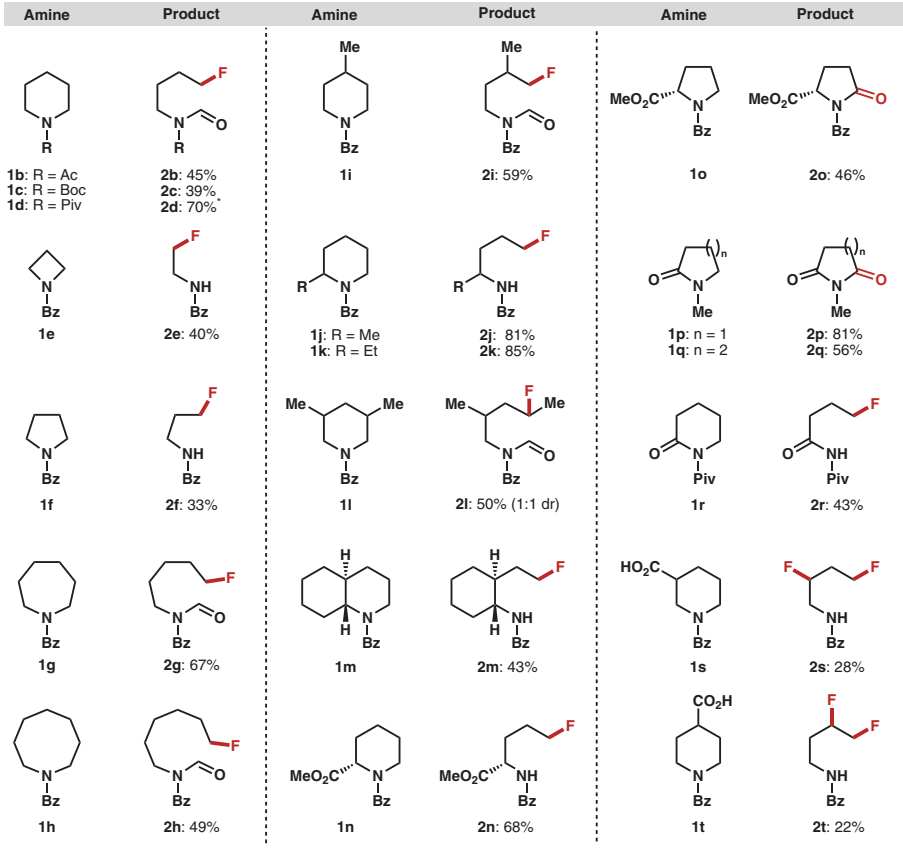


Fig. 2. Deconstructive fluorination: cyclic amine scope. Only isolated yields are shown. Reaction conditions: **1** (0.1 mmol), $AgBF_4$ (4 equivalents), Selectfluor (4 equivalents), acetone:H_2O (1:9), 40°C, 1 hour. *Deformylated product obtained. dr, diastereomeric ratio.

As a demonstration of the utility of this method, we considered functionalizing synthetic peptides, which continue to see widespread use in drug discovery (*30*–*32*). Deconstructive functionalization of peptides can provide orthogonal and complementary skeletal diversification and add profitably to the toolbox of available methods (*33*). When dipeptide **3a**, which has a valine residue, was subjected to our reaction conditions, the fluorination proceeded readily to afford fluorinated dipeptide **4a** in 50% yield along with 25% recovered starting material (Fig. 3). In this case, prolonged reaction times did not lead to an increase in the product yield, presumably owing to product inhibition. Internal peptides such as **3b** also underwent deconstructive fluorination to provide **4b** in 38% yield (accompanied by 40% of recovered starting material). Likewise, tripeptide **3c** was converted to **4c** in 39% yield, along with 29% of recovered starting material. This methodology could also be applied to the selective C5-oxygenation of L-proline–containing peptides. For example, peptide **3d** was oxygenated in 76% yield to provide **4d**. *N*-Benzoyl imide–containing peptide **4d** served as a versatile intermediate for further functionalization. We did not observe racemization of the fluorinated and oxygenated peptides, which were obtained as single diastereomers.

A number of additional experiments were performed to elucidate the reaction mechanism. We began by using nuclear magnetic resonance (NMR) spectroscopy to investigate the interaction of Ag(I) and Selectfluor. Unexpectedly, a ^{19}F NMR spectrum of an equimolar mixture of Selectfluor and $AgBF_4$ in a 1:9 (v/v) mixture of acetone-d_6/D_2O, acquired after stirring at 40°C for 1 hour (fig. S2), displayed no consumption of Selectfluor (*34*, *35*). However, in the presence of an equivalent of **1a**, consumption of Selectfluor was observed, suggesting that the N-protected cyclic

amine substrates play an important role in initiating the ring-opening and fluorination process. In addition, line broadening in the ^{1}H NMR spectrum was observed, suggesting the formation of a paramagnetic Ag(II) complex. Furthermore, downfield shifts of NMR resonances of **1a** were observed in the ^{1}H NMR spectra upon addition of $AgBF_4$ (fig. S3), suggesting the binding of Ag(I) to the amide moiety of **1a** (*36*, *37*). On the basis of these NMR experiments, we propose a mechanism that involves initial coordination of Ag(I) to **1a**, followed by single-electron oxidation by Selectfluor to form Ag(II) and radical dication **5** (Eq. 1 in Fig. 4A) (*35*). The resulting Ag(II) species would undergo single-electron transfer from **1a**. Subsequent hydrogen-atom abstraction by **5** (*38*) would deliver iminium ion **A**, and this would be followed by trapping by H_2O to give hemiaminal **B** (Eq. 2 in Fig. 4A). An alternative mechanistic pathway wherein radical dication **5** undergoes α-amino C–H abstraction of **1a** followed by single-electron transfer by Ag(II) to generate the same iminium ion **A** cannot be ruled out. From hemiaminal **B**, an alkoxy Ag(II) intermediate may form (not shown). Opening of this intermediate to primary radical **C** would achieve the desired C(sp^3)–C(sp^3) bond cleavage, and attendant fluorination of the radical by Selectfluor would yield **2a** (path A in Fig. 4B). However, we recognize that another pathway could be operable based on the fact that deformylated products were obtained in some cases. In this alternate pathway, opening of the hemiaminal to linear aldehyde **D** and subsequent oxidation to the corresponding carboxylic acid would then set the stage for a decarboxylative fluorination—in line with the precedent of (*29*) (path B in Fig. 4B). In a series of experiments to support or refute either mechanism (Fig. 4C), aldehyde **D**, which likely exists in equilibrium with hemiaminal **B**, was subjected to our reaction conditions and gave fluoroamine **2a** in 55% yield, which can only be accessed through path A. In addition, when the reaction was conducted over a prolonged period, the benzoyl amide was obtained as the major product, indicating that the conversion of **2a** to **6** likely occurs through a deformylation process. However, the successful fluorination of *N*-phthaloyl aldehyde **7** demonstrates that fluorination can proceed from the aldehyde, which cannot form the hemiaminal. That **7** gave a higher yield than **1a** and **D** is a result of the relative stability of the product **8** under the oxidative reaction conditions that were used. Subjecting carboxylic acid **9** to the optimized conditions resulted in 23% yield of fluorinated product **6**. On the basis of these experiments, we cannot rule out the possibility that path B is operative for the small subset of substrates that gave exclusively deformylated products. Lastly, on the basis of our mechanistic proposal, we sought to explore the reactivity of enamides under our optimized reaction conditions. We envisioned enamide **10** undergoing electrophilic fluorination followed by trapping of the resulting carbocation with water to yield **11** (*39*). An alkoxy Ag(II) intermediate would follow, leading to C(sp^3)–C(sp^3) cleavage and fluorination to yield gem-difluorinated protected amine **12**. As shown in Fig. 4D, in a reaction using catalytic amounts of silver salts, a variety of enamides related to **10** underwent the desired deconstructive difluorination to yield **12a** to **12c** in 54 to 76% yield. In support of path A (Fig. 4A), the formyl imide products were isolated as the major product under the optimized reaction conditions. These results are productive given the established importance of difluoromethyl groups (*40*, *41*). For example, the difluoromethyl moiety serves as a lipophilic hydrogen bond donor that acts as a bioisostere for thiol and hydroxyl groups.

The simple protocol of the deconstructive fluorination, which proceeds in aqueous solvent mixtures as well as in water alone (*42*, *43*), should lead to its widespread adoption for late-stage skeletal diversification in the pharmaceutical

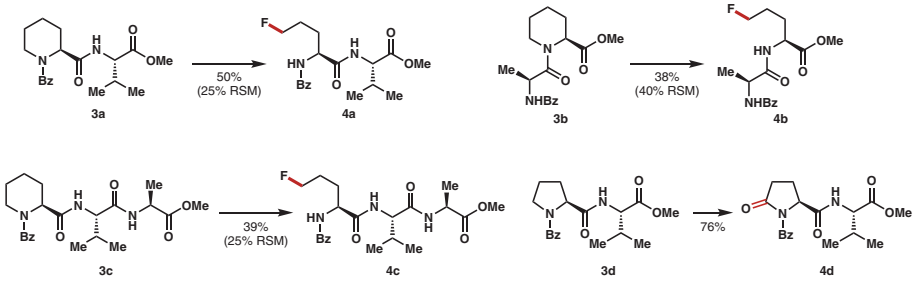


Fig. 3. Diversification of pipecolic acid and proline residues in peptides. Isolated yields are shown. Isolated yields of recovered starting material (RSM) are given in parentheses. Reaction conditions: **3** (0.1 mmol), $AgBF_4$ (4 equivalents), Selectfluor (4 equivalents), acetone:H_2O (1:9), room temperature, 15 hours.

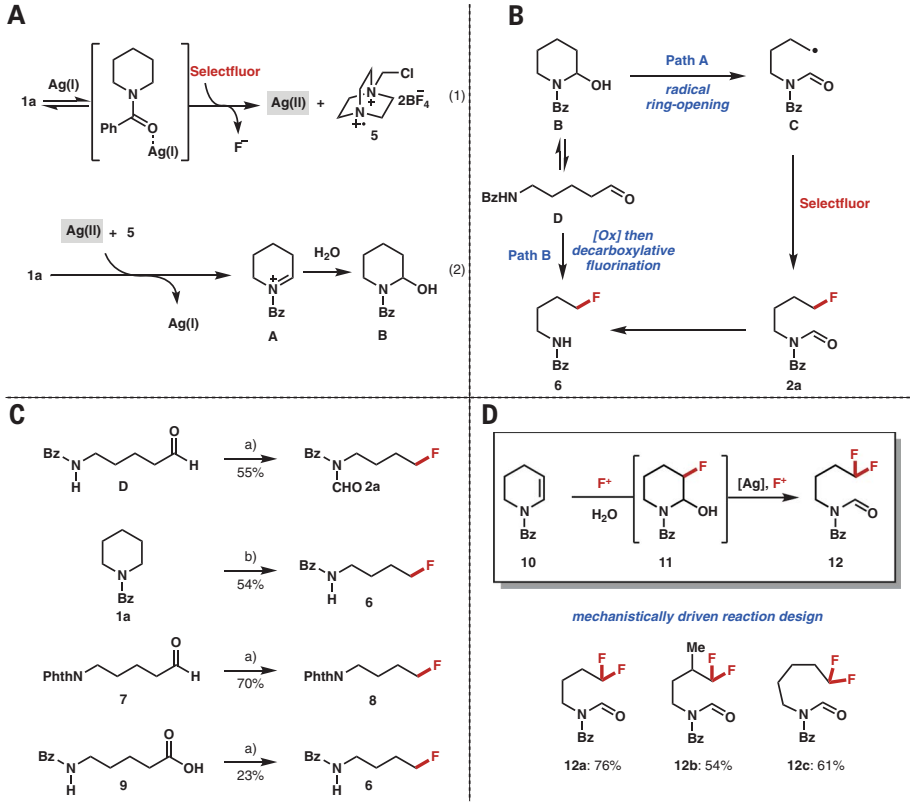


Fig. 4. Mechanistic studies. (**A**) Proposed mechanism for **1a** oxidation. (**B**) Possible mechanisms for fluorination of **B**. (**C**) Mechanistic studies. Reaction conditions: (a) starting material (0.1 mmol), $AgBF_4$ (4 equivalents), Selectfluor (4 equivalents), acetone:H_2O (1:9), 40°C, 1 hour; (b) **1a** (0.5 mmol), $AgBF_4$ (4 equivalents), Selectfluor (4 equivalents), acetone:H_2O (1:9), room temperature, 16 hours. (**D**) Mechanistically driven gem-fluorination of enamide **10**. Reaction conditions: **10** (0.1 mmol), $AgBF_4$ (0.25 equivalents), Selectfluor (4 equivalents), acetone:H_2O (1:1), room temperature, 15 hours. Phth, phthaloyl.

and agrochemical arena. From the retrosynthetic viewpoint, cyclic amines can now be regarded as synthons for amino alkyl radical intermediates, which can be engaged by a variety of coupling partners. Thus, we anticipate that this method will unlock fundamentally different disconnection strategies.

REFERENCES AND NOTES

1. A. H. Hoveyda, A. R. Zhugralin, *Nature* **450**, 243–251 (2007).
2. G. C. Vougioukalakis, R. H. Grubbs, *Chem. Rev.* **110**, 1746–1787 (2010).
3. P. S. Bailey, *Chem. Rev.* **58**, 925–1010 (1958).
4. S. K. Silverman, P. J. Hergenrother, *Curr. Opin. Chem. Biol.* **10**, 185–187 (2006).
5. E. Vitaku, D. T. Smith, J. T. Njardarson, *J. Med. Chem.* **57**, 10257–10274 (2014).
6. S. A. Lawrence, *Amines: Synthesis, Properties and Applications* (Cambridge Univ. Press, 2004).
7. A. P. Shawcross, S. P. Stanforth, *J. Heterocycl. Chem.* **27**, 367–369 (1990).
8. G. Han, M. C. McIntosh, S. M. Weinreb, *Tetrahedron Lett.* **35**, 5813–5816 (1994).
9. G. Cocquet, C. Ferroud, A. Guy, *Tetrahedron* **56**, 2975–2984 (2000).
10. R. Ito, N. Umezawa, T. Higuchi, *J. Am. Chem. Soc.* **127**, 834–835 (2005).
11. M. Kaname, S. Yoshifuji, H. Sashida, *Tetrahedron Lett.* **49**, 2786–2788 (2008).
12. T. J. Osberger, D. C. Rogness, J. T. Kohrt, A. F. Stepan, M. C. White, *Nature* **537**, 214–219 (2016).
13. C. Yu *et al*., *J. Org. Chem.* **82**, 6615–6620 (2017).
14. E. P. Gillis, K. J. Eastman, M. D. Hill, D. J. Donnelly, N. A. Meanwell, *J. Med. Chem.* **58**, 8315–8359 (2015).
15. J. Wang *et al*., *Chem. Rev.* **114**, 2432–2506 (2014).
16. S. Purser, P. R. Moore, S. Swallow, V. Gouverneur, *Chem. Soc. Rev.* **37**, 320–330 (2008).
17. K. Müller, C. Faeh, F. Diederich, *Science* **317**, 1881–1886 (2007).
18. P. A. Champagne, J. Desroches, J.-D. Hamel, M. Vandamme, J.-F. Paquin, *Chem. Rev.* **115**, 9073–9174 (2015).
19. B. Lantaño, A. Postigo, *Org. Biomol. Chem.* **15**, 9954–9973 (2017).
20. M. Murakami, N. Ishida, *Chem. Lett.* **46**, 1692–1700 (2017).
21. M. Yan, J. C. Lo, J. T. Edwards, P. S. Baran, *J. Am. Chem. Soc.* **138**, 12692–12714 (2016).
22. J. Sperry, *Synthesis* **2011**, 3569–3580 (2011).
23. H. N. Po, *Coord. Chem. Rev.* **20**, 171–195 (1976).
24. H. Zhao, X. Fan, J. Yu, C. Zhu, *J. Am. Chem. Soc.* **137**, 3490–3493 (2015).
25. S. Ren, C. Feng, T.-P. Loh, *Org. Biomol. Chem.* **13**, 5105–5109 (2015).
26. Q. Tian, B. Chen, G. Zhang, *Green Chem.* **18**, 6236–6240 (2016).
27. Y. Deng, N. I. Kauser, S. M. Islam, J. T. Mohr, *Eur. J. Org. Chem.* **2017**, 5872–5879 (2017).
28. L. Wang *et al*., *Nucl. Med. Biol.* **39**, 933–943 (2012).
29. F. Yin, Z. Wang, Z. Li, C. Li, *J. Am. Chem. Soc.* **134**, 10401–10404 (2012).
30. Z. Antosova, M. Mackova, V. Kral, T. Macek, *Trends Biotechnol.* **27**, 628–635 (2009).
31. P. Vlieghe, V. Lisowski, J. Martinez, M. Khrestchatisky, *Drug Discov. Today* **15**, 40–56 (2010).
32. A. A. Kaspar, J. M. Reichert, *Drug Discov. Today* **18**, 807–817 (2013).
33. S. Sengupta, G. Mehta, *Tetrahedron Lett.* **58**, 1357–1372 (2017).
34. Selectfluor is reported to react with $AgNO_3$ in acetone-*d6*/D_2O (*35*).
35. N. R. Patel, R. A. Flowers 2nd, *J. Org. Chem.* **80**, 5834–5841 (2015).
36. V. Romanov *et al*., *J. Phys. Chem. A* **112**, 10912–10920 (2008).
37. V. Romanov, C.-K. Siu, U. H. Verkerk, A. C. Hopkinson, K. W. M. Siu, *J. Phys. Chem. A* **114**, 6964–6971 (2010).
38. C. R. Pitts *et al*., *J. Am. Chem. Soc.* **136**, 9780–9791 (2014).
39. S. Singh *et al*., *Synlett* **23**, 2421–2425 (2012).
40. N. A. Meanwell, *J. Med. Chem.* **54**, 2529–2591 (2011).
41. C. D. Sessler *et al*., *J. Am. Chem. Soc.* **139**, 9325–9332 (2017).
42. C. J. Li, L. Chen, *Chem. Soc. Rev.* **35**, 68–82 (2006).
43. M. O. Simon, C. J. Li, *Chem. Soc. Rev.* **41**, 1415–1427 (2012).

ACKNOWLEDGMENTS

We thank J. Derrick for assistance with electrochemical measurements. **Funding:** This work was supported by the National Institutes of Health (NIGMS RO1 086374). J.B.R. thanks the NIH (NIGMS RO1 086374) for a graduate diversity supplement fellowship. Y.K. thanks the Japan Society for the Promotion of Science (JSPS) for an Overseas Research Fellowship. L.T.G. thanks LMU PROSA and DAAD for financial support. **Author contributions:** J.B.R. and Y.K. discovered the reaction and designed the experiments. J.B.R., Y.K., and L.T.G. performed the experiments. R.S. directed the project. J.B.R., Y.K., and R.S. wrote the manuscript. **Competing interests:** J.B.R., Y.K., L.T.G., and R.S. are listed as inventors on an initial patent application describing the Ag-mediated deconstructive functionalization of cyclic amines (052103-515P01US). **Data and materials availability:** Experimental and characterization data are provided in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/171/suppl/DC1
Materials and Methods
Figs. S1 to S3
NMR Spectra
References (*44*–*63*)

19 March 2018; accepted 15 May 2018
10.1126/science.aat6365

EARLY OCEAN

Late inception of a resiliently oxygenated upper ocean

Wanyi Lu[1], Andy Ridgwell[2,3], Ellen Thomas[4,5], Dalton S. Hardisty[6], Genming Luo[7], Thomas J. Algeo[7,8,9], Matthew R. Saltzman[10], Benjamin C. Gill[11], Yanan Shen[12], Hong-Fei Ling[13], Cole T. Edwards[14], Michael T. Whalen[15], Xiaoli Zhou[1], Kristina M. Gutchess[1], Li Jin[16], Rosalind E. M. Rickaby[17], Hugh C. Jenkyns[17], Timothy W. Lyons[2], Timothy M. Lenton[18], Lee R. Kump[19], Zunli Lu[1]*

Rising oceanic and atmospheric oxygen levels through time have been crucial to enhanced habitability of surface Earth environments. Few redox proxies can track secular variations in dissolved oxygen concentrations around threshold levels for metazoan survival in the upper ocean. We present an extensive compilation of iodine-to-calcium ratios (I/Ca) in marine carbonates. Our record supports a major rise in the partial pressure of oxygen in the atmosphere at ~400 million years (Ma) ago and reveals a step change in the oxygenation of the upper ocean to relatively sustainable near-modern conditions at ~200 Ma ago. An Earth system model demonstrates that a shift in organic matter remineralization to greater depths, which may have been due to increasing size and biomineralization of eukaryotic plankton, likely drove the I/Ca signals at ~200 Ma ago.

The evolution and survival of marine animals depended on oxygen availability, particularly in upper ocean waters, ranging from the sea surface to the thermocline, during early Earth history (*1*). The dissolved oxygen concentration ($[O_2]$) in the upper ocean commonly decreases from the well-mixed surface ocean (the top few tens of meters) into deeper subsurface waters (a few hundred meters). This $[O_2]$ gradient is controlled by three key factors: (i) the partial pressure of oxygen in the atmosphere (P_{O_2}), (ii) the intensity of upper-ocean mixing, and (iii) the oxidation of organic matter in the water column, which consumes oxygen (*2*). Atmospheric P_{O_2} changes through time have been estimated via geochemical proxy data and box models (*3*). Oceanic paleoredox proxies typically track the areal extent of euxinic waters (containing H_2S) and the presence or absence of anoxia (positive or zero $[O_2]$) (*4*, *5*). Because most modern marine animals are sensitive to $[O_2]$ changes between ~10 and ~100 μmol/kg (*2*), development of long-term proxy reconstructions for $[O_2]$ in this critical range (oxic-hypoxic) would help elucidate when and how oceanic oxygenation evolved to accommodate the modern ecological landscape.

Carbonate I/Ca is one of the proxies developed for the oxic-hypoxic window that has the potential to reconstruct secular trends in upper-ocean oxygenation (*6*, *7*). The long residence time of iodine (~300,000 years) leads to generally uniform total iodine concentrations in the modern ocean, but speciation changes of iodine between iodate (IO_3^-) and iodide (I^-) are controlled locally (*8*, *9*). IO_3^- is completely reduced to I^- in waters at low $[O_2]$ (*8*, *9*) and reoxidized under well-oxygenated conditions. Because IO_3^- is the only chemical form of iodine incorporated into the carbonate structure (*7*) by replacing the CO_3^{2-} ion (*10*), carbonate I/Ca records of local seawater $[IO_3^-]$ through time can indicate changes in $[O_2]$. Carbonate I/Ca has been shown to be a reliable tracer that responds primarily to $[O_2]$ variations in marine environments over a wide range of geological periods (*6*, *11*–*16*).

