

nature

RAIN STOPS GAIN

Excessive wet weather reduces economic growth rate

Coronavirus
What effect has the pandemic had on child development?

Pulling focus
Use non-profit start-ups to unblock the research pipeline

Social medium
Eggshell beads provided long-range links in early Africa

Nature.2022.01.15

[Sat, 15 Jan 2022]

- [This Week](#)
- [News in Focus](#)
- [Opinion](#)
- [Work](#)
- [Research](#)
- [Amendments & Corrections](#)

This Week

- **[COVID is here to stay: countries must decide how to adapt](#)**

[10 January 2022]

Editorial • The Omicron variant has laid bare the need to live with a disease that throws up an ever-changing set of challenges.

- **[Research evaluation needs to change with the times](#)**

[11 January 2022]

Editorial • The focus on a narrow set of metrics leads to a lack of diversity in the types of leader and institution that win funding.

- **[Stamp out fake clinical data by working together](#)**

[11 January 2022]

World View • How can we make sure that medical trials reported in the scientific literature are real? It is surprisingly hard — but not impossible.

- **[The surprising genes behind a fingerprint's unique swirls](#)**

[06 January 2022]

Research Highlight • Genes that control limb development are looped into the process for generating fingerprint patterns.

- **[The tricky ‘hotspot’ volcanoes that belie the name](#)**

[06 January 2022]

Research Highlight • Seismic waves reveal unexpectedly cool temperatures for certain volcanoes classified as hot.

- **[A mighty antibody is the bane of multiple coronaviruses](#)**

[06 January 2022]

Research Highlight • Potent immune molecule could be used to treat COVID-19 and infections caused by viruses related to SARS-CoV-2.

- **[A cloud named Maggie: space streamer could birth new stars](#)**

[04 January 2022]

Research Highlight • Giant hydrogen filament is one of the longest features of its type in the Galaxy.

- **[Tomb reveals warrior women who roamed the ancient Caucasus](#)**

[04 January 2022]

Research Highlight • Dagger and arrows interred in a grave 3,000 years ago provide evidence of female warriors.

- **A ‘chisel’ of light carves solid shapes out of a liquid** [24
December 2021]
Research Highlight • Ultraviolet light controls the emergence of solid particles from a fluid, creating coral- and blossom-shaped structures.
- **Rivers buffer a vast sea against climate-change impacts** [04
January 2022]
Research Highlight • As global temperatures rise, parts of the Gulf of Mexico are undergoing less acidification than expected — for now.

| [Next section](#) | [Main menu](#) |

- EDITORIAL
- 10 January 2022

COVID is here to stay: countries must decide how to adapt

The Omicron variant has laid bare the need to live with a disease that throws up an ever-changing set of challenges.



As we enter the third year of the pandemic, countries need to find ways of living with SARS-CoV-2 while minimizing its harms. Credit: Rodrigo Paiva/Getty

From a pandemic perspective, 2022 seemed poised to begin with a hefty dose of *déjà vu*, with COVID-19 cases on the rise in many countries in the lead-up to the new year. Meanwhile, a new coronavirus variant seemed

poised to overwhelm health-care systems amid fears that vaccines — from first inoculations to boosters, depending on the country — could not be rolled out quickly enough to stem the impending tsunami of infections.

The welcome news that surges of the Omicron variant are associated with less severe disease in adults than are preceding variants of SARS-CoV-2 suggests that some of pandemic modellers' worst-case scenarios will not come to pass. But