We measured I/Ca in an extensive Phanerozoic collection of shallow marine carbonates likely forming within the top 200 m of the water column and compiled these values with published data (table S1 and Fig. 1A). Maximum I/Ca values for individual localities were generally low in the Proterozoic, except for periods that have been associated with potential atmospheric P_{O_2} rises [e.g., the Great Oxidation Event (*12*)] and some Neoproterozoic carbon isotope excursions (*11*, *15*), when maximum values temporarily increased to Cenozoic levels (3 to 4 μmol/mol) (Fig. 1A). Paleozoic maximum values are comparable to those of the Proterozoic, despite a relatively short spike during the Devonian, at about 400 million years (Ma) ago, when the 75th percentile values reached Cenozoic levels. Break-point analyses indicate a step change at Triassic to early Jurassic (~200 Ma ago, fig. S1), after which maximum values remain above 4 μmol/mol and 75th percentile values are mostly higher than 3 μmol/mol (*6*, *13*). The Devonian I/Ca excursion and the step change at ~200 Ma ago are two key observations in this data compilation.

The stark contrast between predominantly low Paleozoic values and high Meso-Cenozoic values (excluding the Triassic, i.e., <200 Ma ago) cannot be explained by sampling biases. The sample size for the Paleozoic (n = 894) is comparable to that for the Proterozoic (n = 1078) and the Meso-Cenozoic (n = 926). The sampling density (number of samples per unit time) is similar for the Paleozoic and Meso-Cenozoic, although lower in the Proterozoic (Fig. 1B). For Paleozoic samples, we targeted carbonate- and fossil-rich (shallow) continental-shelf locations, i.e., relatively well-oxygenated settings, which are prone to recording high I/Ca values. By contrast, many Mesozoic data were generated from sections recording well-established global oceanic anoxic events, which, if anything, would bias that dataset toward low values. Cenozoic I/Ca values were measured in sediment coarse fractions, which may better preserve primary I/Ca signatures than bulk-rock samples. Existing early Cenozoic (*14*) and Cretaceous I/Ca data (*13*) are generally comparable across different lithologies, although comparisons between the Cenozoic and other periods are more tenuous. The current dataset has relatively denser sample coverage for intervals coinciding with Earth system perturbations (e.g., major carbon-isotope excursions and mass extinctions) than for extended intervals with limited environmental changes, but this should not influence main features of the data compilation.

I/Ca values can potentially be reduced during subaerial exposure, marine burial, and dolomitization, but no postdepositional alterations are known to increase I/Ca (*11*). A variety of diagenesis indicators were considered in previous studies of samples that we used here (*7*, *15*–*17*). In all of those case studies, the number of potentially altered samples was limited and did not influence the central trend of most of the data through time, as represented by the 25th and 75th percentile values (Fig. 1A). High I/Ca values throughout the record (Fig. 1A) are not consistently tied to a specific inferred primary carbonate mineralogy (e.g., calcite versus aragonite seas, fig. S2A). The distinct behaviors of

[1]Department of Earth Sciences, Syracuse University, Syracuse, NY, USA. [2]Department of Earth Sciences, University of California, Riverside, Riverside, CA, USA. [3]School of Geographical Sciences, University of Bristol, Bristol, UK. [4]Department of Geology and Geophysics, Yale University, New Haven, CT, USA. [5]Department of Earth and Environmental Sciences, Wesleyan University, Middletown, CT, USA. [6]Department of Earth and Environmental Sciences, Michigan State University, East Lansing, MI, USA. [7]State Key Laboratory of Biogeology and Environmental Geology and School of Earth Science, China University of Geosciences, Wuhan, China. [8]State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan, China. [9]Department of Geology, University of Cincinnati, Cincinnati, OH, USA. [10]School of Earth Sciences, The Ohio State University, Columbus, OH, USA. [11]Department of Geosciences, Virginia Polytechnic and State University, Blacksburg, VA, USA. [12]School of Earth and Space Sciences, University of Science and Technology of China, Hefei, China. [13]State Key Laboratory for Mineral Deposits Research, School of Earth Sciences and Engineering, Nanjing University, Nanjing, China. [14]Department of Geological and Environmental Sciences, Appalachian State University, Boone, NC, USA. [15]Department of Geosciences, University of Alaska, Fairbanks, Fairbanks, AK, USA. [16]Geology Department, State University of New York College at Cortland, Cortland, NY, USA. [17]Department of Earth Sciences, University of Oxford, Oxford, UK. [18]Earth System Science Group, College of Life and Environmental Sciences, University of Exeter, Exeter, UK. [19]Department of Geosciences, Pennsylvania State University, University Park, PA, USA.
*Corresponding author. Email: zunlilu@syr.edu

I/Ca before and after ~200 Ma ago (Fig. 1A) cannot be explained by secular changes in seawater [Ca^{2+}] (fig. S2B). No evidence suggests that differences in Paleozoic and Mesozoic I/Ca distributions were due to uniformly greater alteration of the Paleozoic samples (fig. S3). Lower relative standard deviations of neighboring samples in each section (i.e., smoother I/Ca profiles; fig. S4) may reflect better preservation of the Paleozoic than the Proterozoic samples (fig. S4).

We interpret I/Ca in marine carbonates primarily as a qualitative indicator for the depth of the oxycline (Fig. 2), that is, that part of the water column where the [O_2] decreases relatively abruptly. Carbonate rocks formed in the upper ocean record surface or near-surface seawater [IO_3^-], which is strongly affected by the presence or absence of a proximal oxygen minimum zone (OMZ) or a shallow oxycline. Owing to the relatively slow oxidation kinetics of I^- (*18*), surface waters may retain a low IO_3^- signal despite high in situ [O_2]. For instance, core-top (modern) planktonic foraminiferal shells exhibit low I/Ca values (~0.5 μmol/mol) in waters above a shallow OMZ in the equatorial Pacific but record higher values (>3 μmol/mol) at other well-oxygenated locations (*6*).

The large I/Ca excursion during the Devonian (Fig. 3) most likely reflects deepening of the oxycline and development of better oxygenated conditions in the upper ocean, consistent with published proxy data and modeling results (*4*, *5*, *19*, *20*). Although different box models yield somewhat divergent interpretations of atmospheric P_{O_2} variation through the Phanerozoic (*21*–*24*), a Devonian rise in P_{O_2} levels is plausible, on the basis of the carbon-oxygen-phosphorus-sulfur evolution (COPSE) model and charcoal proxy reconstructions (Fig. 3A), and was most likely due to increased abundance of vascular land plants (*19*, *20*). Previous work interpreted δ^{98}Mo, iron-speciation, and biological data (Fig. 3C) to reflect oceanic redox changes, supporting the idea of atmospheric P_{O_2} rise during the Devonian (*4*, *5*). The combination of these independent proxies indicates that the Devonian atmospheric P_{O_2} rise affected the whole atmosphere-ocean system, across the entire redox spectrum (Fig. 3).

I/Ca values returned to Proterozoic-like levels after the transient Devonian excursion, but there is no evidence for a P_{O_2} decrease to pre-Devonian levels between the Carboniferous and the Triassic. Instead, the post-Devonian atmosphere was probably relatively O_2 rich (Fig. 3A). High atmospheric P_{O_2} likely altered terrestrial weathering feedbacks and enhanced nutrient delivery to the ocean (*25*, *26*), leading to intensified O_2 consumption in the upper ocean, a generally shallow oxycline, and low I/Ca values between ~400 and 200 Ma ago (Fig. 3B). Under such conditions, carbonates formed in surface oceans rapidly equilibrated with the high-P_{O_2} atmosphere would record low in situ [IO_3^-] because of the slow oxidation of I^- during mixing between surface and subsurface waters (Fig. 2B). If the oxycline were indeed shallow, marine animals on continental shelves at that time (~400 to 200 Ma ago) would have been living in a thin layer of well-oxygenated surface water directly underlain by an OMZ (Fig. 2B). Our dataset (Fig. 1A) implies that well-oxygenated upper-ocean conditions became persistent and resilient only by the Triassic-Jurassic (~200 Ma ago), much later than previously inferred (*27*). The prerequisites for achieving such well-oxygenated upper-ocean conditions are a combination of high atmospheric P_{O_2} and a generally deep oxycline (Fig. 2C). The position of the oxycline is strongly controlled by the depth of organic-matter

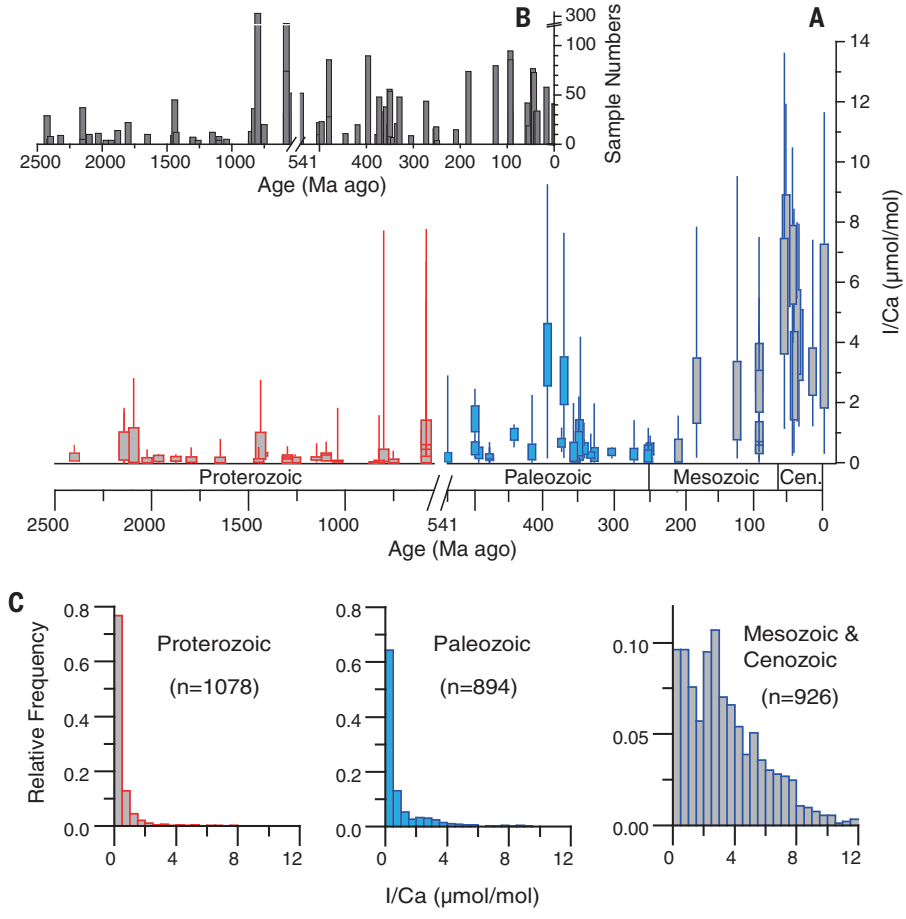


Fig. 1. Carbonate I/Ca through time. (**A**) Candlestick plot showing ranges of I/Ca values for Proterozoic (red) (*11*, *15*), Paleozoic (blue), and Meso-Cenozoic (purple). Boxes mark the 25th and 75th percentiles of values at each locality, and the whiskers show the maximum and minimum. The Proterozoic values from dolostones are I/(Mg + Ca). (**B**) Number of samples measured at each section. (**C**) Relative frequency distributions of I/Ca.

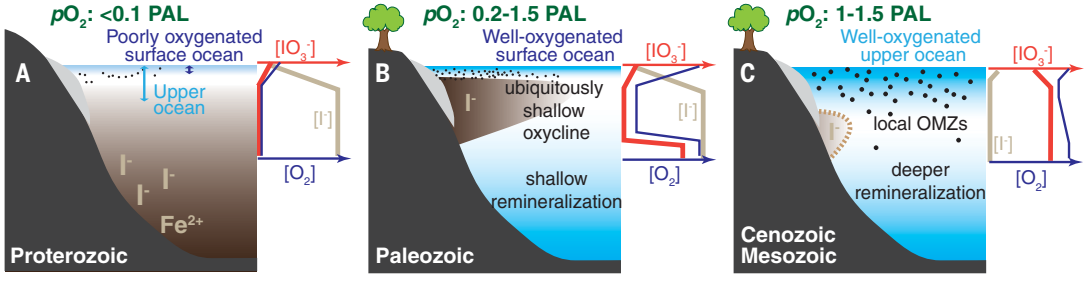


Fig. 2. Schematic illustrations for the evolution of oxygenation conditions. These simplified cartoons are not intended to capture all temporal and spatial variations. (**A**) Proterozoic. (**B**) Paleozoic. (**C**) Mesozoic and Cenozoic.

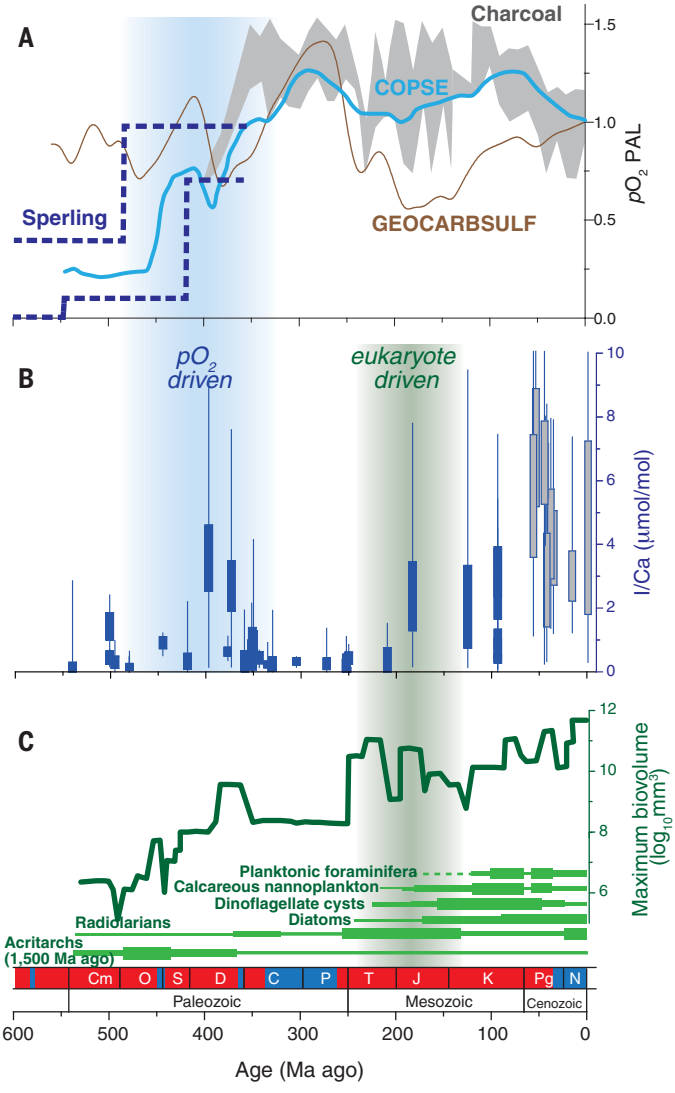


Fig. 3. Phanerozoic I/Ca compared with atmospheric P_{O_2}, oceanic oxygenation, and animal evolution trends. (**A**) Modeled atmospheric P_{O_2} curves (*5*, *21*, *23*, *24*) in comparison with the charcoal proxy record (*20*). Dashed lines (*5*) mark a broadly defined ocean-atmospheric O_2 level, not just atmospheric P_{O_2}. (**B**) I/Ca records through the Phanerozoic. Blue boxes indicate bulk carbonate rock; gray boxes indicate bulk coarse fraction of Ocean Drilling Project samples (>63 µm). (**C**) Marine animal body-size record (*36*). The thickness of the green bars indicates relative generic diversity modified from literature (*31*, *32*). The red versus blue bars mark greenhouse versus icehouse climate conditions, respectively. Cm, Cambrian; O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; T, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene; N, Neogene.

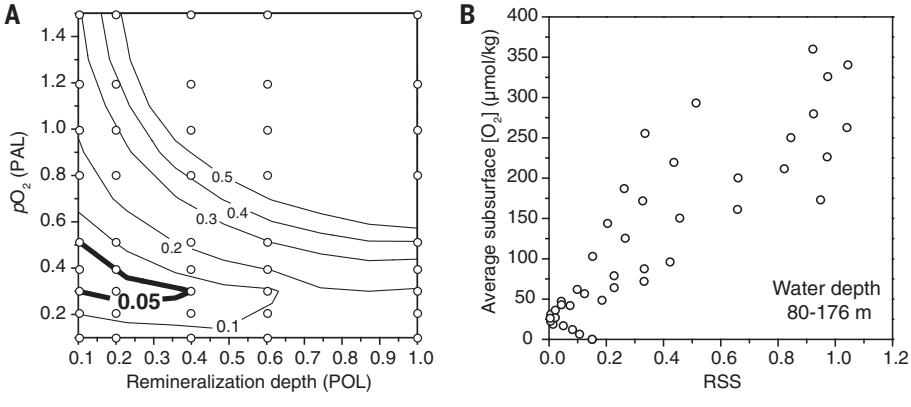


Fig. 4. Residual sum of squares and subsurface [O_2] at different P_{O_2} levels and remineralization depths. (**A**) Shallow remineralization depths produce the best model fit (the smallest RSS < 0.05) to Paleozoic I/Ca distribution, at 1x CO_2 condition. The circles represent 45 cGENIE simulations defining the contours. (**B**) Averaged [O_2] in the shallowest subsurface layer in each cGENIE run as an indicator of oxycline depth correlating with the RSS.

remineralization, which is dependent on the efficiency of organic-matter export from the photic zone, and has been proposed as a governing parameter for OMZs during the Phanerozoic (*28*).

We hypothesize that changes in remineralization of organic matter strongly influenced the upper-ocean I/Ca signature (Fig. 2), and we tested this hypothesis by simulating the marine iodine cycle in the cGENIE Earth system model (see methods and fig. S5) (*29*). We aimed to identify possible causes for low I/Ca during the Paleozoic through ensembles of model runs using a range of values for atmospheric P_{O_2}, the depth of organic-matter remineralization in the water column, and the mean concentrations of iodine and phosphate in seawater. For each Paleozoic model run, surface-water [IO_3^-] values along continental margins were extracted to calculate a relative frequency distribution (fig. S6). The modeled IO_3^- distributions were compared with observed Paleozoic I/Ca distributions (Fig. 1C) to obtain the residual sum of squares (RSS) (see materials and methods, Fig. 4A, and fig. S7).

We found that the lowest RSS values (<0.05), representing the best data-model fits, were achieved at shallow remineralization depths [i.e., <0.5 present oceanic level (POL)]. In the same set of cGENIE runs (Fig. 4B), lower RSS values correlated with lower average [O_2] in the subsurface layer (80 to 176 m), which is consistent with a shallower oxycline. Even as Paleozoic oceans experienced transitions between greenhouse and icehouse climate conditions, P_{CO_2} levels appear to have had minimal influence on IO_3^- distributions (fig. S8). Global-scale changes in ocean circulation and continent configuration also do not notably influence the oxycline depths independently of P_{O_2} and subsurface oxygen consumption (fig. S9). The RSS contours differed only slightly when the Paleozoic I/Ca distribution was compared with modeled [IO_3^-] distributions in the top four layers in the upper ocean (from 0 to 410 m, fig. S10). Thus, a lack of precise constraints on the paleodepths of carbonate formation is unlikely to have affected the main conclusions of our data-model comparison. Additional model runs also suggest that oceanic nutrient concentrations and total iodine concentrations are unlikely to dominate the secular trends in proxy data (figs. S11 and S12). Our data-model comparison (Fig. 4A) should not be viewed as a precise estimate of the atmospheric P_{O_2} for any single time slice, because the data were compiled over the entire Paleozoic under varying P_{O_2} levels. Thus, the lower RSS values at P_{O_2} below 1 present atmospheric level (PAL) suggest that some portions of the Paleozoic may have had P_{O_2} levels lower than those of today (*5*, *23*).

On the basis of our data compilation and model analyses, we attribute the transition at ~200 Ma ago from Proterozoic-like low I/Ca values in the Paleozoic (except for the mid-Devonian) to modern-like high values in the Meso-Cenozoic to a profound increase in the average remineralization depth of organic matter in the water column. The timing of this transition is consistent

with the proliferation of eukaryotic phyto- and zooplankton after the Permian-Triassic extinction (Fig. 3A) (*30*, *31*), which eventually shaped the ecological landscape of the modern ocean (*32*). The larger size of primary producers (*33*), grazing and repackaging of organic matter into fecal pellets (*34*), and/or the advent of mineralized plankton (*32*) may have led to faster sinking of organic matter, which reduced O_2 utilization in the upper water column and caused a pervasive deepening of the oxycline (*28*).

The rise of oxygen levels over geological time has been linked to increases in animal body size (*24*, *35*). A comprehensive compilation of Phanerozoic marine animal body-size data (*36*) shows that maximum biovolume probably covaried with I/Ca to some extent (Fig. 3B), indicating that O_2 availability in the global upper ocean may have been an important factor in Phanerozoic metazoan evolution. New forms of organisms (e.g., mineralized plankton, larger animals) fundamentally influenced oceanic environments, which in turn affected the evolving biosphere, representing a prime example of the coevolution of life and planet.

REFERENCES AND NOTES

1. A. H. Knoll, *Cold Spring Harbor Perspect. Biol.* **6**, a016121 (2014).
2. R. E. Keeling, A. Körtzinger, N. Gruber, *Ann. Rev. Mar. Sci.* **2**, 199–229 (2010).
3. T. W. Lyons, C. T. Reinhard, N. J. Planavsky, *Nature* **506**, 307–315 (2014).
4. T. W. Dahl *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17911–17915 (2010).
5. E. A. Sperling *et al*., *Nature* **523**, 451–454 (2015).
6. Z. Lu *et al*., *Nat. Commun.* **7**, 11146 (2016).
7. Z. Lu, H. C. Jenkyns, R. E. M. Rickaby, *Geology* **38**, 1107–1110 (2010).
8. G. W. Luther III, T. Campbell, *Deep-Sea Res. A, Oceanogr. Res. Pap.* **38**, S875–S882 (1991).
9. E. L. Rue, G. J. Smith, G. A. Cutter, K. W. Bruland, *Deep Sea Res. Part I Oceanogr. Res. Pap.* **44**, 113–134 (1997).
10. J. Podder *et al*., *Geochim. Cosmochim. Acta* **198**, 218–228 (2017).
11. D. S. Hardisty *et al*., *Earth Planet. Sci. Lett.* **463**, 159–170 (2017).
12. D. S. Hardisty *et al*., *Geology* **42**, 619–622 (2014).
13. X. Zhou *et al*., *Paleoceanography* **30**, 510–526 (2015).
14. X. L. Zhou, E. Thomas, R. E. M. Rickaby, A. M. E. Winguth, Z. L. Lu, *Paleoceanography* **29**, 964–975 (2014).
15. W. Lu *et al*., *Geochem. Perspect. Lett.* **5**, 53–57 (2017).
16. C. T. Edwards, D. A. Fike, M. R. Saltzman, W. Lu, Z. Lu, *Earth Planet. Sci. Lett.* **481**, 125–135 (2018).
17. G. R. Loope, L. R. Kump, M. A. Arthur, *Chem. Geol.* **351**, 195–208 (2013).
18. R. Chance, A. R. Baker, L. Carpenter, T. D. Jickells, *Environ. Sci. Process. Impacts* **16**, 1841–1859 (2014).
19. T. M. Lenton *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9704–9709 (2016).
20. I. J. Glasspool, A. C. Scott, *Nat. Geosci.* **3**, 627–630 (2010).
21. R. A. Berner, *Geochim. Cosmochim. Acta* **70**, 5653–5664 (2006).
22. N. M. Bergman, T. M. Lenton, A. J. Watson, *Am. J. Sci.* **304**, 397–437 (2004).
23. T. M. Lenton, S. J. Daines, B. J. Mills, *Earth Sci. Rev.* **178**, 1–28 (2018).
24. P. G. Falkowski *et al*., *Science* **309**, 2202–2204 (2005).
25. L. R. Kump, *Nature* **335**, 152–154 (1988).
26. T. J. Algeo, R. A. Berner, J. B. Maynard, S. E. Scheckler, *GSA Today* **5**, 64–66 (1995).
27. H. D. Holland, *Philos. Trans. R. Soc. London B Biol. Sci.* **361**, 903–915 (2006).
28. K. M. Meyer, A. Ridgwell, J. L. Payne, *Geobiology* **14**, 207–219 (2016).
29. A. Ridgwell *et al*., *Biogeosciences* **4**, 87–104 (2007).
30. M. E. Katz, Z. V. Finkel, D. Grzebyk, A. H. Knoll, P. G. Falkowski, *Annu. Rev. Ecol. Evol. Syst.* **35**, 523–556 (2004).
31. R. E. Martin, *Global Planet. Change* **11**, 1–23 (1995).
32. P. G. Falkowski *et al*., *Science* **305**, 354–360 (2004).
33. N. J. Butterfield, *Geobiology* **7**, 1–7 (2009).
34. G. A. Logan, J. M. Hayes, G. B. Hieshima, R. E. Summons, *Nature* **376**, 53–56 (1995).
35. J. L. Payne *et al*., *Photosynth. Res.* **107**, 37–57 (2011).
36. N. A. Heim, M. L. Knope, E. K. Schaal, S. C. Wang, J. L. Payne, *Science* **347**, 867–870 (2015).
37. W. Lu *et al*., (Table S2) I/Ca record of phanerozoic carbonates. PANGAEA (2018).

ACKNOWLEDGMENTS

Funding: We acknowledge funding from grants NSF EAR-1349252, OCE-1232620, and OCE-1736542 (Z.L.); NERC NE/J01043X/1, ERC-2013-CoG-617313, and NSF OCE-1736771 (A.R.); NSF OCE-1736538 (E.T.); National Key R&D Project of China 2016YFA0601104 and NSFC 41290260 (G.L.); NSFC 41330102 and 41721002 and the "111" project (Y.S.); Wolfson Research Merit award from the Royal Society and EU award (R.E.M.R.); and NERC NE/N018508/1 and NE/P013643/1 (T.M.L.). **Author contributions:** Z.L. conceived the study. W.L. and Z.L. carried out the data analysis, model simulations, and wrote the paper. A.R. modified cGENIE to incorporate the iodine cycle. All authors contributed to data interpretation and the writing of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All I/Ca data are available in the supplementary materials. Table S2 is additionally provided online at www.pangaea.de/ (*37*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/174/suppl/DC1
Materials and Methods
Figs. S1 to S12
References (*38*–*52*)
Tables S1 and S2

20 November 2017; accepted 17 May 2018
Published online 31 May 2018
10.1126/science.aar5372

EVOLUTIONARY COGNITION

Sensitivity to "sunk costs" in mice, rats, and humans

Brian M. Sweis[1,2], Samantha V. Abram[3], Brandy J. Schmidt[2], Kelsey D. Seeland[2], Angus W. MacDonald III[3], Mark J. Thomas[2,3], A. David Redish[2]*

Sunk costs are irrecoverable investments that should not influence decisions, because decisions should be made on the basis of expected future consequences. Both human and nonhuman animals can show sensitivity to sunk costs, but reports from across species are inconsistent. In a temporal context, a sensitivity to sunk costs arises when an individual resists ending an activity, even if it seems unproductive, because of the time already invested. In two parallel foraging tasks that we designed, we found that mice, rats, and humans show similar sensitivities to sunk costs in their decision-making. Unexpectedly, sensitivity to time invested accrued only after an initial decision had been made. These findings suggest that sensitivity to temporal sunk costs lies in a vulnerability distinct from deliberation processes and that this distinction is present across species.

Traditional economic theory suggests that decisions should be based on valuations of future expectations that ignore spent resources that cannot be recovered [sunk costs (*1*)]. However, extensive evidence shows that humans factor such sunk costs into prospective decisions, even when faced with better alternatives (*2*, *3*). Although early reports claimed that humans are uniquely sensitive to sunk costs (*2*, *3*), it is becoming increasingly clear that nonhuman animals exhibit parallel behaviors (*4*).

Previous nonhuman animal studies that attempted to model the sunk cost phenomenon have yielded conflicting evidence (*4*, *5*). Observational and experimental field studies in swallows, sparrows, mice, and bluegills have found evidence both for and against the sunk cost effect in behaviors relating to parental investment and willingness to care for young (*6*–*10*). Yet in such studies, it has been difficult to disentangle influences of investment history from those of future prospects. Laboratory operant conditioning paradigms in pigeons and rats that control for future expectations when looking at reinforcement learning behaviors have demonstrated that nonhuman animals show increased work ethic or suboptimal perseverative reward-seeking behaviors that escalate with prior investment amount (*11*, *12*). However, these observations often relied on situations of information uncertainty, where subjects overworked in the absence of progress-indicating cues. These observations also often relied on automation or habit-like behaviors (e.g., repetitive lever pressing) driving continued reward pursuit. Such confounding factors in nonhuman animal studies obscure translation to human sunk cost effects, which do not depend on these mechanisms.

Laboratory foraging tasks provide an alternative approach to study decision-making by using naturalistic behaviors that carry both ecological validity and evolutionary significance and are translatable across species (*13*). Foraging tasks rely on optimizing reward-seeking under conditions of limited resources, making them economic tasks.

We designed a foraging task in which subjects spent time from a limited time budget waiting for rewards [Fig. 1, Restaurant Row (*14*) and webSurf (*15*)]. The division of time spent during the task reveals the economic preferences of the subjects. All three species learned to forage in a way that revealed preferences for certain rewards, and all species used reliable subjective valuation strategies to decide between multiple competing reward offers (supplementary text S1). Our neuroeconomic tasks directly test sensitivity to sunk costs across species.

Flavors and genres of rewards (Fig. 1) allowed us to measure subjective preferences as a function of cost, avoiding the confounding possibility that different reward sizes might require different consumption times. Multiple zones allowed us to characterize multiple valuation processes involved in decisions: initial commitment valuations (offer zone), secondary reevaluations (wait zone), and postconsumption hedonic valuations (supplementary text S1 and S2). In this task, two key factors minimized information uncertainty and automated reward-seeking behavior as potential confounding factors: (i) Subjects were provided full information on cost and investment progress (tones counting down or download bar shrinking), and (ii) earning rewards required subjects to wait and withhold quitting after making an initial acceptance decision (rather than requiring a repetitive action).

To address susceptibility to sunk costs, we examined quit decisions in the wait zone. These behaviors involve the abandonment of continued reward pursuit despite having made prior investments (partial waiting) while on a limited budget (time). We parameterized the probability of earning a reward in the wait zone as a function of the remaining time investment required to earn a reward (future costs) and the prior time investment already spent waiting in the wait zone [past (sunk) costs; fig. S3]. The data yielded many samples across all conditions of time remaining and time spent (fig. S4), which allowed us to measure the extent to which irrecoverable prior investments (sunk costs) escalated wait zone commitment (fig. S5).

We found that mice, rats, and humans demonstrated robust sunk cost effects [analysis of variance (ANOVA) collapsing across all sunk cost conditions: mice, $F = 30.75$, $P < 0.0001$; rats, $F = 45.65$, $P < 0.0001$; humans, $F = 3.95$, $P < 0.0001$] (Fig. 2). Importantly, increasing prior investment amounts generated a continuously stronger sunk cost effect (example post-hoc comparison between +1-s and +5-s sunk costs: mice, $F = 45.49$, $P < 0.0001$; rats, $F = 54.41$, $P < 0.0001$; humans, $F = 4.21$, $P < 0.05$) (Fig. 2)—a critical tenet of the sunk cost fallacy (*2*, *4*).

Time spent in the offer zone also detracts from the total time budget, and a similar analysis can be performed (fig. S6). In contrast to our findings for reevaluation processes in the wait zone, we found no effect of time spent in the offer zone. That is, the amount of time spent in the offer zone did not alter the probability of earning rewards once in the wait zone (ANOVA collapsing across all offer conditions: mice, $F = 1.55$, $P = 0.23$; rats, $F = 0.77$, $P = 0.39$; humans: $F = 0.12$, $P = 0.74$) (Fig. 3). Importantly, the delay to reward did not start counting down while the subject remained in the offer zone. This meant that the animal was choosing between distant options and had not yet invested in the offer. This lack of an effect of time spent in the offer zone on progress abandonment once committed suggests that waste avoidance, overall resource depletion, and loss aversion are insufficient explanations of sunk cost–driven escalation of reward-seeking behavior (figs. S7 to S11, supplementary text S2 and S3, and table S1). This also suggests that the offer zone and wait zone may access separable valuation processes and reveals a previously unknown determinant of susceptibility to sunk costs rooted in dissociable decision-making algorithms that are conserved across species.

A sensitivity to sunk costs defies optimality considerations (fig. S12). So, why has this cognitive bias persisted across evolution (supplementary text S4)? Three plausible psychological mechanisms that support sunk cost biases include (i) that it may be more advantageous to calculate reward value through effort expended, (ii) state-dependent valuation learning (SDVL), and (iii) within-trial contrast (WTC) processes (*12*, *16*–*20*). We discuss each of these below.

Because predicting valuations that depend on future outcomes is complex and difficult, animals may have evolved processes in which valuation is measured from effort spent rather than

[1]Graduate Program in Neuroscience and Medical Scientist Training Program, University of Minnesota, Minneapolis, MN 55455, USA. [2]Department of Neuroscience, University of Minnesota, Minneapolis, MN 55455, USA. [3]Department of Psychology, University of Minnesota, Minneapolis, MN 55455, USA.
*Corresponding author. Email: redish@umn.edu

calculated explicitly as an estimate from constructed imaginations of potential future outcomes. Past effort is easy to measure but has a limited (though nonzero) correlation with future value. In contrast, calculating value from expected future outcomes has its own estimation uncertainties. If the correlation between past efforts and future value provides better predictions than the uncertainties of future outcomes, then animals may have evolved processes that use past effort as a proxy to estimate future value (*16*, *19*). This can explain our observation that the postconsumption evaluation increases proportionally to the time spent waiting for the reward in all three species (fig. S13 and supplementary text S5). The fact that susceptibility to sunk costs only accrued in the wait zone implies that valuations in the offer zone depend on different processes that do not include measures of effort spent, but that may be more related to direct estimates of future value.

The SDVL theory hypothesizes that energy spent working toward reward receipt moves the individual into a poorer energy state, enhancing the perceived value of the yet-to-be-obtained reward (*19*, *20*). This continued work can thus escalate commitment of continued reward pursuit with growing sunk costs. Similarly, the WTC theory describes the sunk cost phenomenon as an increasing contrast between the decision-maker's current physical state and the goal (*21*). SDVL and WTC propose that either physiological or psychological states could drive added value, leading to a susceptibility to sunk costs. However, we did not observe sunk costs accruing during the offer zone, even though time spent in the offer zone is equivalent in physical and cognitive demands to time spent in the wait zone. Simple explanations from the WTC and SDVL theories would predict sunk costs to accrue in the offer zone as well.

Past-effort heuristics, SDVL, and WTC can indeed be prominent drivers of the sunk cost effect in our data when sunk cost effects are present. Therefore, our work brings up an intriguing question: How do decision-making processes differ between the wait zone (susceptible to sunk costs) and the offer zone (not susceptible to sunk costs)?

One possibility is that decisions made in the offer zone and wait zone may rely on separate processes that calculate value in distinct ways through dissociable neural circuits (*22*–*24*). Recent findings from other foraging tasks suggest that choosing to remain committed to already accepted options accesses different valuation algorithms than deliberating between distant options (*16*, *25*–*27*). We suggest that wait zone

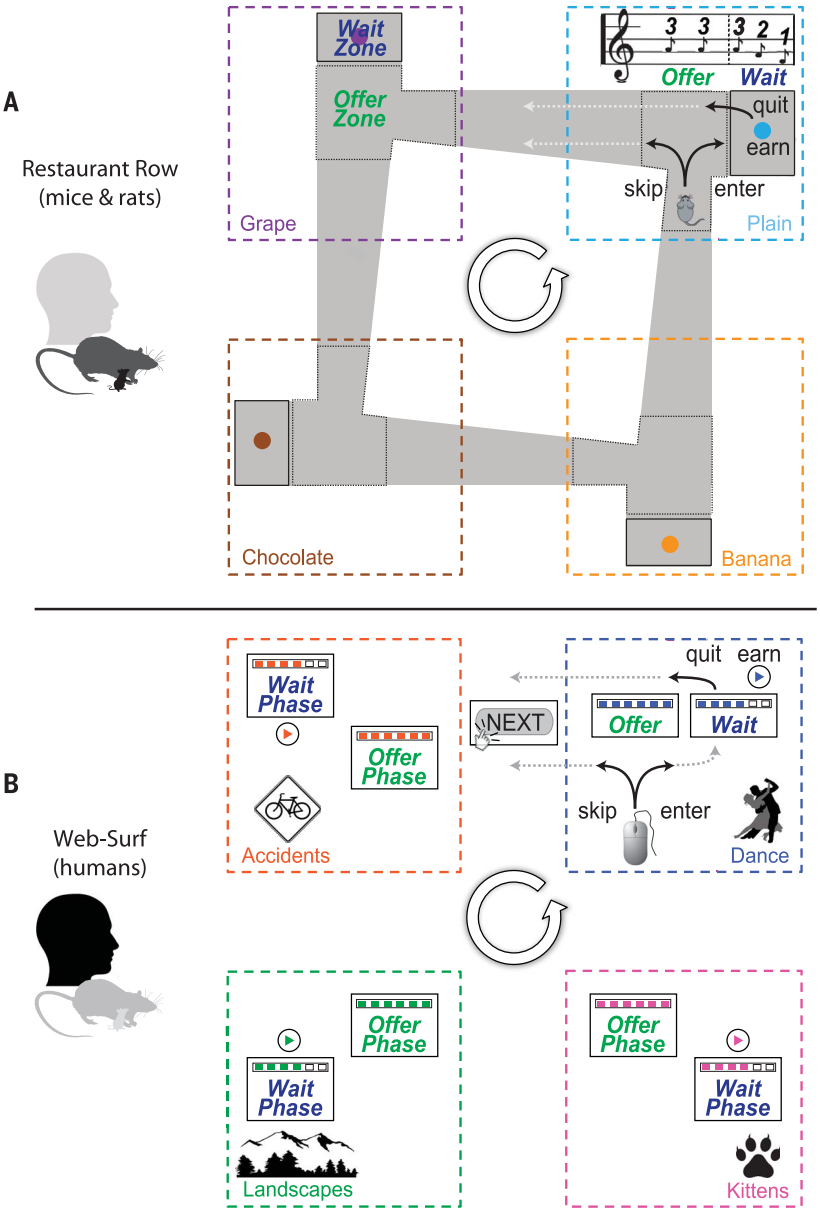


Fig. 1. Task schematics. (**A**) In the Restaurant Row task, food-restricted rodents were trained on a maze in which they encountered serial offers for flavored rewards in four "restaurants." Each restaurant contained a separate offer zone and wait zone. Tones sounded in the offer zone; a fixed tone pitch indicated the delay for which rodents would have to wait in the wait zone (1 to 30 s, random on offer entry). Tone pitch descended in the wait zone during the delay "countdown." Rodents could quit the wait zone for the next restaurant during the countdown, terminating the trial. (**B**) In the web-Surf task, humans performed an analogous 30-min computer-based foraging paradigm in which they encountered serial offers for short entertaining videos from four "galleries." A static "download bar" appeared in the offer phase indicating delay length (1 to 30 s, random on offer entry), which did not begin downloading until after entering the wait phase. Downloads could be quit during the wait phase. Humans were also asked to rate each video on a scale from 1 (least enjoyable) to 4 (most enjoyable) after viewing and to rank the genres at the end of the session.

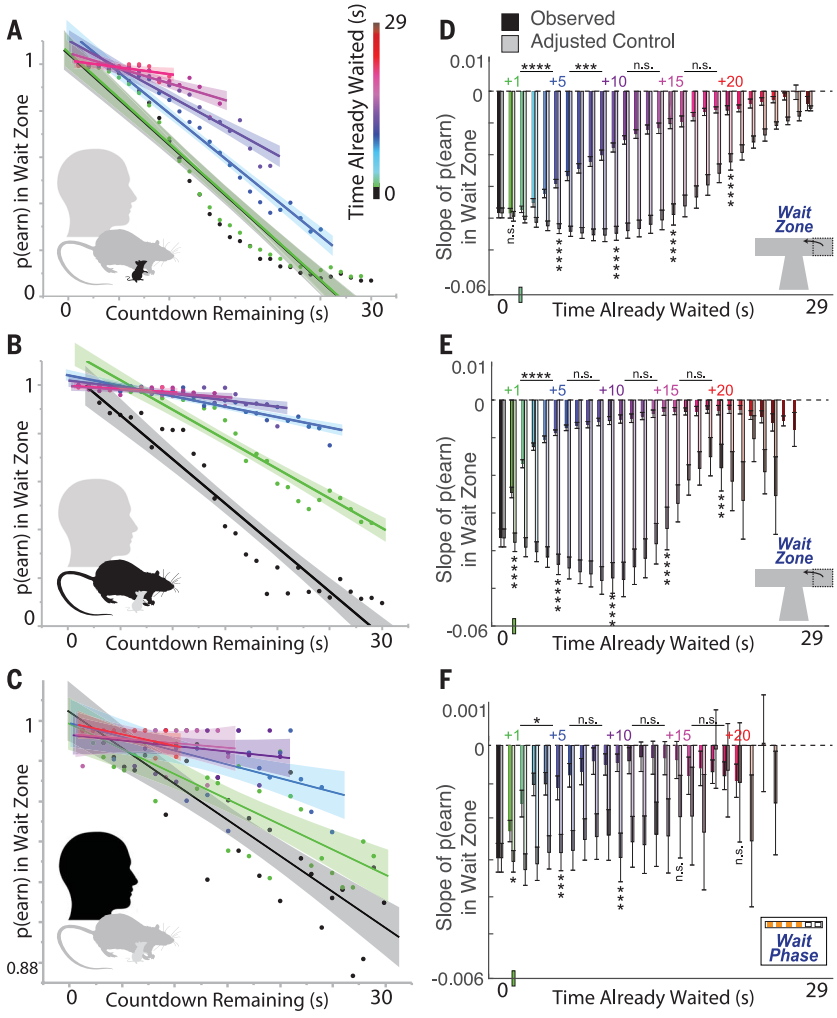


Fig. 2. The amount of time spent waiting increases commitment to continuing reward pursuit in mice, rats, and humans. (**A** to **C**) Probability of earning a reward in the wait zone as a function of countdown time remaining in (A) mice, (B) rats, and (C) humans. Black data points indicate trials in which subjects had just entered the wait zone. Colored data points indicate time remaining in the countdown after subjects had already waited varying times (fig. S3). Linear regressions are plotted with 95% confidence interval shadings. (**D** to **F**) Slopes calculated from each linear regression in (A) to (C) ("observed") and slopes recalculated iteratively from black data points to match colored data ranges in (A) to (C) ("adjusted controls") (fig. S5) are plotted ±1 SEM. The colored tick on the *x* axis indicates time in the wait zone until the first significant sunk cost effect was observed. ANOVAs were used to compare slopes of linear regression models, testing for interactions with sunk cost conditions and controls, correcting for multiple comparisons. Not significant (n.s.), $P > 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

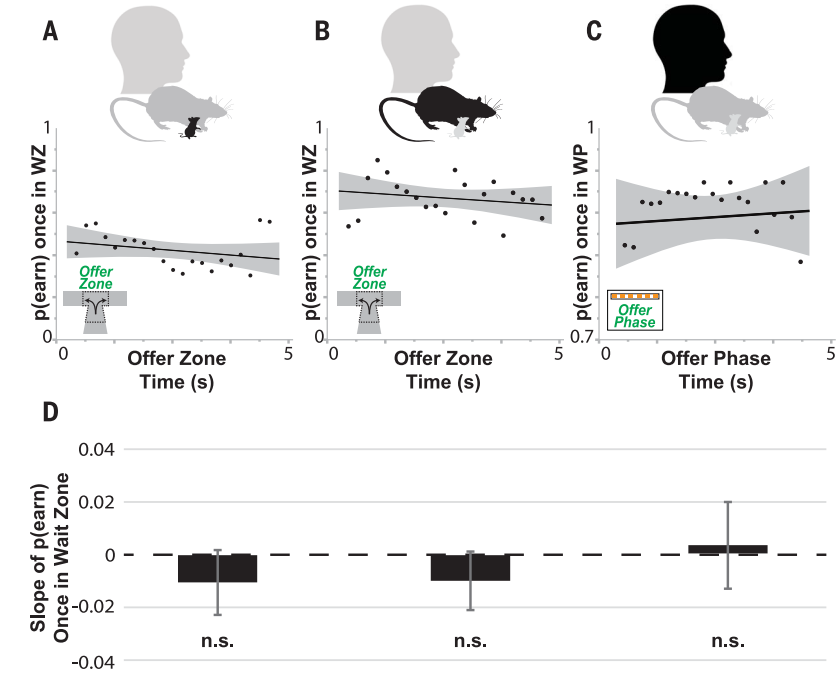


Fig. 3. Resources spent while deliberating do not contribute to the sunk cost effect. (**A** to **C**) Amount of time spent in the offer zone choosing to skip versus to enter did not influence the probability of earning versus quitting once in the wait zone after subjects chose to enter (fig. S6). Linear regressions are plotted with 95% confidence interval shadings. (**D**) Slopes calculated from linear regressions are plotted ±1 SEM and are not significantly different from each other or zero in mice ($F = 1.545$, $P = 0.229$), rats ($F = 0.767$, $P = 0.392$), or humans ($F = 0.117$, $P = 0.737$), on the basis of an ANOVA with post-hoc comparisons against zero. n.s., $P > 0.05$.

decisions are driven by distinct mechanisms that depend on recently accumulated states, whereas offer zone decisions are driven by deliberation mechanisms that simulate future outcomes constructed from a more extended knowledge base of past experiences. There is strong neural evidence across these species to suggest that competing simulated future alternatives are being represented, evaluated, and compared in deliberation algorithms, whereas other decision-making systems depend on more immediate sensory signals that likely include interoceptive signals of effort expended and representations of internal state (*23*–*24*, *28*–*31*) (fig. S8 and supplementary text S3). Each of these decision-making systems provides computational advantages better suited for different situations. Thus, these multiple valuation algorithms can each confer independent evolutionary advantages and can coexist and persist across time and species (*22*–*24*).

The sunk cost fallacy, by definition, arises from valuing spent resources that cannot be recovered. Our data finds that these sunk costs only accrue under specific situations in mice, rats, and humans. We suggest that multiple, parallel decision-making valuation algorithms implemented in dissociable neural circuits have persisted across species and over time through evolution. Our data imply that these different valuation algorithms are differentially susceptible to sunk costs. Past studies that reported conflicting findings across species may have failed to consider how different decision systems drive behavior (fig. S14 and supplementary text S6 and S7). Studies identifying differences in sensitivity to sunk costs should consider which decision-making processes are being accessed by the individual in a given task. Because these processes could change between species, or within species but across aging, stages of development, or circumstances, so too could sensitivity to sunk costs.

Using a translational approach in mice, rats, and humans, we find direct evidence in parallel tasks that the sunk cost phenomenon is conserved across species. Our findings highlight the utility of economic paradigms that can dissociate decision-making computations, using naturalistic tasks that are translatable across various species and that can be expanded to survey individuals of varying ages or psychiatric populations. These tasks and findings may aid future research in education or neuropsychiatry by shedding light on diagnostic or intervention strategies and revealing the roles of neurally distinct decision systems.

REFERENCES AND NOTES

1. R. H. Thaler, *J. Behav. Decis. Making* **12**, 183–206 (1999).
2. H. Arkes, P. Ayton, *Psychol. Bull.* **125**, 591–600 (1999).
3. F. Höffler, *Why Humans Care About Sunk Costs While Animals Don't. An Evolutionary Explanation* (Max Planck Institute for Research on Collective Goods, 2005).
4. P. Magalhães, K. G. White, *J. Exp. Anal. Behav.* **105**, 339–361 (2016).
5. J. Arantes, R. C. Grace, *Learn. Behav.* **36**, 1–11 (2008).
6. R. Coleman, M. Gross, R. Sargent, *Behav. Ecol. Sociobiol.* **18**, 59–66 (1985).
7. D. Maestripieri, E. Alleva, *Behav. Processes* **23**, 223–230 (1991).
8. D. Winkler, *Behav. Ecol.* **2**, 133–142 (1991).
9. R. Dawkins, T. R. Carlisle, *Nature* **262**, 131–133 (1976).
10. P. Weatherhead, *Behav. Ecol. Sociobiol.* **5**, 373–381 (1979).
11. P. Magalhães, K. G. White, T. Stewart, E. Beeby, W. van der Vliet, *Learn. Behav.* **40**, 195–206 (2012).
12. K. F. Pattison, T. R. Zentall, S. Watanabe, *J. Comp. Psychol.* **126**, 1–9 (2012).
13. T. Kalenscher, M. van Wingerden, *Front. Neurosci.* **5**, 82 (2011).
14. A. P. Steiner, A. D. Redish, *Nat. Neurosci.* **17**, 995–1002 (2014).
15. S. V. Abram, Y. A. Breton, B. Schmidt, A. D. Redish, A. W. MacDonald III, *Cogn. Affect. Behav. Neurosci.* **16**, 37–50 (2016).
16. A. M. Wikenheiser, D. W. Stephens, A. D. Redish, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8308–8313 (2013).
17. J. Gibbon, R. M. Church, "Sources of variance in an information processing theory of timing," in *Animal Cognition*, H. L. Roitblat, T. G. Bever, H. S. Terrace, Eds. (Erlbaum, 1984), pp. 465–488.
18. A. Kacelnik, B. Marsh, *Anim. Behav.* **63**, 245–250 (2002).
19. L. Pompilio, A. Kacelnik, S. T. Behmer, *Science* **311**, 1613–1615 (2006).
20. J. M. Aw, M. Vasconcelos, A. Kacelnik, *Anim. Behav.* **81**, 1117–1128 (2011).
21. R. A. Singer, T. R. Zentall, *Learn. Motiv.* **42**, 255–271 (2011).
22. B. J. Casey, R. M. Jones, T. A. Hare, *Ann. N. Y. Acad. Sci.* **1124**, 111–126 (2008).
23. M. van der Meer, Z. Kurth-Nelson, A. D. Redish, *Neuroscientist* **18**, 342–359 (2012).
24. A. D. Redish, *The Mind within the Brain: How We Make Decisions and How Those Decisions Go Wrong* (Oxford Univ. Press, 2013).
25. A. D. Redish, N. W. Schultheiss, E. C. Carter, *Curr. Top. Behav. Neurosci.* **27**, 313–333 (2015).
26. E. C. Carter, A. D. Redish, *J. Exp. Psychol. Gen.* **145**, 1093–1101 (2016).
27. D. Stephens, J. Krebs, *Foraging Theory* (Princeton Univ. Press, 1987).
28. A. M. Wikenheiser, A. D. Redish, *Nat. Neurosci.* **18**, 289–294 (2015).
29. A. E. Papale, M. C. Zielinski, L. M. Frank, S. P. Jadhav, A. D. Redish, *Neuron* **92**, 975–982 (2016).
30. B. Bushong, L. M. King, C. F. Camerer, A. Rangel, *Am. Econ. Rev.* **100**, 1556–1571 (2010).
31. P. Dayan, Y. Niv, B. Seymour, N. D. Daw, *Neural Netw.* **19**, 1153–1160 (2006).

ACKNOWLEDGMENTS

We thank the Redish, Thomas, and MacDonald laboratories for assistance. **Funding:** R01 DA019666 (M.J.T.), R01 DA030672 (A.D.R.), R01 MH080318 (A.D.R.), R01 DA041808 (M.J.T.), R01 MH084861 (A.W.M.), MnDRIVE Neuromodulation Research Fellowships (B.M.S.), the Breyer-Longden Family Research Foundation (M.J.T.), MSTP NIGMS 5T32GM008244-25 (B.M.S.), GPN NIGMS 5T32GM008471-22 (B.M.S.), K02 DA035459 (M.J.T.), F30 DA043326 NRSA (B.M.S.), F31 DA040335 NRSA (S.V.A.), and F32 DA038392 NRSA (B.J.S.). **Author contributions:** B.M.S., S.V.A., and B.J.S. and K.D.S. performed the experiments in mice, humans, and rats, respectively. B.M.S. analyzed the data and wrote the manuscript. A.D.R., M.J.T., and A.W.M. supervised the project and cowrote the manuscript. **Competing interests:** The authors have no conflicting interests. **Data and materials availability:** Data are included in supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/178/suppl/DC1
Materials and Methods
Supplementary Text S1 to S7
Figs. S1 to S14
Table S1
References (*32*–*75*)
Data S1

26 December 2017; accepted 29 May 2018
10.1126/science.aar8644

PLANT SCIENCE

Ethylene-gibberellin signaling underlies adaptation of rice to periodic flooding

Takeshi Kuroha[1]*†, Keisuke Nagai[2]*, Rico Gamuyao[2]*‡, Diane R. Wang[3]*§, Tomoyuki Furuta[2], Masanari Nakamori[2], Takuya Kitaoka[2], Keita Adachi[2], Anzu Minami[2], Yoshinao Mori[2], Kiyoshi Mashiguchi[1], Yoshiya Seto[1]||, Shinjiro Yamaguchi[1], Mikiko Kojima[4], Hitoshi Sakakibara[4,5], Jianzhong Wu[6], Kaworu Ebana[7], Nobutaka Mitsuda[8], Masaru Ohme-Takagi[8,9], Shuichi Yanagisawa[10], Masanori Yamasaki[11], Ryusuke Yokoyama[1], Kazuhiko Nishitani[1], Toshihiro Mochizuki[12], Gen Tamiya[13,14], Susan R. McCouch[3]†, Motoyuki Ashikari[2]†

Most plants do poorly when flooded. Certain rice varieties, known as deepwater rice, survive periodic flooding and consequent oxygen deficiency by activating internode growth of stems to keep above the water. Here, we identify the gibberellin biosynthesis gene, *SD1* (*SEMIDWARF1*), whose loss-of-function allele catapulted the rice Green Revolution, as being responsible for submergence-induced internode elongation. When submerged, plants carrying the deepwater rice–specific *SD1* haplotype amplify a signaling relay in which the *SD1* gene is transcriptionally activated by an ethylene-responsive transcription factor, OsEIL1a. The SD1 protein directs increased synthesis of gibberellins, largely GA_4, which promote internode elongation. Evolutionary analysis shows that the deepwater rice–specific haplotype was derived from standing variation in wild rice and selected for deepwater rice cultivation in Bangladesh.

Deepwater rice varieties, grown mainly in Asian lowland areas, respond to months-long deep flooding by increasing plant height as water levels rise. The deepwater growth response is due to rapid internode elongation, allowing leaves to remain above the water surface; internode elongation is accomplished by both cell proliferation and elongation (*1*). Only *SNORKEL1* and *SNORKEL2* (*SK1/2*), encoding transcription factors with high similarity to the flash flood–tolerant regulator SUB1A (*2*), have previously been identified as causal genes for the deepwater response (*3*). Although ethylene and gibberellins are thought to trigger this response (*1*, *4*), the molecular mechanism(s) underlying the interaction among these hormones during internode elongation remains elusive.

To identify factors regulating the deepwater response in rice, we carried out a genome-wide association study (GWAS) with a diversity panel of Asian rice and deepwater rice varieties (table S1). We measured nine deepwater traits (Fig. 1A, fig. S1, and table S2) and selected total internode length measured 7 days after submergence at the 10-leaf stage (TILS) as a representative trait (Fig. 1B and figs. S2 and S3). The GWAS with the TILS value as a proxy for the deepwater response (figs. S4 to S6) revealed six quantitative trait loci (QTLs) exceeding the significance threshold (Fig. 1C and table S3). The GWAS peak on chromosome 1 was located within a QTL, $qTIL1^{C9285}$, which we had detected in a previous analysis of the deepwater response (*5*) (fig. S7). To identify the gene(s) responsible for $qTIL1^{C9285}$, we conducted high-resolution linkage analysis by using a mapping population from a cross between near-isogenic lines NIL-12 and NIL-1+12, both containing chromosomal introgressions from deepwater rice (C9285 variety) in a nondeepwater rice (T65 variety) background (figs. S8 to S10). The candidate region was delimited to a 5.5-kb stretch that included the *OsGA20ox2* (*Oryza sativa gibberellin 20–oxidase 2*) gene. Also known as *SD1* (*SEMIDWARF1*), this gene encodes a gibberellin biosynthesis enzyme; its null allele produces a semidwarf phenotype and was selected during the rice Green Revolution (*6*) (Fig. 1D and fig. S11). Introgression of the C9285 *SD1* allele ($SD1^{C9285}$) into NIL-12 increased total internode length in response to submergence (Fig. 1E), and near-constitutive expression of $SD1^{C9285}$ increased internode elongation even without submergence (fig. S12). Furthermore, NIL-12 with a null allele of *SD1* showed shorter total internode length in response to submergence than either NIL-12 or NIL-1+12 (fig. S13). These results suggest that $SD1^{C9285}$ is responsible for $qTIL1^{C9285}$, which contributes to the promotion of internode elongation in response to submergence.

Next, we compared polymorphisms in the GWAS panel across the $qTIL1^{C9285}$ candidate region and identified six *SD1* haplogroups (fig. S14 and table S4). The haplogroup belonging to C9285 (Hap-6) had the highest TILS value in the presence of *SK1/2* (Fig. 2A and fig. S15). Hereafter, we refer to the $SD1^{C9285}$ haplotype, including the 17 specific polymorphisms in the promoter and second intron, as the deepwater rice–specific haplotype (DWH) (fig. S14). In response to submergence, *SD1* transcript accumulation in C9285 was higher than that in T65 (Fig. 2B) and diminished from 6 hours until 12 hours after submergence, consistent with negative feedback of gibberellin suppressing *SD1* transcription (*7*) (fig. S16). Accumulation of *SD1* transcripts was observed in the elongating internode (Fig. 2C). NIL-1 also showed induced *SD1* transcript accumulation (figs. S8A and S17); we found a significant association between the level of *SD1* transcript accumulation and internode elongation in recombinant lines from the linkage analysis (fig. S18). Furthermore, varieties harboring the DWH and Hap-5 showed higher *SD1* transcript accumulation than varieties containing other *SD1* haplotypes, regardless of the presence of *SK1/2* (fig. S19). Our results suggest that the DWH potentiates *SD1* transcript accumulation in response to submergence, independently of *SK1/2*, and that this induction works in synergy with *SK1/2* to enhance internode elongation in response to submergence.

We next tested *SD1* transcription in response to ethylene (*1*, *4*). Ethylene application showed higher *SD1* transcript accumulation in C9285 and NIL-1 than in T65 (Fig. 2D and fig. S20) even in the presence of a protein synthesis inhibitor (fig. S21), suggesting that ethylene-inducible *SD1* transcription is mediated by the DWH and that *SD1* is regulated by the ethylene signaling pathway without de novo protein synthesis. The ethylene response involves stabilization and accumulation of the transcription factor EIN3 (ETHYLENE INSENSITIVE 3) family to activate the transcription of ethylene-responsive genes (*8*), so we investigated the transcriptional regulation of *SD1* by an EIN3 homolog, OsEIL1a protein (fig. S22). Near-constitutive expression of OsEIL1a fused with a repression domain (*9*) resulted in suppression of *SD1* induction in response to ethylene treatment (Fig. 2E and fig. S23). A transactivation assay demonstrated the direct activation of the $SD1^{C9285}$ promoter by OsEIL1a, whereas SK1/2 was not able to transactivate the $SD1^{C9285}$ promoter (fig. S24). Specific binding of OsEIL1a to the $SD1^{C9285}$ promoter is supported by assays using a chemical induction

[1]Graduate School of Life Sciences, Tohoku University, Miyagi 890-8577, Japan. [2]Bioscience and Biotechnology Center, Nagoya University, Aichi 464-8601, Japan. [3]Plant Breeding and Genetics, Cornell University, Ithaca, NY 14853, USA. [4]RIKEN Center for Sustainable Resource Science, Kanagawa 230-0045, Japan. [5]Graduate School of Bioagricultural Sciences, Nagoya University, Aichi 464-8601, Japan. [6]Institute of Crop Science, NARO, Ibaraki 305-8518, Japan. [7]Genetic Resources Center, NARO, Ibaraki 305-8518, Japan. [8]Bioproduction Research Institute, AIST, Ibaraki 305-8566, Japan. [9]Graduate School of Science and Engineering, Saitama University, Saitama 338-8570, Japan. [10]Biotechnology Research Center, The University of Tokyo, Tokyo 113-8657, Japan. [11]Graduate School of Agricultural Science, Kobe University, Hyogo 675-2103, Japan. [12]Faculty of Agriculture, Kyushu University, Fukuoka 812-8581, Japan. [13]Tohoku Medical Megabank Organization, Tohoku University, Miyagi 980-8575, Japan. [14]RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan.

*These authors contributed equally to this work.

†Corresponding author. Email: tkuroha@m.tohoku.ac.jp (T.K.); srm4@cornell.edu (S.R.M.); ashi@agr.nagoya-u.ac.jp (M.A.)

‡Present address: Department of Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA. §Present address: Department of Geography, University at Buffalo, Buffalo, NY 14261, USA. ||Present address: Department of Agricultural Chemistry, School of Agriculture, Meiji University, 1-1-1, Higashimita, Tama-ku, Kawasaki, Kanagawa 214-8571, Japan.

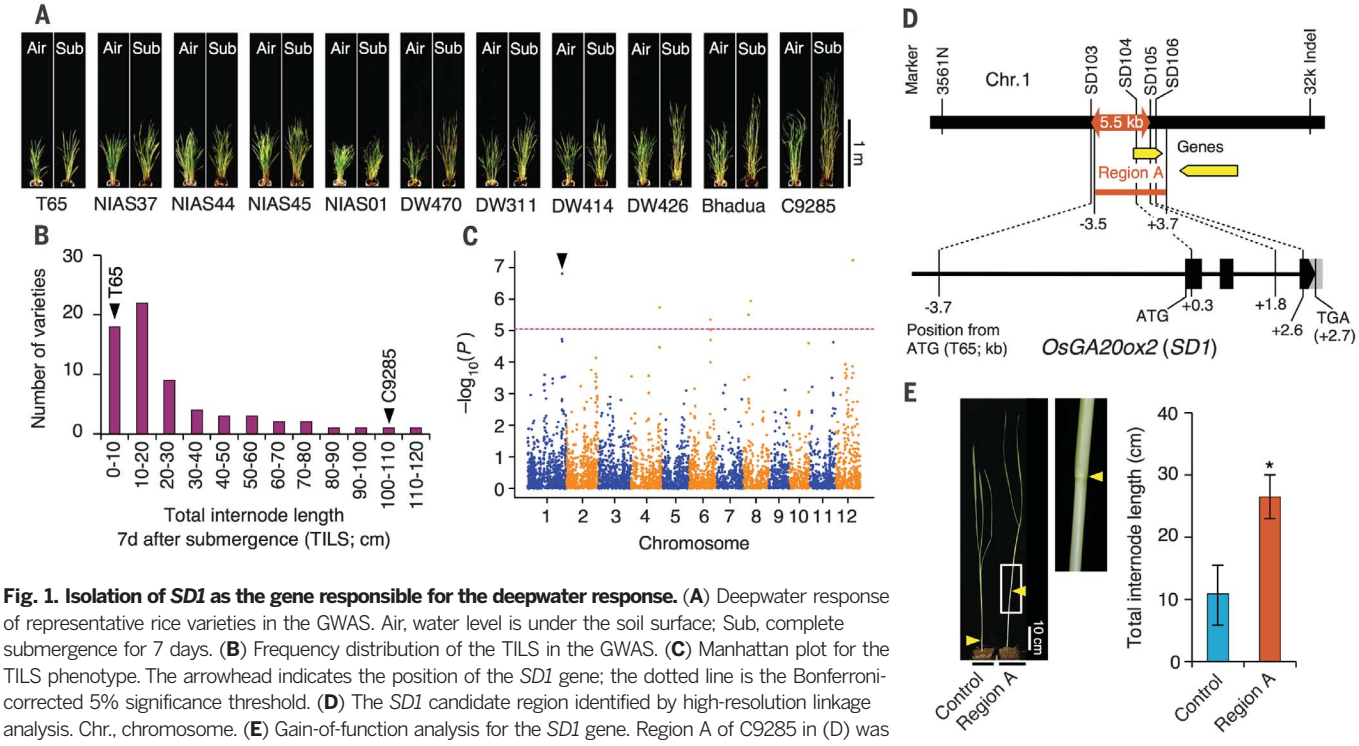


Fig. 1. Isolation of *SD1* as the gene responsible for the deepwater response. (**A**) Deepwater response of representative rice varieties in the GWAS. Air, water level is under the soil surface; Sub, complete submergence for 7 days. (**B**) Frequency distribution of the TILS in the GWAS. (**C**) Manhattan plot for the TILS phenotype. The arrowhead indicates the position of the *SD1* gene; the dotted line is the Bonferroni-corrected 5% significance threshold. (**D**) The *SD1* candidate region identified by high-resolution linkage analysis. Chr., chromosome. (**E**) Gain-of-function analysis for the *SD1* gene. Region A of C9285 in (D) was transformed into NIL-12 (mean ± SD, $n \geq 3$ replicates). *$P < 0.05$ (Student's *t* test). Arrowheads indicate positions of nodes. The boxed area is shown to the right at greater magnification.

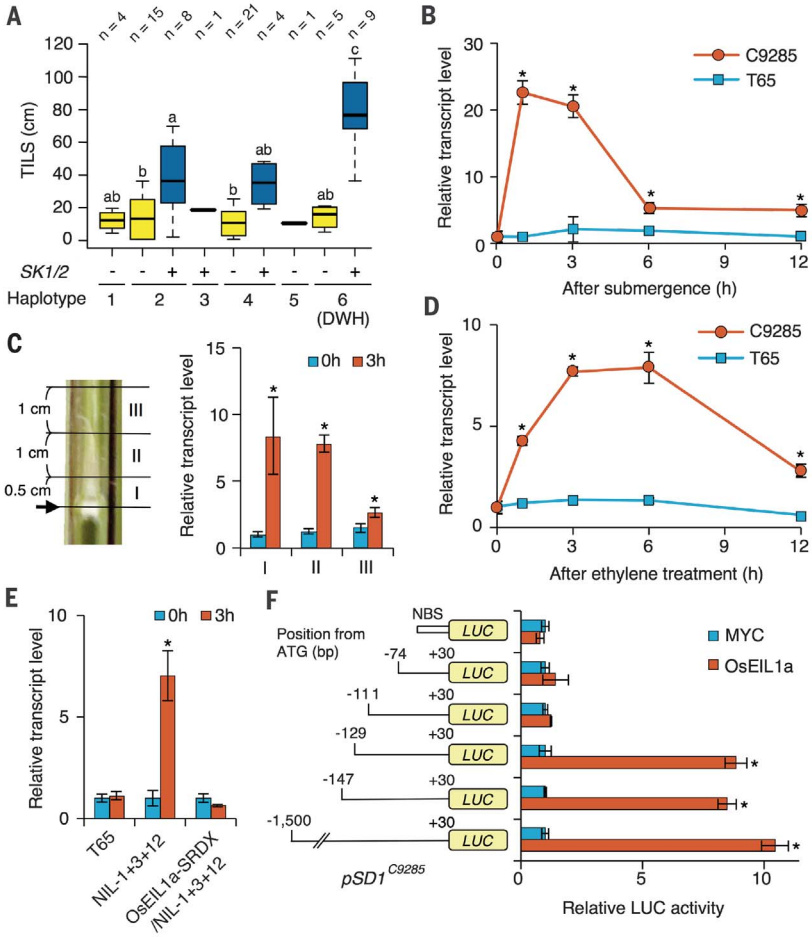


Fig. 2. Contribution of *SD1* to internode elongation in response to submergence via ethylene signaling. (**A**) Box plots of TILS values for six *SD1* haplotypes and absence (−) or presence (+) of *SK1/2*. Box edges represent the 0.25 quantile and 0.75 quantile, with median values shown by bold lines; whiskers indicate 1.5 times the interquartile range. Different letters denote significant differences according to the Tukey-Kramer test ($P < 0.05$). (**B**) Quantitative PCR (qPCR) analysis of *SD1* transcription under submergence treatment. (**C**) qPCR analysis of submergence-induced *SD1* transcription in different tissue regions (I to III) above the top node (arrow). (**D**) qPCR analysis of *SD1* transcription under ethylene treatment. (**E**) qPCR analysis determining the effect of *OsEIL1a-SRDX* near-constitutive expression on the induction of *SD1* transcription by ethylene. Values in (B) to (E) are the mean ± SD ($n = 3$). *$P < 0.05$ (Student's *t* test) versus 0 hours. (**F**) Transactivation of the *SD1* promoter by OsEIL1a in rice protoplasts (mean ± SD, $n = 3$). *$P < 0.05$ (Student's *t* test) versus the MYC control. NBS, 35S minimal promoter; LUC, luciferase.

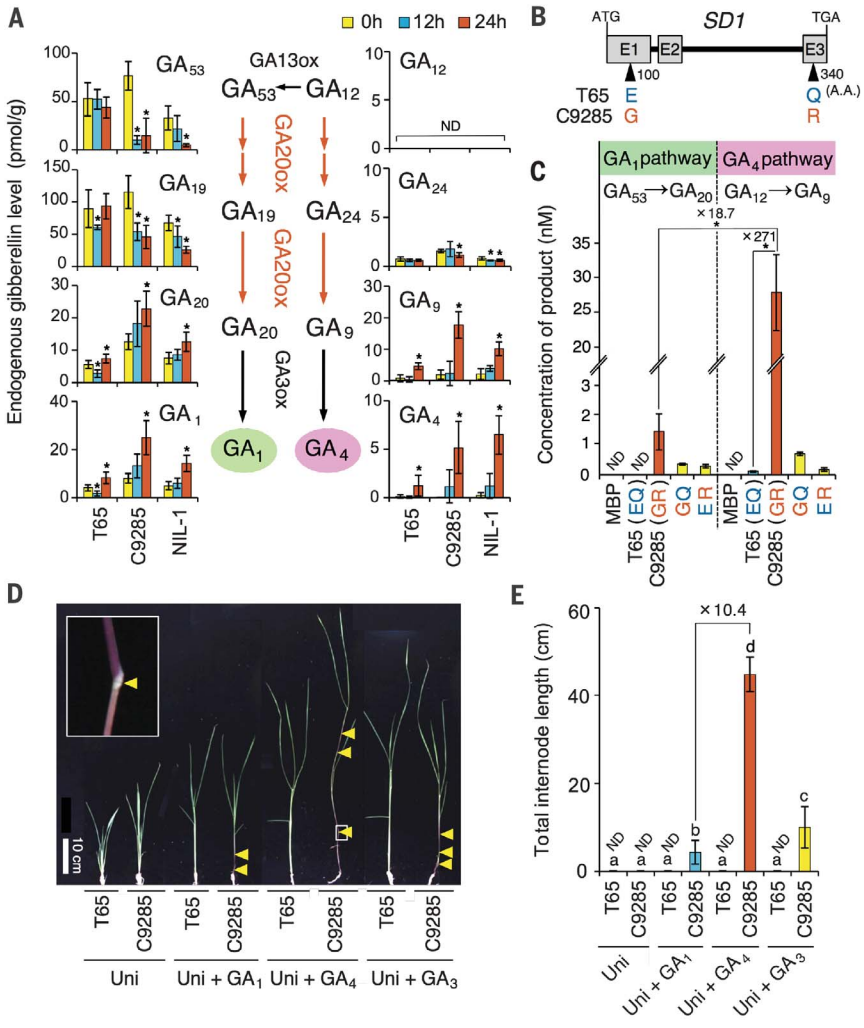


Fig. 3. Contribution of SD1-derived GA_4 for internode elongation. (**A**) Gibberellin levels in elongating internodes under submergence treatment (mean ± SD, $n \geq 3$). ox, -oxidase; ND, not detected. *$P < 0.05$ (Student's t test) versus 0 hours. (**B**) Exonic SNPs of *SD1* between T65 and C9285. E1 to E3, exons 1 to 3; E, Glu; Q, Gln; G, Gly; R, Arg; A.A., amino acid. (**C**) Enzymatic activities of recombinant SD1 proteins for 10-min reactions against GA_{53} or GA_{12} (mean ± SD, $n = 3$). MBP, maltose-binding protein. *$P < 0.05$ (Student's t test). (**D** and **E**) Effects of different active gibberellin species (10^{-5} M) on internode elongation (means ± SD, $n \geq 9$). Uni, 10^{-6} M uniconazole; arrowheads, positions of nodes. Different letters denote significant differences according to the Tukey-Kramer test ($P < 0.05$).

system, the OsEIL1a effector fused with an activation domain, and a mutant form of OsEIL1a (figs. S25 and S26). Promoter deletion, scanning mutagenesis, and synthetic promoter analyses revealed that a 13–base pair (bp) sequence of the $SD1^{C9285}$ promoter region (positions −123 to −111 from the start codon) is required for OsEIL1a transactivation (Fig. 2F and fig. S27, A and B). We confirmed direct binding of the OsEIL1 protein to the $SD1^{C9285}$ promoter region by an in vitro protein-DNA interaction assay (fig. S28). $SD1^{T65}$ and $SD1^{C9285}$ promoters both harbor the 13-bp sequence (table S4) and were each activated by OsEIL1a (fig. S27C), suggesting that the regulatory element(s) for ethylene-induced transcription is conserved in $SD1^{T65}$ and $SD1^{C9285}$ promoters. As *SD1* transcription in T65 was not induced by ethylene treatment (Fig. 2D and fig. S20), these observations imply the existence of an unknown suppressor(s) for the ethylene-inducible transactivation. It has been reported that *GA20-oxidase* transcription is under negative feedback regulation by a transcription factor(s) or transcriptional co-regulators (*10*). In deepwater rice harboring the DWH, this negative feedback may be partially diminished, leading to higher *SD1* transcript accumulation relative to that in other rice in response to submergence.

SD1 encodes a GA20-oxidase for gibberellin biosynthesis (*11*) (Fig. 3A). GA_{20} and GA_9, the products of GA20-oxidases in the early-13-hydroxylation pathway (GA_1 pathway) and non-13-hydroxylation pathway (GA_4 pathway), are converted into bioactive gibberellin species GA_1 and GA_4, respectively (*11*). In rice vegetative tissues, endogenous GA_1 is predominant whereas GA_4 is minor (*12*). Contrastingly, in response to submergence, levels of both GA_1 and GA_4 increased in C9285 but not in T65 (*3*). To test whether this increase is due to the $SD1^{C9285}$ harbored by the DWH, internode gibberellin levels were quantified in T65, C9285, and NIL-1 under submergence treatment (Fig. 3A). Levels of GA_{20}, GA_9, GA_1, and GA_4 increased in a time-dependent fashion in both C9285 and NIL-1, whereas these trends were not observed in T65, suggesting that the DWH contributes to the increase in bioactive gibberellin species GA_1 and GA_4 after submergence.

The *SD1* coding region contains two nonsynonymous single-nucleotide polymorphisms (SNPs) in the first and third exons among rice varieties (*13*) (Fig. 3B and fig. S14). We compared enzymatic activities of recombinant $SD1^{T65}$ and $SD1^{C9285}$ proteins in the GA_1 and GA_4 pathways. $SD1^{C9285}$ showed higher enzymatic activity than $SD1^{T65}$ in both the GA_1 (*13*) and GA_4 pathways (Fig. 3C, fig. S29, and table S5). In the GA_4 pathway, $SD1^{C9285}$ displayed enzymatic activity ~270 times as high as that of $SD1^{T65}$ and ~19 times as high as that in the GA_1 pathway. Moreover, near-constitutive expression of $SD1^{C9285}$ in planta increased GA_9 and GA_4 levels (fig. S30). Further phenotypic analyses revealed that the exonic SNPs and the DWH together contribute to internode elongation in response to submergence (fig. S31). Taken together, our results indicate that the main activity of the $SD1^{C9285}$ protein is the conversion of GA_{12} into GA_9 in the GA_4 pathway rather than of GA_{53} into GA_{20} in the GA_1 pathway and that $SD1^{C9285}$ contributes to the increase of the GA_4 level in response to submergence. The reason GA_1, in addition to GA_4, increases in response to submergence in C9285 can be explained by the existence of GA13-oxidase activity in rice internodes (*14*), involving a metabolic flow from GA_{12} to GA_{53} (Fig. 3A). We further compared physiological activities of gibberellin species for elongation of internodes in T65 and C9285 (Fig. 3, D and E). C9285 showed internode

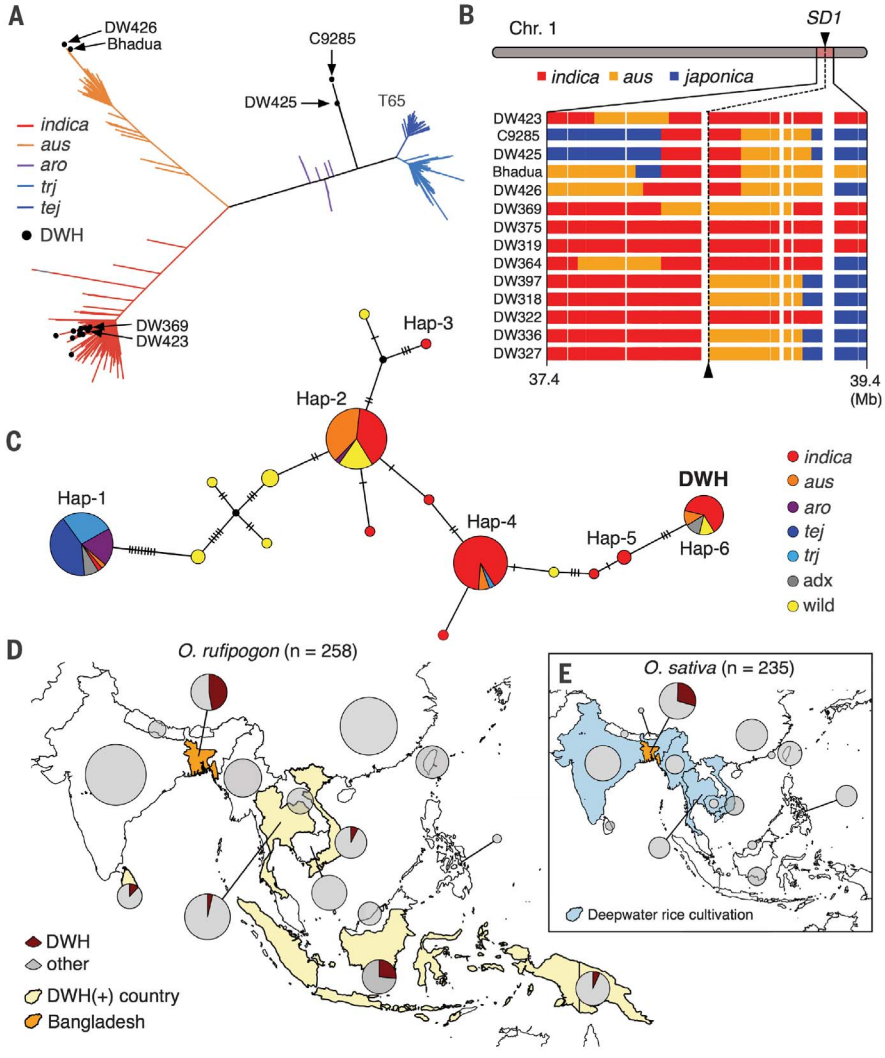


Fig. 4. The DWH was derived from standing variation in *O. rufipogon* and selected in *O. sativa*. (**A**) Distance tree of diverse *O. sativa* varieties in the GWAS panel. Black dots, DWH(+) varieties. Arrows, varieties with accentuated deepwater response. (**B**) Admixture analysis of the 2-Mb region flanking *SD1* in DWH(+) varieties. (**C**) Haplotype network of the *SD1* gene. Each haplotype is separated by mutational changes, with hatches indicating mutational differences between linked haplotypes. *aro*, *aromatic*; *tej*, *temperate japonica*; *trj*, *tropical japonica*; adx, admix; wild, *O. rufipogon*. (**D** and **E**) Geographical distribution of the DWH in *O. rufipogon* and *O. sativa*. The pie chart size is proportional to the number of accessions.

elongation in response to gibberellins, whereas T65 did not (*15*) (Fig. 3, D and E). GA_4 promoted internode elongation 10 times as much as GA_1 did, resulting in the highest increase of total plant height among the gibberellin species (Fig. 3, D and E, and fig. S32). The higher bioactivity of GA_4 than of GA_1, due to differences in binding affinity for the rice gibberellin receptor (*16*), may amplify the effect of gibberellins on cell proliferation and the elongation of the internodes in deepwater rice (*1*). Taken together, our results suggest that the physiological activity of GA_4 impacts internode elongation and that the increase in GA_4 levels under submergence induces rapid internode elongation in deepwater rice.

We next investigated the genetic relationships among deepwater rices. Haplotype analysis of 149 *O. sativa* varieties (*17*) revealed that the DWH is present only in Bangladeshi deepwater rice (tables S6A and S7A). Principal components analysis showed that DWH-positive [DWH(+)] varieties with the most accentuated deepwater responses (table S6A) belonged to *indica*, *aus*, and admixed subpopulations (Fig. 4A and figs. S33 to S35). Other DWH(+) varieties all clustered within *indica* (figs. S33B and S34B); nucleotide diversity analysis suggested that *indica* DWH(+) varieties underwent a population bottleneck via selection for cultivation in Bangladesh (fig. S36). Admixture analysis revealed that the DWH originated in *indica* or *aus* (Fig. 4B). We further investigated DWH presence in *O. sativa*'s wild ancestor, *Oryza rufipogon*. Gene haplotype network analysis showed that wild and cultivated rice shared the DWH (Fig. 4C and tables S6B and S7B). We analyzed wild rice resequencing (fig. S37 and tables S6B and S7B) and indel (table S8) data revealing 16 DWH(+) wild accessions (table S9A). W1 and W4 subpopulations in wild rice have been reported to be likely ancestral populations for *indica* and *aus*, respectively (*18*). All but one DWH(+) wild accession were classified as W1, W4, or W1 admixed (table S9A). Of nine known chloroplast haplotypes (*18*), cpGroup-III had the highest DWH frequency (table S9B). These cpGroup-III DWH(+) accessions originated from Bangladesh (fig. S38A and table S8), suggesting that the DWH emerged in cpGroup-III Bangladeshi wild rice. DWH presence was more broadly distributed in wild rice than in cultivated rice (Fig. 4, D and E); nonetheless, occurrence was more frequent in Bangladesh (table S9C). Comparison of nucleotide diversity of the *SD1* region in DWH(+) accessions versus DWH(−) accessions from the *O. rufipogon* panel (fig. S38A) resulted in a ratio of ~1 (fig. S38B), providing evidence that the DWH evolved first in *O. rufipogon* populations. Three W4 Bangladeshi wild accessions clustering in the same clade (fig. S38A, red circle) shared all 48 polymorphisms in *SD1* with C9285, suggesting that this group may be direct ancestors of Bangladeshi DWH(+) cultivated deepwater rice (table S10). The absence of DWH(+)-diagnostic indels in other wild rice species indicated that the DWH emerged after the speciation of *O. rufipogon* (table S11). Taken together, our results support the hypothesis that DWH emerged in *O. rufipogon* during W1-W4 differentiation; this conditionally functional haplotype was then a target of selection for the cultivation of *O. sativa* under deepwater environments in Bangladesh.

On the basis of our present study, we suggest a model of signaling pathways underlying rice internode elongation in response to submergence via a

direct molecular link between ethylene signaling and gibberellin biosynthesis, ethylene-gibberellin relay (fig. S39). We suggest that the DWH mediates amplification of *SD1* transactivation via direct binding of OsEIL1a, resulting in increased GA_1 and GA_4; this amplification is key to allowing deepwater rice to withstand severe flooding conditions, in combination with *SK1/2*. Finally, we propose a model of evolution and domestication for nondeepwater and deepwater rice whereby the DWH emerged in *O. rufipogon* and was selected for deepwater rice cultivation in Bangladesh (fig. S40).

As contemporary climate change triggers radical shifts in weather patterns, cryptic genetic variation found in wild rice gene pools may offer adaptive solutions to help breeders fine-tune modern rice varieties. Here we reveal that a transcriptional gain-of-function allele of the Green Revolution semidwarf gene triggers rapid stem elongation in deepwater rice, enabling it to survive adverse flooding conditions. Thus, the same gene has been co-opted several times to permit rice cultivation in highly contrasting production systems via different molecular responses—decreasing enzymatic activity in one case and enhancing transactivation in the other. The capacity of *SD1* to function in such diverse roles in cultivated rice highlights the intrinsic complexity and molecular plasticity of plant adaptation strategies.

REFERENCES AND NOTES

1. H. Kende, E. van der Knaap, H.-T. Cho, *Plant Physiol.* **118**, 1105–1110 (1998).
2. K. Xu *et al*., *Nature* **442**, 705–708 (2006).
3. Y. Hattori *et al*., *Nature* **460**, 1026–1030 (2009).
4. L. A. Voesenek, J. Bailey-Serres, *New Phytol.* **206**, 57–73 (2015).
5. Y. Hattori *et al*., *Breed. Sci.* **58**, 39–46 (2008).
6. A. Sasaki *et al*., *Nature* **416**, 701–702 (2002).
7. H. Itoh, M. Ueguchi-Tanaka, Y. Sato, M. Ashikari, M. Matsuoka, *Plant Cell* **14**, 57–70 (2002).
8. C. Ju, C. Chang, *Plant Physiol.* **169**, 85–95 (2015).
9. K. Hiratsu, K. Matsui, T. Koyama, M. Ohme-Takagi, *Plant J.* **34**, 733–739 (2003).
10. J. Fukazawa *et al*., *Plant Physiol.* **175**, 1395–1406 (2017).
11. S. Yamaguchi, *Annu. Rev. Plant Biol.* **59**, 225–251 (2008).
12. K. Hirano *et al*., *Plant Cell Physiol.* **49**, 1429–1450 (2008).
13. K. Asano *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11034–11039 (2011).
14. H. Magome *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1947–1952 (2013).
15. K. Nagai *et al*., *AoB Plants* **6**, plu028 (2014).
16. M. Ueguchi-Tanaka *et al*., *Nature* **437**, 693–698 (2005).
17. K. Zhao *et al*., *Nat. Commun.* **2**, 467 (2011).
18. H. Kim *et al*., *Rice (N.Y.)* **9**, 56 (2016).

ACKNOWLEDGMENTS

We thank M. Ueguchi-Tanaka, K. Doi, K. Yano, and E. Koketsu for helpful suggestions. We also thank Y. Niimi, M. Ito, T. Noda, Y. Kurokawa, K. Akther, F. Agosto-Perez, and Y. Shi for technical assistance. C9285 was provided by the National Institute of Genetics (NIG), supported by the National BioResource Project, Agency for Medical Research and Development, Japan. The *O. rufipogon* species complex accessions were provided by the IRGC at the International Rice Research Institute, Philippines. The pSMAHdN643UGWRD vector was kindly provided by T. Mayama-Tsuchida and H. Ichikawa (NARO). **Funding:** This work was supported by JST CREST (JPMJCR13B1), JICA-JST SATREPS, the Japan Advanced Plant Science Network, The Canon Foundation, MEXT/JSPS KAKENHI (22119007, 24114001, 24114005, 14F03386, 16K18565, 16H01464, 17H06473, 17H06474, and 18K06274), the NSF Graduate Research Fellowship (DGE-1144153), and USDA NIFA (2014-67003-21858). **Author contributions:** T.Ku., S.R.M., and M.A. designed the research; T.Ku., R.G., and K.Na. assessed the deepwater response for the GWAS; T.Ku., R.G., K.Na., M.N., and K.A. carried out genetic linkage analysis; T.Ku., D.R.W., and S.R.M. carried out evolution analysis; T.Ku. analyzed transgenic rice; T.Ku., R.G., D.R.W., and K.E. genotyped rice varieties; T.F., G.T., and M.Y. performed the GWAS; M.K. and H.S. quantified gibberellin levels; Y.S., K.M., and S.Yam. evaluated enzymatic activities; T.Ku., T.Ki., M.N., A.M., and Y.M. analyzed other molecular studies; N.M., M.O.-T., J.W., S.Yan., R.Y., K.Ni., and T.M. contributed materials and tools; and T.Ku., R.G., D.R.W., S.R.M., and M.A. wrote the paper with input from other coauthors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Varieties designated by the prefix "NIAS" are available from K.E. under a material transfer agreement (MTA) with NARO, Genebank. All data are available in the main text or supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/181/suppl/DC1
Materials and Methods
Figs. S1 to S40
Tables S1 to S12
References (*19*–*39*)

5 February 2018; accepted 18 May 2018
10.1126/science.aat1577

GREENHOUSE GASES

Assessment of methane emissions from the U.S. oil and gas supply chain

Ramón A. Alvarez[1]*, Daniel Zavala-Araiza[1], David R. Lyon[1], David T. Allen[2], Zachary R. Barkley[3], Adam R. Brandt[4], Kenneth J. Davis[3], Scott C. Herndon[5], Daniel J. Jacob[6], Anna Karion[7], Eric A. Kort[8], Brian K. Lamb[9], Thomas Lauvaux[3], Joannes D. Maasakkers[6], Anthony J. Marchese[10], Mark Omara[1], Stephen W. Pacala[11], Jeff Peischl[12,13], Allen L. Robinson[14], Paul B. Shepson[15], Colm Sweeney[13], Amy Townsend-Small[16], Steven C. Wofsy[6], Steven P. Hamburg[1]

Methane emissions from the U.S. oil and natural gas supply chain were estimated by using ground-based, facility-scale measurements and validated with aircraft observations in areas accounting for ~30% of U.S. gas production. When scaled up nationally, our facility-based estimate of 2015 supply chain emissions is 13 ± 2 teragrams per year, equivalent to 2.3% of gross U.S. gas production. This value is ~60% higher than the U.S. Environmental Protection Agency inventory estimate, likely because existing inventory methods miss emissions released during abnormal operating conditions. Methane emissions of this magnitude, per unit of natural gas consumed, produce radiative forcing over a 20-year time horizon comparable to the CO_2 from natural gas combustion. Substantial emission reductions are feasible through rapid detection of the root causes of high emissions and deployment of less failure-prone systems.

Methane (CH_4) is a potent greenhouse gas, and CH_4 emissions from human activities since preindustrial times are responsible for 0.97 W m^{-2} of radiative forcing, as compared to 1.7 W m^{-2} for carbon dioxide (CO_2) (*1*). CH_4 is removed from the atmosphere much more rapidly than CO_2; thus, reducing CH_4 emissions can effectively reduce the near-term rate of warming (*2*). Sharp growth in U.S. oil and natural gas (O/NG) production beginning around 2005 (*3*) raised concerns about the climate impacts of increased natural gas use (*4*, *5*). By 2012, disagreement among published estimates of CH_4 emissions from U.S. natural gas operations led to a broad consensus that additional data were needed to better characterize emission rates (*4*–*7*). A large body of field measurements made between 2012 and 2016 (table S1) has markedly improved understanding of the sources and magnitude of CH_4 emissions from the industry's operations. Brandt *et al.* summarized the early literature (*8*); other assessments incorporated elements of recent data (*9*–*11*). This work synthesizes recent studies to provide an improved overall assessment of emissions from the O/NG supply chain, which we define to include all operations associated with O/NG production, processing, and transport (materials and methods, section S1.0) (*12*).

Measurements of O/NG CH_4 emissions can be classified as either top-down (TD) or bottom-up (BU). TD studies quantify ambient methane enhancements using aircraft, satellites, or tower networks and infer aggregate emissions from all contributing sources across large geographies. TD estimates for nine O/NG production areas have been reported to date (table S2). These areas are distributed across the U.S. (fig. S1) and account for ~33% of natural gas, ~24% of oil production, and ~14% of all wells (*13*). Areas sampled in TD studies also span the range of hydrocarbon characteristics (predominantly gas, predominantly oil, or mixed), as well as a range of production characteristics such as well productivity and maturity. In contrast, BU studies generate regional, state, or national emission estimates by aggregating and extrapolating measured emissions from individual pieces of equipment, operations, or facilities, using measurements made directly at the emission point or, in the case of facilities, directly downwind.

Recent BU studies have been performed on equipment or facilities that are expected to represent the vast majority of emissions from the O/NG supply chain (table S1). In this work, we integrate the results of recent facility-scale BU studies to estimate CH_4 emissions from the U.S. O/NG supply chain, and then we validate the results using TD studies (materials and methods). The probability distributions of our BU methodology are based on observed facility-level emissions, in contrast to the component-by-component approach used for conventional inventories. We thus capture enhancements produced by all sources within a facility, including the heavy tail of the distribution. When the BU estimate is developed in this manner, direct comparison of BU and TD estimates of CH_4 emissions in the nine basins for which TD measurements have been reported indicates agreement between methods, within estimated uncertainty ranges (Fig. 1).

Our national BU estimate of total CH_4 emissions in 2015 from the U.S. O/NG supply chain is 13 (+2.1/−1.6, 95% confidence interval) Tg CH_4/year (Table 1). This estimate of O/NG CH_4 emissions can also be expressed as a production-normalized emission rate of 2.3% (+0.4%/−0.3%) by normalizing by annual gross natural gas production [33 trillion cubic feet (*13*), with average CH_4 content of 90 volume %]. Roughly 85% of national BU emissions are from production, gathering, and processing sources, which are concentrated in active O/NG production areas.

Our assessment does not update emissions from local distribution and end use of natural gas, owing to insufficient information addressing this portion of the supply chain. However, recent studies suggest that local distribution emissions exceed the current inventory estimate (*14*–*16*), and that end-user emissions might also be important. If these findings prove to be representative, overall emissions from the natural gas supply chain would increase relative to the value in Table 1 (materials and methods, section S1.5).

Our BU method and TD measurements yield similar estimates of U.S. O/NG CH_4 emissions in 2015, and both are significantly higher than the corresponding estimate in the U.S. Environmental Protection Agency's Greenhouse Gas Inventory (EPA GHGI) (Table 1 and materials and methods, section S1.3) (*17*). Discrepancies between TD estimates and the EPA GHGI have been reported previously (*8*, *18*). Our BU estimate is 63% higher than the EPA GHGI, largely due to a more than twofold difference in the production segment (Table 1). The discrepancy in production sector emissions alone is ~4 Tg CH_4/year, an amount larger than the emissions from any other O/NG supply chain segment. Such a large difference cannot be attributed to expected uncertainty in either estimate: The extremal ends of the 95% confidence intervals for each estimate differ by 20% (i.e., ~12 Tg/year for the lower bound of our BU estimate can be compared to ~10 Tg/year for the upper bound of the EPA GHGI estimate).

We believe the reason for such large divergence is that sampling methods underlying conventional inventories systematically underestimate total emissions because they miss high emissions caused by abnormal operating conditions (e.g., malfunctions). Distributions of measured emissions from production sites in BU studies are invariably "tail-heavy," with large emission rates measured at a small subset of sites at any single point in time (*19*–*22*). Consequently, the most likely hypothesis for the difference between the EPA GHGI and BU estimates derived from facility-level measurements is that measurements used to develop GHGI emission factors

[1]Environmental Defense Fund, Austin, TX, USA. [2]University of Texas at Austin, Austin, TX, USA. [3]The Pennsylvania State University, University Park, PA, USA. [4]Stanford University, Stanford, CA, USA. [5]Aerodyne Research Inc., Billerica, MA, USA. [6]Harvard University, Cambridge, MA, USA. [7]National Institute of Standards and Technology, Gaithersburg, MD, USA. [8]University of Michigan, Ann Arbor, MI, USA. [9]Washington State University, Pullman, WA, USA. [10]Colorado State University, Fort Collins, CO, USA. [11]Princeton University, Princeton, NJ, USA. [12]University of Colorado, CIRES, Boulder, CO, USA. [13]NOAA Earth System Research Laboratory, Boulder, CO, USA. [14]Carnegie Mellon University, Pittsburgh, PA, USA. [15]Purdue University, West Lafayette, IN, USA. [16]University of Cincinnati, Cincinnati, OH, USA.
*Corresponding author. Email: ralvarez@edf.org

undersample abnormal operating conditions encountered during the BU work. Component-based inventory estimates like the GHGI have been shown to underestimate facility-level emissions (*23*), probably because of the technical difficulty and safety and liability risks associated with measuring large emissions from, for example, venting tanks such as those observed in aerial surveys (*24*).

Abnormal conditions causing high CH_4 emissions have been observed in studies across the O/NG supply chain. An analysis of site-scale emission measurements in the Barnett Shale concluded that equipment behaving as designed could not explain the number of high-emitting production sites in the region (*23*). An extensive aerial infrared camera survey of ~8000 production sites in seven U.S. O/NG basins found that ~4% of surveyed sites had one or more observable high–emission rate plumes (*24*) (detection threshold of ~3 to 10 kg CH_4/hour was two to seven times higher than mean production site emissions estimated in this work). Emissions released from liquid storage tank hatches and vents represented 90% of these sightings. It appears that abnormal operating conditions must be largely responsible, because the observation frequency was too high to be attributed to routine operations like condensate flashing or liquid unloadings alone (*24*). All other observations were due to anomalous venting from dehydrators, separators, and flares. Notably, the two largest sources of aggregate emissions in the EPA GHGI—pneumatic controllers and equipment leaks—were never observed from these aerial surveys. Similarly, a national survey of gathering facilities found that emission rates were four times higher at the 20% of facilities where substantial tank venting emissions were observed, as compared to the 80% of facilities without such venting (*25*). In addition, very large emissions from leaking isolation valves at transmission and storage facilities were quantified by means of downwind measurement but could not be accurately (or safely) measured by on-site methods (*26*). There is an urgent need to complete equipment-based measurement campaigns that capture these large-emission events, so that their causes are better understood.

In contrast to abnormal operational conditions, alternative explanations such as outdated component emission factors are unlikely to explain the magnitude of the difference between our facility-based BU estimate and the GHGI. First, an equipment-level inventory analogous to the EPA GHGI but updated with recent direct measurements of component emissions (materials and methods, section S1.4) predicts total production emissions that are within ~10% of the EPA GHGI, although the contributions of individual source categories differ significantly (table S3). Second, we consider unlikely an alternative hypothesis that systematically higher emissions during daytime sampling cause a high bias in TD methods (materials and methods, section S1.6). Two other factors may lead to low bias in EPA GHGI and similar inventory estimates. Operator cooperation is required to obtain site access for emission measurements (*8*). Operators with lower-emitting sites are plausibly more likely to cooperate in such studies, and workers are likely to be more careful to avoid errors or fix problems when measurement teams are on site or about to arrive. The potential bias due to this "opt-in" study design is very challenging to determine. We therefore rely primarily on site-level, downwind measurement methods with limited or no operator forewarning to construct our BU estimate. Another possible source of bias is measurement error. It has been suggested that malfunction of a measurement instrument widely used in the O/NG industry contributes to underestimated emissions in inventories (*27*); however, this cannot explain the more than twofold difference in production emissions (*28*).

Table 1. Summary of this work's bottom-up estimates of CH_4 emissions from the U.S. oil and natural gas (O/NG) supply chain (95% confidence interval) and comparison to the EPA Greenhouse Gas Inventory (GHGI).

Industry segment	2015 CH_4 emissions (Tg/year)	
	This work (bottom-up)	EPA GHGI (*17*)
Production	7.6 (+1.9/−1.6)	3.5
Gathering	2.6 (+0.59/−0.18)	2.3
Processing	0.72 (+0.20/−0.071)	0.44
Transmission and storage	1.8 (+0.35/−0.22)	1.4
Local distribution*	0.44 (+0.51/−0.22)	0.44
Oil refining and transportation*	0.034 (+0.050/−0.008)	0.034
U.S. O/NG total	13 (+2.1/−1.7)	8.1 (+2.1/−1.4)†

*This work's emission estimates for these sources are taken directly from the GHGI. The local distribution estimate is expected to be a lower bound on actual emissions and does not include losses downstream of customer meters due to leaks or incomplete combustion (materials and methods, section S1.5).
†The GHGI only reports industry-wide uncertainties.

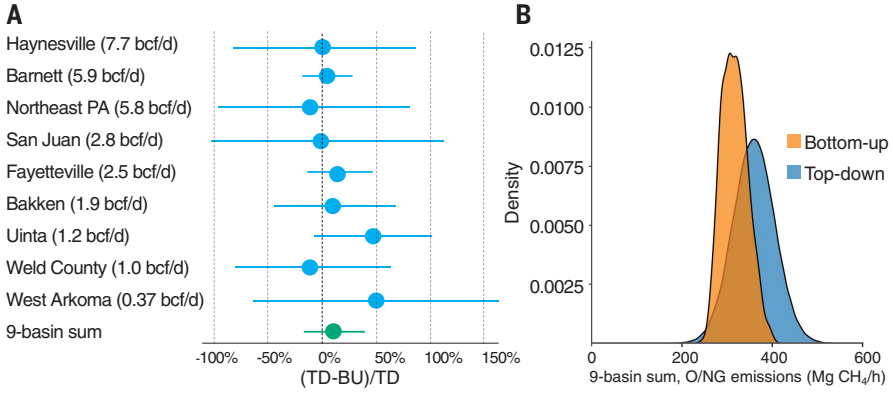


Fig. 1. Comparison of this work's bottom-up (BU) estimates of methane emissions from oil and natural gas (O/NG) sources to top-down (TD) estimates in nine U.S. O/NG production areas. (**A**) Relative differences of the TD and BU mean emissions, normalized by the TD value, rank ordered by natural gas production in billion cubic feet per day (bcf/d, where 1 bcf = 2.8×10^7 m^3). Error bars represent 95% confidence intervals. (**B**) Distributions of the nine-basin sum of TD and BU mean estimates (blue and orange probability density, respectively). Neither the ensemble of TD-BU pairs (A) nor the nine-basin sum of means (B) are statistically different [$p = 0.13$ by a randomization test, and mean difference of 11% (95% confidence interval of −17 to 41%)].

The tail-heavy distribution for many O/NG CH_4 emission sources has important implications for mitigation because it suggests that most sources—whether they represent whole facilities or individual pieces of equipment—can have lower emissions when they operate as designed. We anticipate that significant emissions reductions could be achieved by deploying well-designed emission detection and repair systems that are capable of identifying abnormally operating facilities or equipment. For example, pneumatic controllers and equipment leaks are the largest emission sources in the O/NG production segment exclusive of missing emission sources (38 and 21%, respectively; table S3), with malfunctioning controllers contributing 66% of total pneumatic controller emissions (materials and methods, section S1.4) and equipment leaks 60% higher than the GHGI estimate.

Gathering operations, which transport unprocessed natural gas from production sites to processing plants or transmission pipelines, produce ~20% of total O/NG supply chain CH_4 emissions. Until the publication of recent measurements (*29*), these emissions were largely unaccounted by the EPA GHGI. Gas processing, transmission and storage together contribute another ~20% of total O/NG supply chain emissions, most of which come from ~2500 processing and compression facilities.

Our estimate of emissions from the U.S. O/NG supply chain (13 Tg CH_4/year) compares to the EPA estimate of 18 Tg CH_4/year for all other anthropogenic CH_4 sources (*17*). Natural gas losses are a waste of a limited natural resource (~$2 billion/year), increase global levels of surface ozone pollution (*30*), and substantially erode the potential climate benefits of natural gas use. Indeed, our estimate of CH_4 emissions across the supply chain, per unit of gas consumed, results in roughly the same radiative forcing as does the CO_2 from combustion of natural gas over a 20-year time horizon (31% over 100 years). Moreover, the climate impact of 13 Tg CH_4/year over a 20-year time horizon roughly equals that from the annual CO_2 emissions from all U.S. coal-fired power plants operating in 2015 (31% of the impact over a 100-year time horizon) (materials and methods, section S1.7).

We suggest that inventory methods would be improved by including the substantial volume of missing O/NG CH_4 emissions evident from the large body of scientific work now available and synthesized here. Such empirical adjustments based on observed data have been previously used in air quality management (*31*).

The large spatial and temporal variability in CH_4 emissions for similar equipment and facilities (due to equipment malfunction and other abnormal operating conditions) reinforces the conclusion that substantial emission reductions are feasible. Key aspects of effective mitigation include pairing well-established technologies and best practices for routine emission sources with economically viable systems to rapidly detect the root causes of high emissions arising from abnormal conditions. The latter could involve combinations of current technologies such as on-site leak surveys by company personnel using optical gas imaging (*32*), deployment of passive sensors at individual facilities (*33*, *34*) or mounted on ground-based work trucks (*35*), and in situ remote-sensing approaches using tower networks, aircraft, or satellites (*36*). Over time, the development of less failure-prone systems would be expected through repeated observation of and further research into common causes of abnormal emissions, followed by re-engineered design of individual components and processes.

REFERENCES AND NOTES

1. G. Myhre *et al*., in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, Cambridge, UK, 2013); www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_Chapter08_FINAL.pdf.
2. J. K. Shoemaker, D. P. Schrag, M. J. Molina, V. Ramanathan, *Science* **342**, 1323–1324 (2013).
3. U.S. Energy Information Administration (EIA), "Annual Energy Outlook 2017" (EIA, 2017); www.eia.gov/outlooks/aeo/.
4. R. W. Howarth, R. Santoro, A. Ingraffea, *Clim. Change* **106**, 679–690 (2011).
5. R. A. Alvarez, S. W. Pacala, J. J. Winebrake, W. L. Chameides, S. P. Hamburg, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6435–6440 (2012).
6. U.S. Department of Energy (DOE), "Ninety-day report of the Secretary of Energy Advisory Board's Shale Gas Subcommittee" (2011); https://energy.gov/downloads/90-day-interim-report-shale-gas-production-secretary-energy-advisory-board.
7. National Petroleum Council (NPC), "Prudent Development: Realizing the Potential of North America's Abundant Natural Gas and Oil Resources" (NPC, 2011); www.npc.org.
8. A. R. Brandt *et al*., *Science* **343**, 733–735 (2014).
9. D. T. Allen, *J. Air Waste Manag. Assoc.* **66**, 549–575 (2016).
10. P. Balcombe, K. Anderson, J. Speirs, N. Brandon, A. Hawkes, *ACS Sustain. Chem.& Eng.* **5**, 3–20 (2017).
11. J. A. Littlefield, J. Marriott, G. A. Schivley, T. J. Skone, *J. Clean. Prod.* **148**, 118–126 (2017).
12. See supplementary materials.
13. Drillinginfo, Inc., Drillinginfo Production Query (2015); https://info.drillinginfo.com/.
14. K. McKain *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1941–1946 (2015).
15. B. K. Lamb *et al*., *Environ. Sci. Technol.* **50**, 8910–8917 (2016).
16. D. Wunch *et al*., *Atmos. Chem. Phys.* **16**, 14091–14105 (2016).
17. U.S. Environmental Protection Agency (EPA), "Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2015" (EPA, 2017); www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2015.
18. D. Zavala-Araiza *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15597–15602 (2015).
19. C. W. Rella, T. R. Tsai, C. G. Botkin, E. R. Crosson, D. Steele, *Environ. Sci. Technol.* **49**, 4742–4748 (2015).
20. M. Omara *et al*., *Environ. Sci. Technol.* **50**, 2099–2107 (2016).
21. A. M. Robertson *et al*., *Environ. Sci. Technol.* **51**, 8832–8840 (2017).
22. A. R. Brandt, G. A. Heath, D. Cooley, *Environ. Sci. Technol.* **50**, 12512–12520 (2016).
23. D. Zavala-Araiza *et al*., *Nat. Commun.* **8**, 14012 (2017).
24. D. R. Lyon *et al*., *Environ. Sci. Technol.* **50**, 4877–4886 (2016).
25. A. L. Mitchell *et al*., *Environ. Sci. Technol.* **49**, 3219–3227 (2015).
26. D. J. Zimmerle *et al*., *Environ. Sci. Technol.* **49**, 9374–9383 (2015).
27. T. Howard, T. W. Ferrara, A. Townsend-Small, *J. Air Waste Manag. Assoc.* **65**, 856–862 (2015).
28. R. A. Alvarez, D. R. Lyon, A. J. Marchese, A. L. Robinson, S. P. Hamburg, *Elem. Sci. Anth.* **4**, 000137 (2016).
29. A. J. Marchese *et al*., *Environ. Sci. Technol.* **49**, 10718–10727 (2015).
30. A. M. Fiore *et al*., *Geophys. Res. Lett.* **29**, 21-1–25-4 (2002).
31. Texas Commission on Environmental Quality (TCEQ), "Houston-Galveston-Brazoria Attainment Demonstration State Implementation Plan Revision for the 1997 Eight-Hour Ozone Standard" (2010), pp. 3–18; www.tceq.texas.gov/assets/public/implementation/air/sip/hgb/hgb_sip_2009/09017SIP_completeNarr_ado.pdf.
32. A. P. Ravikumar, J. Wang, A. R. Brandt, *Environ. Sci. Technol.* **51**, 718–724 (2017).
33. U.S. Department of Energy (DOE) Advanced Research Projects Agency – Energy, (ARPA-E, 2014), "ARPA-E MONITOR Program" (ARPA-E); https://arpa-e.energy.gov/?q=programs/monitor.
34. Environmental Defense Fund (EDF), "Methane Detectors Challenge" (EDF, 2014); www.edf.org/energy/natural-gas-policy/methane-detectors-challenge.
35. J. D. Albertson *et al*., *Environ. Sci. Technol.* **50**, 2487–2497 (2016).
36. D. J. Jacob *et al*., *Atmos. Chem. Phys.* **16**, 14371–14396 (2016).

ACKNOWLEDGMENTS

The authors are grateful to R. Harriss for support in the design and conduct of studies. We thank D. Zimmerle, A. Robertson, and A. Pintar for helpful discussions, and the scores of researchers that contributed to the body of work assessed here. **Funding:** Alfred P. Sloan Foundation, Fiona and Stan Druckenmiller, Heising-Simons Foundation, Bill and Susan Oberndorf, Betsy and Sam Reeves, Robertson Foundation, TomKat Charitable Trust, and the Walton Family Foundation (for EDF authors as well as support of related studies involving D.T.A, S.C.H., A.K., E.A.K., B.K.L., A.J.M., A.L.R., P.B.S., C.S., A.T.-S., S.C.W.); DOE National Energy Technology Laboratory (Z.R.B., K.J.D., T.L., A.L.R.); NASA Earth Science Division (D.J.J., E.A.K., J.D.M.); NOAA Climate Program Office (E.A.K., J.P., A.L.R., C.S.). **Author contributions:** R.A.A., D.Z-A., D.R.L., and S.P.H. conceived the study; R.A.A., D.Z-A., D.R.L., E.A.K., S.W.P. and S.P.H. designed the study and interpreted results with input from all authors; each author contributed to the collection, analysis, or assessment of one or more datasets necessary to perform this study; D.Z-A, D.R.L, and S.W.P, performed the analysis, with contributions from R.A.A., A.R.B., A.K., and M.O.; R.A.A., D.Z-A., D.R.L., S.W.P., S.C.W., and S.P.H. wrote the manuscript with input from all authors. **Competing interests**: None declared. **Data and materials availability:** All data and methods needed to reproduce the results in the paper are provided in the paper or as supplementary materials. Additional author disclosures (affiliations, funding sources, financial holdings) are provided in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/186/suppl/DC1
Materials and Methods
Additional Author Disclosures
Figs. S1 to S11
Tables S1 to S12
References (*37*–*77*)
Databases S1 and S2

19 December 2017; accepted 18 May 2018
Published online 21 June 2018
10.1126/science.aar7204

EARLY DEVELOPMENT

Dual-spindle formation in zygotes keeps parental genomes apart in early mammalian embryos

Judith Reichmann[1], Bianca Nijmeijer[1], M. Julius Hossain[1], Manuel Eguren[1], Isabell Schneider[1], Antonio Z. Politi[1], M. Julia Roberti[1], Lars Hufnagel[1], Takashi Hiiragi[2], Jan Ellenberg[1]*

At the beginning of mammalian life, the genetic material from each parent meets when the fertilized egg divides. It was previously thought that a single microtubule spindle is responsible for spatially combining the two genomes and then segregating them to create the two-cell embryo. We used light-sheet microscopy to show that two bipolar spindles form in the zygote and then independently congress the maternal and paternal genomes. These two spindles aligned their poles before anaphase but kept the parental genomes apart during the first cleavage. This spindle assembly mechanism provides a potential rationale for erroneous divisions into more than two blastomeric nuclei observed in mammalian zygotes and reveals the mechanism behind the observation that parental genomes occupy separate nuclear compartments in the two-cell embryo.

After fertilization, the haploid genomes of egg and sperm come together to form the genome of a new diploid organism, a moment that is of fundamental biological importance. In mammals, parental chromosomes meet for the first time upon entry into the first zygotic mitosis after nuclear envelope breakdown (NEBD). It has been assumed that, as observed in oocytes, a single bipolar microtubule system would self-assemble around both parental genomes in zygotes (*1*–*6*). However, owing to the extreme light sensitivity of the mammalian embryo, the details of the dynamic process of zygotic spindle assembly remain unclear.

To examine how parental genomes join for the first time, we imaged live embryos in which the maternal and paternal centromeres were differentially labeled (*7*) using our recently developed inverted light-sheet microscope, which allows fast three-dimensional (3D) imaging of embryonic development owing to its low phototoxicity (*8*). This revealed that the two genomes remain spatially separate throughout the first mitosis (fig. S1 and movie S1). To understand why the genomes are not mixed, we next imaged spindle assembly using fluorescently labeled microtubule organizing centers (MTOCs) and spindle microtubules (Fig. 1A and movie S2). Newly nucleated microtubules self-organized into two separate bipolar spindles after NEBD, attracting a subset of the cytoplasmic MTOCs that had accumulated around each pronucleus to their poles (Fig. 1, A and B). Subsequently, the two spindles aligned and came into close apposition to form a compound barrel–shaped system. This structure typically had two clusters of MTOCs at at least one of its poles, suggesting that the two spindles were aligned closely but not completely merged (Fig. 1, A and B, and movie S2). To probe this further, we performed 3D immunofluorescence analysis of zygotes, visualizing endogenous spindle poles, microtubules, kinetochores, and DNA. In early and mid pro-metaphase, two separate bipolar spindles were formed in in vivo–developed zygotes (Fig. 1, C and D). Given the delayed and partial association of MTOCs with the microtubule mass, we hypothesized that dual-spindle formation might be driven by self-assembly of microtubules nucleated by chromosomes. To test this, we assayed microtubule regrowth after washing out the microtubule-depolymerizing drug Nocodazole. Indeed, a large proportion of microtubules was nucleated on chromosomes, particularly at kinetochores (fig. S2), whereas MTOCs became associated with microtubules only later. This observation prompted us to investigate the organization of K-fibers in zygotic pro-metaphase by means of high-resolution immunofluorescence of zygotes fixed after brief cold treatment in order to highlight stable microtubules (Fig. 1D and fig. S3). Two bipolar arrays of K-fibers started assembling in early pro-metaphase and were stably organized in mid pro-metaphase, and their remnants were clearly recognizable by their slightly offset centers and split poles even after parallelization in metaphase, during which the two spindles were no longer separated by a large gap.

To characterize the kinetics of zygotic spindle assembly in live embryos, we next imaged maternal and paternal centromeres in relation to growing spindle microtubule tips and observed three phases of zygotic spindle assembly (Fig. 2A). A transient first phase (~3 min; 10.3 ± 3.5 to 13. 4 ± 4 min after NEBD), characterized by the clustering of growing microtubules around the two pronuclei, was followed by phase 2 (~16 min; 14.5 ± 4 to 30.7 ± 6.5 min after NEBD), in which individual bipolar spindles assembled around each parental genome, and subsequently, phase 3 (~83 min; 46.7 ± 17 to 129.2 ± 16.5 min after NEBD), when the two spindles aligned and combined into a compound barrel–shaped structure.

To test whether the two zygotic spindles are functionally independent, we measured the timing and direction of maternal and paternal chromosome congression (fig. S4, A and B, and supplementary materials, materials and methods). Congression started in pro-metaphase (phase 2), while the spindles were clearly separated (fig. S4, A and B). Parental chromosome congression was not correlated in time until shortly before anaphase, suggesting that they are moved by different microtubule systems (fig. S4C). Furthermore, the parental genomes were congressed in different directions along separate spindle axes, as evidenced by the large difference between the angles of the two forming metaphase plates, which became parallel only later during dual spindle alignment in phase 3 (Fig. 2, B to D, and fig. S4, D and E). Tracking of growing microtubule tips showed two major directions of microtubule flow during phase 2 and one main direction during phase 3, as indicated by the corresponding kymograph profiles (fig. S5 and movies S3 to S8). The independent congression frequently led to an offset in the final bioriented position of the paternal and maternal metaphase plates at the end of phase 3 before and during segregation (Fig. 2A, arrows). Thus, each of the two spindles around the parental genomes functions independently for chromosome congression and may even function separately in chromosome segregation.

In in vitro fertilization (IVF) clinics, the zygotic division is error prone and often leads to embryos with blastomeres that contain two nuclei (*9*–*13*). We hypothesized that failure to align the two parental spindles before anaphase could explain this enigmatic phenotype. To test this, we increased the distance between the two pronuclei with transient treatment with Nocodazole, which led to a larger gap between the two self-assembling spindles (Fig. 3 and movies S9 to S11). Indeed, such embryos frequently failed to fully align the parental spindles at one or both poles. This did not delay anaphase but resulted in chromosome segregation by two spindles into different directions, leading to two cell embryos with one or two binucleated blastomeres (Fig. 3 and movies S10 and S11). By contrast, embryos that did align the two spindles parallel to each other before anaphase cleaved into two blastomeres with single nuclei (movie S9). Thus, failure to align the two zygotic spindles gives rise to multinucleated two-cell embryos, phenocopying frequently observed errors in human embryonic development in IVF clinics.

Dual-spindle assembly in the mammalian zygote would also offer a mechanistic explanation for the long-standing observation that the parental genomes occupy separate compartments inside the nuclei of two- and four-cell embryo

[1]Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. [2]Developmental Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany.
*Corresponding author. Email: jan.ellenberg@embl.de

blastomeres (*14*, *15*). If dual-spindle assembly around two pronuclei was responsible for genome compartmentalization, parental genomes should mix in subsequent divisions, in which only one nucleus is present per cell. Imaging of the metaphase plate of live hybrid mouse embryos from the zygote to the eight-cell stage showed that parental genomes were separated in zygotes but rapidly became mixed in the subsequent developmental stages, as predicted (fig. S6, A to D). This loss of genome compartmentalization was also seen in in vivo–developed isogenic embryos (fig. S6, E to I). Thus, parental genomes are kept separate by two spindles only during the first mitosis but then mix during subsequent divisions, driven by a single common spindle.

If dual-spindle assembly is the mechanism for parental genome compartmentalization (fig. S7A), formation of a single spindle around both genomes in the zygote should already mix them in the first division. To test this prediction, we redirected spindle assembly with two small-molecule inhibitors of microtubule polymerization (Nocodazole) and the motor protein Eg5 (Monastrol). Transient treatment with Monastrol collected both genomes in a single microtubule aster, and subsequent depolymerization of microtubules with Nocodazole

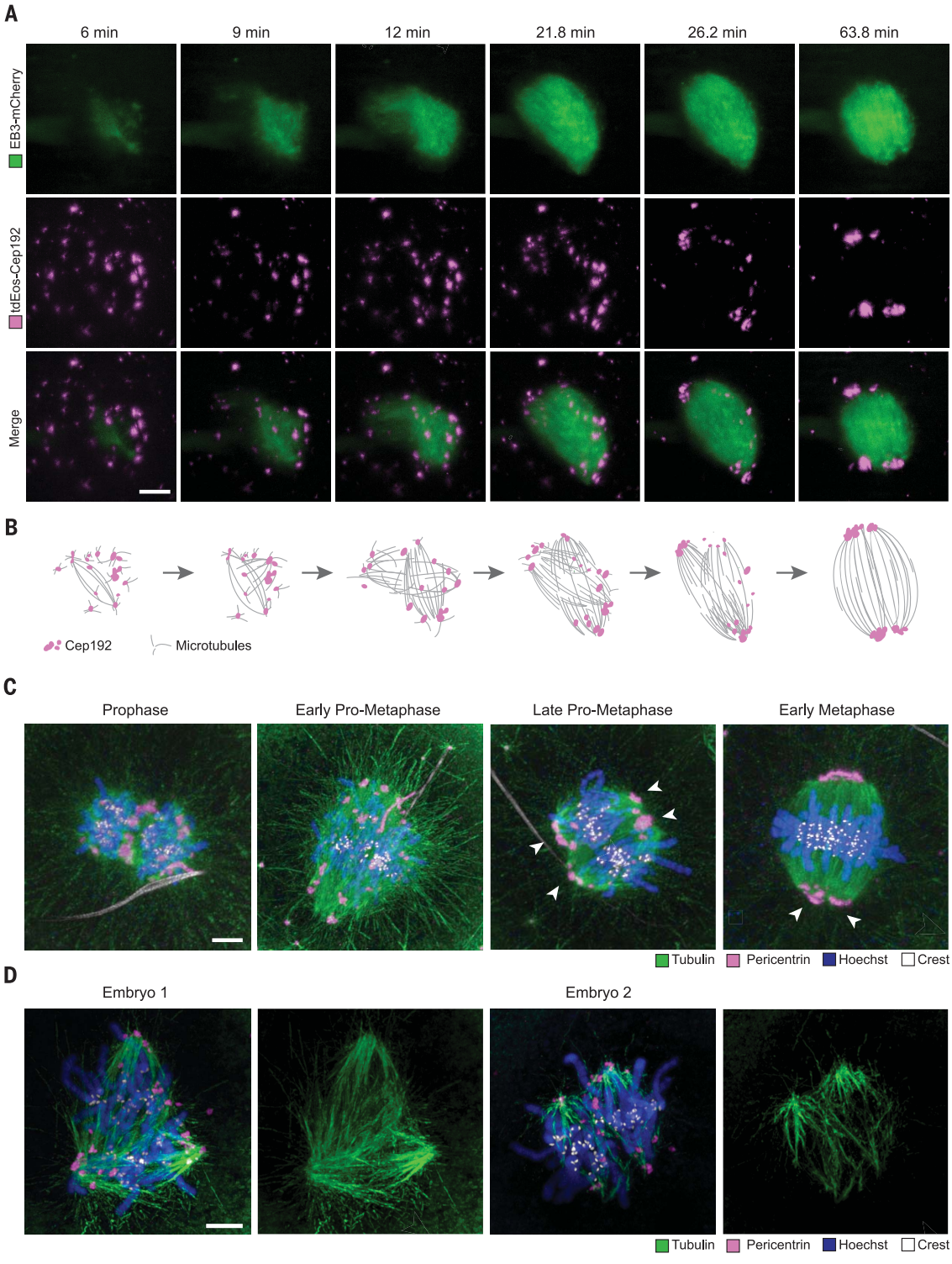


Fig. 1. Individual bipolar spindle formation around each pronucleus. (**A**) Time-lapse imaging of *Mus musculus* x *Mus musculus* (MMU x MMU) zygotes expressing EB3-mCherry (marker for microtubules; green) and tdEos-Cep192 (marker for MTOCs; magenta). Scale bar, 10 μm. In 10 out of 13 zygotes, both or at least one of the dual-spindle poles remained clearly split after the spindles had parallelized. (**B**) Schematic diagram showing progression of dual-spindle formation based on the data presented in (A) and data shown subsequently in the manuscript. Microtubules, gray; MTOCs, magenta. (**C**) Immunofluorescence staining of MMU x MMU zygotes fixed at consecutive stages of development. Maximum *z* projections of confocal sections of zygotes at prophase, early pro-metaphase, late pro-metaphase, and early metaphase. White arrowheads indicate poles. (**D**) Immunofluorescence staining of cold-treated MMU x MMU pro-metaphase zygotes. Maximum *z* projections of confocal sections. Microtubules, α-Tubulin (green); MTOCs, Pericentrin (magenta); kinetochores, Crest (white); and DNA, Hoechst (blue). Scale bars, 5 μm in (C) and (D).

followed by regrowth after washout then resulted in one bipolar spindle around both genomes (figs. S7B and S8, MoNoc-treated zygotes). Such MoNoc-treated embryos captured and congressed chromosomes within a single spindle and showed a high degree of parental genome mixing in the first mitotic metaphase (Fig. 4). This was substantially different from untreated or control zygotes in which the order of drug treatments is reversed (NocMo-treated zygotes), which maintained dual-spindle formation and genome separation (Fig. 4 and figs. S7C and S8). Thus, dual-spindle formation in the zygote is required for parental genome separation in mammals.

Having this experimental method to induce mixing of the parental genomes in hand allowed us to demonstrate that genome separation was not required for parental genome epigenetic asymmetry and its resolution (*16*–*19*), as proposed previously (figs. S9 and S10) (*15*, *20*–*22*).

Here, we showed that two spindles form around pronuclei in mammalian zygotes, which individually collect the parental genomes and then position them next to each other before the

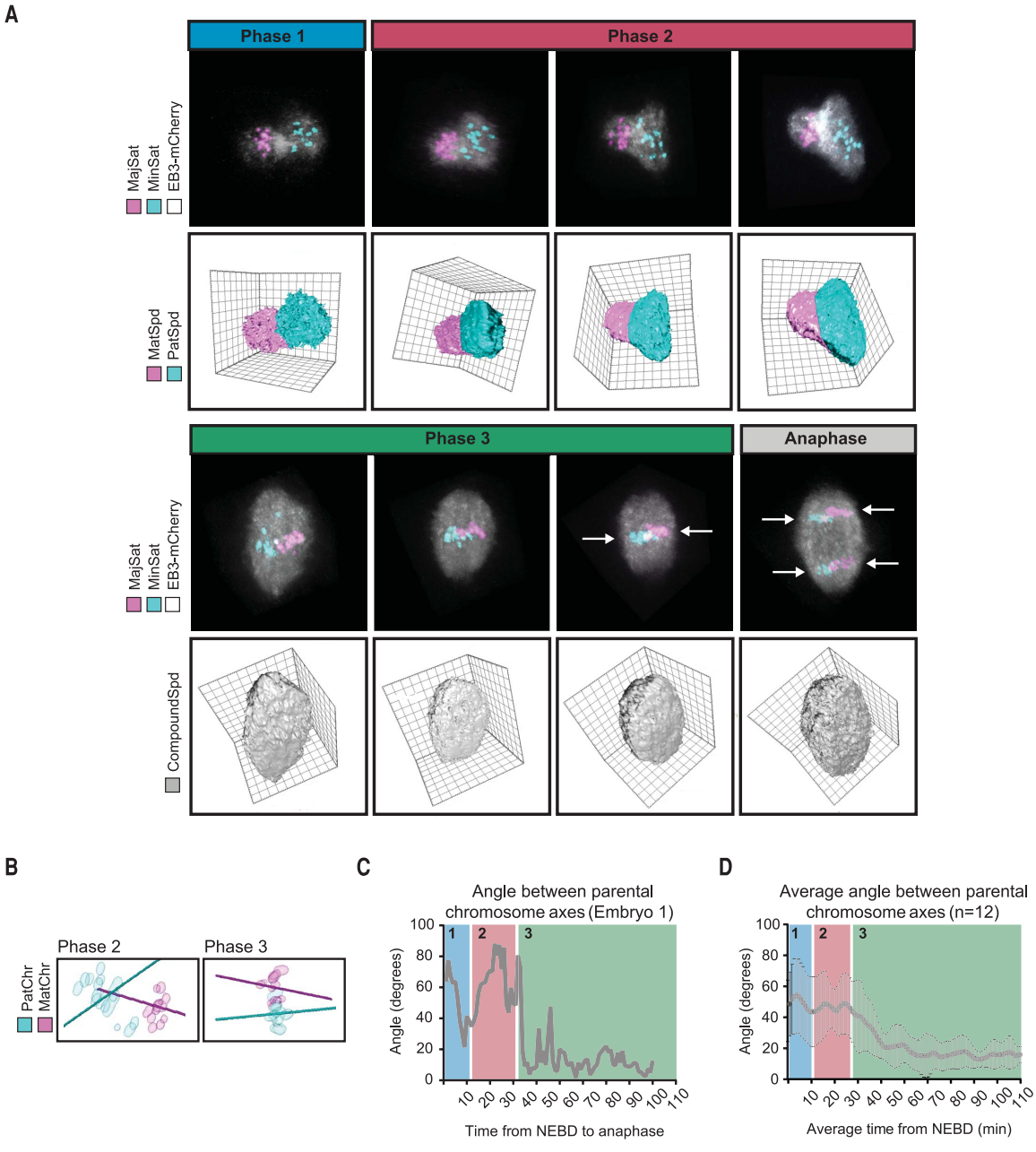


Fig. 2. Spindle assembly and chromosome dynamics in the zygote. (**A**) Time-lapse imaging of *Mus musculus* x *Mus Spretus* (MMU x MSP) zygotes expressing EB3-mCherry and fluorescent TALEs to label maternal and paternal chromosomes though distinction of Major satellites (MajSat) and Minor satellites (MinSat). Phase 1: Microtubule ball formation around pronuclei. Phase 2: Bipolarization of maternal and paternal spindle. Phase 3: Formation of single barrel–shaped spindle. (Top and top middle) 3D-rotated images of the whole spindle volume. Maternal chromosomes, MajSat (magenta); paternal chromosomes, MinSat (cyan); microtubules, EB3-mCherry (white). (Bottom middle and bottom) Segmentation of maternal (MatSpd; magenta) and paternal (PatSpd; cyan) spindles in phase 1 and phase 2 and single bipolar spindle in phase 3 (CompoundSpd; gray). Offset between maternal and paternal chromosomes at metaphase and anaphase is indicated with white arrows. (**B**) Schematic of measurements on maternal (MatChr) and paternal chromatin masses (PatChr) in phases 2 and 3. (**C** and **D**) Angle between maternal and paternal chromosome axis over time for (C) a single embryo and (D) averaged for 12 embryos (mean ± SD) is shown. Phase 1, blue; phase 2, red; phase 3, green.

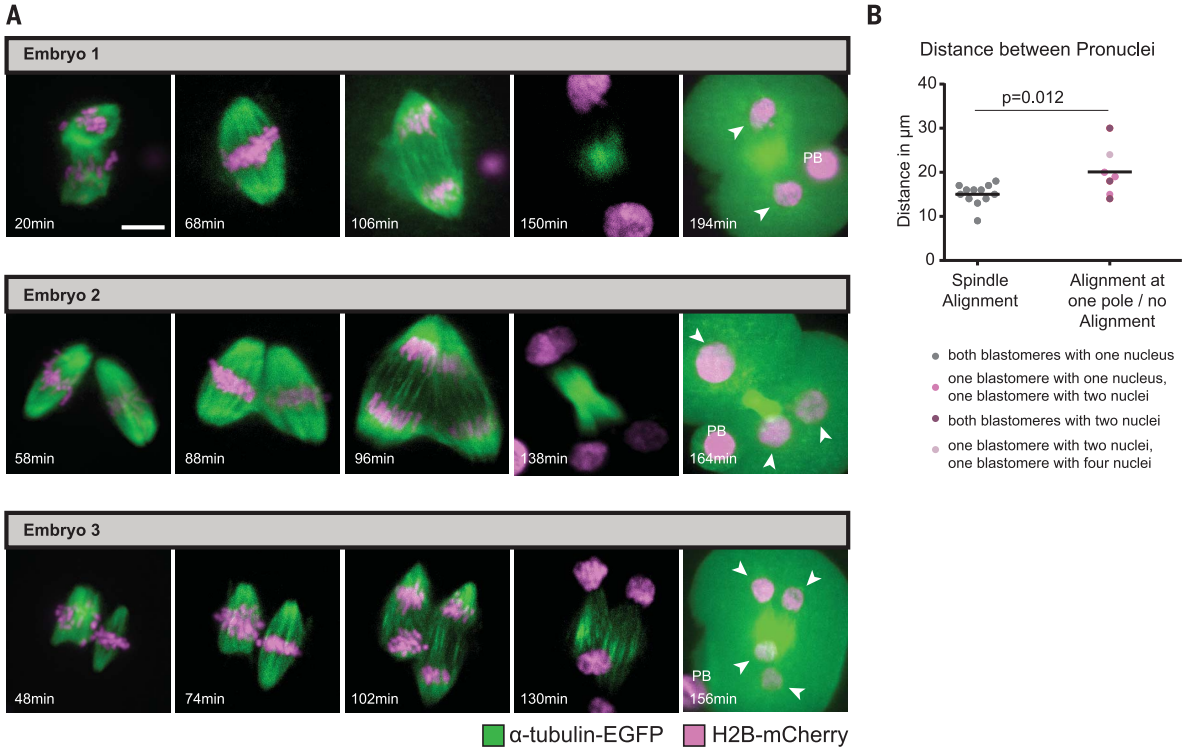


Fig. 3. Proximity dependency of bipolar spindle fusion. (**A**) Time-lapse imaging of MMU x MMU zygotes expressing H2B-mCherry (chromatin; magenta) and αTubulin–enhanced green fluorescent protein (EGFP) (microtubules; green). Shown is spindle morphology from pro-metaphase to postmitosis in three representative zygotes treated with Nocodazole for >10 hours. Maximum *z* projections are of pro-metaphase, metaphase, anaphase, telophase, and postmitosis. Arrowheads and "PB" indicate nuclei and polar body, respectively. Scale bar, 10 µm. In the absence of NEBD as a timing reference, anaphase onset was set at 90 min (average time from NEBD to anaphase in MMU x MMU zygotes), and the other times were calculated accordingly. (**B**) Initial distance of pro-nuclei (n = 19). Statistics, Student's t test.

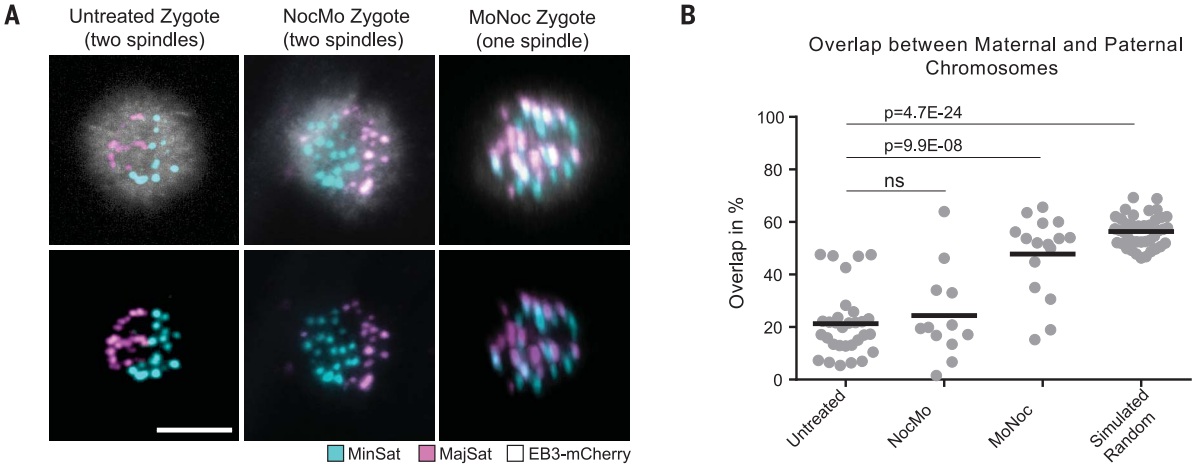


Fig. 4. Distribution of parental centromeres in control, NocMo-treated, and MoNoc-treated zygotes. (**A**) Differential labeling of maternal (MajSat; magenta) and paternal (MinSat; cyan) centromeres through distinction of single-nucleotide polymorphisms by means of fluorescent TALEs. Mitotic spindle is labeled with EB3-mCherry (white). Representative *z*-projected images of parental chromosome distribution in untreated, MoNoc, and NocMo MMU x MSP zygotes. Scale bar, 10 µm. (**B**) Degree of overlap between 3D convex hulls of parental chromosomes for untreated (n = 31), MoNoc (n = 16), and NocMo (n = 12) zygotes and embryos with in silico randomized distribution (n = 40) (fig. S1 and supplementary materials, materials and methods). Statistics, Student's t test.

first anaphase by aligning the two spindles parallel to each other. Our data explain how parental genome separation is achieved in mammalian embryos. To date, the formation of physically distinct mitotic spindles around the two pronuclei has been thought to be specific to certain arthropod species (*23*, *24*). Our finding that this occurs also during mammalian pro-metaphase suggests that two zygotic spindles might be characteristic for many species that maintain separate pronuclei after fertilization. Failure to align the two spindles produced errors in the zygotic division that closely resemble clinical phenotypes of human embryos in IVF procedures, suggesting that a similar mechanism of dual

zygotic spindle assembly also occurs in humans. This view is supported by the spatial separation of parental chromosomes reported in human zygotes (*25*) and divisions of human zygotes into multinucleated blastomeres, as well as by reports that identified a mix of paternal, maternal, and diploid cells in eight-cell cattle embryos. A dual zygotic spindle would provide a mechanistic basis for this parental genome segregation (*9*–*13*, *26*, *27*). These severe and relatively frequent zygotic division errors in human and agriculturally used mammals thus find their likely mechanistic explanation in a failure of the close alignment of the two zygotic spindles before anaphase. If a similar mechanism of microtubule-driven parental genome separation indeed occurs in human zygotes, it would be important from an ethical and legal perspective because "pronuclear fusion," a process that strictly speaking does not occur in mouse zygotes, is used to define the beginning of embryonic life as protected by law in several countries (for example, Germany, *§ 8 Abs. 1* Embryonenschutzgesetz).

REFERENCES AND NOTES

1. A. Courtois, M. Schuh, J. Ellenberg, T. Hiiragi, *J. Cell Biol.* **198**, 357–370 (2012).
2. G. Fitzharris, *Development* **136**, 2111–2119 (2009).
3. P. A. Coelho *et al*., *Dev. Cell* **27**, 586–597 (2013).
4. M. Schuh, J. Ellenberg, *Cell* **130**, 484–498 (2007).
5. T. S. Kitajima, M. Ohsugi, J. Ellenberg, *Cell* **146**, 568–581 (2011).
6. Z. Holubcová, M. Blayney, K. Elder, M. Schuh, *Science* **348**, 1143–1147 (2015).
7. Y. Miyanari, C. Ziegler-Birling, M. E. Torres-Padilla, *Nat. Struct. Mol. Biol.* **20**, 1321–1324 (2013).
8. P. Strnad *et al*., *Nat. Methods* **13**, 139–142 (2016).
9. B. Kalatova, R. Jesenska, D. Hlinka, M. Dudas, *Acta Histochem.* **117**, 111–125 (2015).
10. H. Balakier, K. Cadesky, *Hum. Reprod.* **12**, 800–804 (1997).
11. J. Meriano, C. Clark, K. Cadesky, C. A. Laskin, *Reprod. Biomed. Online* **9**, 511–520 (2004).
12. A. Egashira *et al*., *J. Reprod. Dev.* **61**, 595–600 (2015).
13. I. Kligman, C. Benadiva, M. Alikani, S. Munne, *Hum. Reprod.* **11**, 1492–1498 (1996).
14. W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **403**, 501–502 (2000).
15. W. Mayer, A. Smith, R. Fundele, T. Haaf, *J. Cell Biol.* **148**, 629–634 (2000).
16. M. Puschendorf *et al*., *Nat. Genet.* **40**, 411–420 (2008).
17. K. Iqbal, S. G. Jin, G. P. Pfeifer, P. E. Szabó, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3642–3647 (2011).
18. A. Inoue, Y. Zhang, *Science* **334**, 194 (2011).
19. M. Wossidlo *et al*., *Nat. Commun.* **2**, 241 (2011).
20. A. V. Probst, G. Almouzni, *Differentiation* **76**, 15–23 (2008).
21. R. Duffié, D. Bourc'his, *Curr. Top. Dev. Biol.* **104**, 293–328 (2013).
22. R. De La Fuente, C. Baumann, M. M. Viveiros, *Development* **142**, 1806–1817 (2015).
23. R. R. Snook, D. J. Hosken, T. L. Karr, *Reproduction* **142**, 779–792 (2011).
24. N. Kawamura, *Dev. Growth Differ.* **43**, 343–349 (2001).
25. C. van de Werken *et al*., *Nat. Commun.* **5**, 5868 (2014).
26. A. Destouni *et al*., *Genome Res.* **26**, 567–578 (2016).
27. A. Destouni, J. R. Vermeesch, *BioEssays* **39**, 1600226 (2017).

ACKNOWLEDGMENTS

We thank N. Daigle for cloning the EB3-mCherry plasmid; P. Neveu and M. Schuh for providing tdiRFP670 and tdEos-Cep192, respectively; and N. Galjart for providing full-length *Homo sapiens* EB3 cDNA (NM_001303050.1). We thank EMBL's laboratory of animal resources for support, P. Strnad for development and assistance of the inverted light-sheet microscope, A. Rybina for assistance with the Zeiss LSM 880 Airy microscope, Arivis for support in image analysis, and the EMBL Advanced Light Microscopy Facility for support in image acquisition and analysis. We thank J. Reddington and S. Alexander for critical reading of the manuscript. **Funding:** This work was supported by funds from the European Research Council (ERC Advanced Grant "Corema," grant agreement 694236) to J.E. and by the European Molecular Biology Laboratory (all authors). J.R. was further supported by the EMBL Interdisciplinary Postdoc Programme (EIPOD) under Marie Curie Actions COFUND; M.E. by the EMBO long-term postdoctoral fellowship and EC Marie Slodowska-Curie postdoctoral fellowship; I.S. by a Boehringer Ingelheim Fonds Ph.D. fellowship; and M.J.R. by a Humboldt Foundation postdoctoral fellowship. I.S. is a candidate for a joint Ph.D. between EMBL and Heidelberg University, Faculty of Biosciences. **Author contributions:** J.E. and J.R. conceived the project and designed the experiments. J.R., B.N., M.E., and I.S. performed the experiments. M.J.R. supported the mouse EDU experiments. J.R., M.J.H., and A.Z.P. analyzed the data. T.H. and L.H. contributed to conception and design of the work. J.E. and J.R. wrote the manuscript. All authors contributed to the interpretation of the data and read and approved the final manuscript. **Competing interests:** L.H. and J.E. are scientific cofounders and advisers of Luxendo GmbH (part of Bruker), which makes light-sheet–based microscopes commercially available. **Data and materials availability:** All images processed in this study are available in the Image Data Resource (IDR), accession number idr0045 (https://idr.openmicroscopy.org/webclient/?show=project-405). All code is available from EMBL's git depository (https://git.embl.de/grpellenberg/dual_zygotic_spindles_reichmann_2018).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6398/189/suppl/DC1
Materials and Methods
Figs. S1 to S10
Table S1
Movies S1 to S11
References (*28*–*34*)

15 December 2017; accepted 8 June 2018
10.1126/science.aar7462

PINS

RESTORING HOPE

Since the first Deep Brain Stimulation initiative of Tsinghua Univeristy in 2000, PINS Medical has gradually established a multinational corporation with headquarter based in Beijing and international business center in Singapore. As an innovative high-tech enterprise with focus on neuromodulation, a variety of clinical products have been developed to date, which include stimulators for deep brain, vagus nerve, spinal cord and sacral nerve stimulation therapies. PINS Medical devotes itself to providing cutting-edge treatments for patients who suffer from neurological disorders such as Parkinson's Disease, Epilepsy, Chronic Pain and OAB, etc.

As part of the "National Engineering Laboratory for Neuromodulation", PINS Medical works in close cooperation with Tsinghua University and the numerous affiliated clinical centers, becoming a center of attraction for a wide range of professional talents in areas of clinical research, innovative R&D and business management. Since 2008, PINS Medical has developed rapidly in becoming a leading brand in neuromodulation within the Chinese market, due to the success of its creative research platform that efficiency links basic research, R&D of novel products, clinical testing and market entry.

With an outstanding reputation as a high-tech healthcare corporation, PINS Medical has a primary mission for providing innovative, high-quality products and services for patients to improve quality of life. PINS, which stands for Programmable Implanted Neuromodulation Stimulator, is also an abbreviation of "Patient Is No.1 alwayS". This clearly presents the goal of PINS Medical for "restoring hope", not simply as an innovation company but also across society to citizens.

Looking into the future with the continuous rise in incidence of neuropsychiatric disorders and increased social burden across the globe, PINS Medical along with local governments, research centers, companies and top academic scientists, are now developing and promoting innovative therapies worldwide.

www.pinsmedical.com info@pinsmedical.cn

GMP Cytokines

PeproTech's PeproGMP Cytokines help researchers meet the challenges of cell therapies. They are manufactured and tested in compliance with U.S. Food and Drug Administration Good Manufacturing Practices (GMP) regulations and the ISO 9001 quality management systems standard, without the use of animal-derived materials. This line of GMP cytokines is used in some of today's most exciting clinical and investigational efforts, from cell therapy to gene therapy to creating tissue-engineered products. PeproGMP Cytokines are not intended for direct administration to humans.

PeproTech
For info: 800-436-9910
www.peprotech.com/gmp

Flow Reactor

The FlowSyn is a fully integrated, continuous-flow reactor designed for easy, safe, and efficient operation. This two-channel benchtop system combines two high-pressure pump channels and two reactor modules in a single compact unit. Its high-resolution user interface quickly guides you through the process of setting up and running flow chemistry reactions. FlowSyn includes models for performing single or multiple homogeneous or heterogeneous reactions, either manually or automatically. Typical examples of flow chemistry applications include hydrogenation, nitration, bromination, metalation, molecular rearrangements, and synthesis of compounds such as dihydropyridine, indole, pyrazole, quinolinone, and benzimidazole.

Uniqsis
For info: +44-(0)-845-864-7747
www.uniqsis.com

Clinical Genome Analysis Software Platform

Sapientia is a clinical genome analysis software platform that evaluates genome-scale DNA data to produce a comprehensive diagnostic report that can be linked to patients' symptoms, supporting clinical decision-making about rare genetic diseases. The platform is based on pioneering research from the UK's Wellcome Trust Sanger Institute, National Health Service clinicians, and regional genetic testing laboratories. Its underlying technology has been validated by leading independent institutes and clinicians, including Genomics England. Sapientia is also being used for the advancement of personalized medicine by the global pharmaceutical industry to create disease registries, identify patient populations for clinical studies, and support the discovery of novel drug targets and biomarkers.

Congenica
For info: 44-(0)-1223-499-965
www.congenica.com/about-sapientia

Library Prep Solution

Agilent Technologies' SureSelect$^{XT\ HS}$ is a complete research solution that provides total workflow management for laboratories, from quality control to target enrichment, to analysis and interpretation. SureSelect$^{XT\ HS}$ is a streamlined, high-sensitivity system, optimized for labs with a requirement to sequence DNA from formalin-fixed paraffin-embedded (FFPE) samples, which may have degraded over time. By incorporating molecular barcodes to assist error correction, the system improves overall precision and produces higher-complexity libraries than competing products, on a broad range of tissue types and low- and high-quality FFPE samples. SureSelect$^{XT\ HS}$ libraries provide higher-percentage reads in targeted regions and use as little as 10 ng of starting DNA. Additionally, they offer faster, more efficient processing with master-mixed reagents that require less hands-on time, coupled with 90-minute hybridization—now enabling labs to move from sample to sequencer in a single day.

Agilent Technologies
For info: 800-227-9770
www.agilent.com

System Microscope

Olympus' BX53 microscope features an innovative LED light source that produces a light output greater than the 100-watt halogen reference, and does not introduce any color casting, providing a true representation of the sample. Users can clearly identify commonly used dyes in pathology and other life science applications. With a color temperature that stays constant, mimicking the halogen reference lamp, the BX53 saves time by not requiring users to adjust color filters. Its True Color LED brightness makes this the reference platform for multiple viewing head discussion systems—up to 26 observation heads can be connected. Its integrated Light Intensity Manager streamlines observations by immediately adjusting the LED brightness level when a user changes magnification. Users can specify and save their desired preset brightness levels at any time for each objective, and accessories allow the most comfortable positioning for objective change and camera snapshots.

Olympus
For info: 704-877-8801
www.olympus-lifescience.com

ddPCR Genome Edit Detection Assays

ddPCR Genome Edit Detection Assays offer a fast, precise, simple, and cost-effective method for detection of genome editing events. Assays can be obtained for any target using Bio-Rad's easy-to-use, intuitive Digital Assay Site. Droplet Digital PCR (ddPCR) enables rapid assessment of homology-directed repair (HDR) and non-homology end-joining (NHEJ) edits generated by CRISPR/Cas9 or other genome editing tools, detection of editing events present at frequencies of <0.5%, absolute quantification of genome events from as little as 5 ng of total genomic DNA, and distinction between homozygous and heterozygous edits in clonal populations.

Bio-Rad Laboratories
For info: 510-741-1000
www.bio-rad.com/digital-assays

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

Present your work to the world

Are you a representative of the upcoming generation of thought leaders in your field? Together we look forward to your application for the new Sartorius & *Science* Prize for Regenerative Medicine & Cell Therapy.

Apply now!

www.passionforscience.com/prize

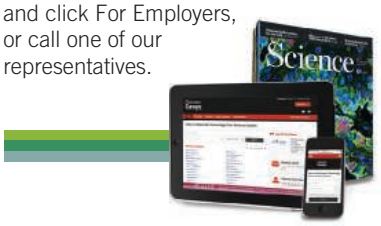
The Sartorius & *Science* Prize for Regenerative Medicine & Cell Therapy

Awarded by sartorius Science

Science Careers

SCIENCE CAREERS ADVERTISING

For full advertising details, go to ScienceCareers.org and click For Employers, or call one of our representatives.



AMERICAS

+1 202 326-6577
+1 202 326-6578
advertise@sciencecareers.org

EUROPE, INDIA, AUSTRALIA, NEW ZEALAND, REST OF WORLD

+44 (0) 1223 326527
advertise@sciencecareers.org

CHINA, KOREA, SINGAPORE, TAIWAN, THAILAND

+86 131 4114 0012
advertise@sciencecareers.org

JAPAN

+81 3-6459-4174
advertise@sciencecareers.org

CUSTOMER SERVICE

AMERICAS

+1 202 326-6577

REST OF WORLD

+44 (0) 1223 326528

advertise@sciencecareers.org

All ads submitted for publication must comply with applicable U.S. and non-U.S. laws. *Science* reserves the right to refuse any advertisement at its sole discretion for any reason, including without limitation for offensive language or inappropriate content, and all advertising is subject to publisher approval. *Science* encourages our readers to alert us to any ads that they feel may be discriminatory or offensive.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

ScienceCareers.org

TECHNISCHE UNIVERSITÄT DARMSTADT

In 2019, the position of

President

(Code. No. 286)

is to be filled at Technische Universität Darmstadt.

Technische Universität Darmstadt is one of Germany's leading technical universities. Its main areas of research, teaching and transfer are engineering and natural sciences, which cooperate closely with selected humanities and social sciences. TU Darmstadt with its approximately 25,000 students is the only technical university in Hessen. It is integrated into the economically strong Rhine/Main/Neckar region and has an effective international network.

We are looking for an academically proven leader with international experience who will develop the profile of TU Darmstadt in a creative, committed and competent manner, taking into account the legally stipulated autonomy. The notable strengths of TU Darmstadt in interdisciplinary cooperation between engineering, natural sciences, humanities and social sciences are to be expanded and there is to be a particular focus on actively shaping the excellence and internationalisation strategy. TU Darmstadt lives a culture of open communication and expects the applicant to be able to integrate and cooperate both internally and externally.

In accordance with the requirements of Section 39 Hessisches Hochschulgesetz [Higher Education Act of the State of Hessen], the President is elected by the University Assembly on the proposal of the University Council. Adequate knowledge of German and a sound knowledge of the German higher education system are expected. The remuneration is negotiated with the Hessen State Ministry for Higher Education, Research and the Arts.

TU Darmstadt expressly welcomes applications from women. Persons with a degree of disability of at least 50 or those of an equal status are given preference if they are of the same suitability.

Applications including substantiating documents must be sent to the Chairman of the Selection Committee, Professor Ferdi Schüth, TU Darmstadt, Referat Qualitätsmanagement und Gremien (IB), Karolinenplatz 5, 64289 Darmstadt or by e-mail to findungskommission@tu-darmstadt.de.

Application deadline: September 1, 2018

UR MEDICINE | WILMOT CANCER INSTITUTE

Postdoctoral Fellow / Staff Scientist Positions in Cancer Epigenomics

The newly established Laboratory of Cancer Genomics in the Wilmot Cancer Institute at the University of Rochester Medical Center is seeking productive and highly motivated individuals interested in cancer epigenetics/genomics. Current studies make use of molecular genetic and genomic approaches to understand epigenetic control of transcription and its role in phenotypic plasticity, intratumor heterogeneity and the emergence of invasive or drug resistant properties. Translational epigenomics studies in human cancers are focusing on defining epigenomic/genomic signatures associated with environmental exposures and patient outcomes with the goal of identifying novel routes for cancer therapy.

Applicants should hold a PhD in the biomedical sciences, computational biology or related discipline with experience in molecular/cancer biology, (epi) genetic/omics, and/or bioinformatics. Wet and/or 'dry' lab experience with epigenomics technology/data (WGBS, ChIP-seq, ATAQ-seq, Pro-seq/RIP-seq), single-cell approaches, RNA biology, gene editing, or cancer models (3D culture, spheroid/organoid) are highly desirable, but applicants with a strong background in a related area, track-record of accomplishment, and desire to work in a multidisciplinary team of investigators focused on cancer epigenetics will be considered. Computational scientists (MS/PhD) with expertise in bioinformatics, systems biology and machine learning approaches are also encouraged to apply. Candidates should have excellent scientific writing and oral communication skills, as well as the ability to work effectively and collaboratively with others. Information about living/working at University of Rochester can be found at https://www.rochester.edu/working/hr/relocation/

Interested applicants should submit a cover letter, CV and contact information for three references to: **Ms. Elva Mikk, Wilmot Cancer Institute, University of Rochester, 601 Elmwood Avenue, Rochester, NY 14642, Box 704** or e-mail at: **elva_mikk@urmc.rochester.edu**.

The University of Rochester is an Equal Opportunity Employer and has a strong commitment to diversity and actively encourages applications from candidates from groups underrepresented in higher education.

By **Tracy Evans**

Heed the call to change

I needed a change. Just a few years earlier, I stood at the edge of the swamp under a massive hollow cypress tree reading *Winnie-the-Pooh* to a group of 5- and 6-year-olds. As a naturalist at a state park, my goal was to introduce these children to the nature found in their own backyard. I played a tape of owl calls to accompany the story. Just then, an owl swooped down over our heads, startling and delighting us. "I can't believe I'm actually getting paid to have this much fun!" I thought to myself. But over time, my job changed. Sitting at my desk, staring at environmental impact reports and grant applications on my computer screen, I began to think, "They cannot pay me enough to do this job." It was another turning point in the winding road that led me, at age 66, to earn a Ph.D.

My first career had been audiology—a pragmatic choice based on job opportunities in a rural area, and one that allowed me to work part time while tending my growing family. But when my children grew up and left home, I found myself longing to be outside, not in a soundproof room in the hospital basement.

I began attending ornithology classes at the local university, hoping that might lead me to new career prospects. I also went to see a career counselor, offered as part of a university program for nontraditional students. "What gave you pleasure as a child?" the counselor asked. I thought for a while before responding, "Playing outside in the woods."

Our discussion eventually led to a second master's degree, this one in zoology, and the job as a naturalist at a state park. It was an entry-level position. I was 44 years old. And I was delighted to once again experience the joy of being outdoors that I had felt as a child.

But as I climbed the career ladder at my state's Department of Natural Resources, from naturalist to grant compliance officer to managing environmental reviews of publicly funded conservation projects, my time in the field diminished and finally disappeared. I was chained to a computer screen. In the back of my mind, the career counselor's question smoldered. "What gave you pleasure as a child?"

As my discontent grew, life went on. At a dinner party, my friends and I discussed the perfect retirement. We agreed that, if money were not an issue, we would each purchase a home in a different part of the world and rotate among them. Shortly thereafter, an opportunity presented itself: a 4-month sabbatical at a university in the Netherlands as part of a field research project. It wasn't quite the house-share plan, but it had the appeal of being someplace new.

"I was 60 years old. ... But I wanted to follow my new passion."

I found that field research allowed me to play in the woods again, but with a new purpose. I was able to combine my childhood delight in nature with the intellectual stimulation that comes with figuring things out—pleasures that had become all too rare in my job.

At the end of my stay, I was invited to participate in the "sandwich" Ph.D. program, in which students do research in their home country while receiving support and instruction in the Netherlands. I was 60 years old. I didn't think that earning a Ph.D. would further my career. But I wanted to follow my new passion.

I figured that I could work full time at my day job in the state government and conduct my research during evenings and weekends. Still, my colleagues raised their eyebrows. "What do you intend to do with a Ph.D.?" they asked. "It is all about the journey," I responded.

I successfully defended my dissertation 6 years later, 3 years after retiring from my government position.

My career path has taken some sharp zigs and zags over my 40 years in the workforce. And each decision I made was the right one for me at the time. But when they were no longer right, I changed direction. No career decision has to be final. As one of my favorite authors, Ursula K. Le Guin, wrote, "It is good to have an end to journey towards, but it is the journey that matters in the end." ■

Tracy Evans is an associate researcher at the Illinois State Museum Research and Collections Center in Springfield.
Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER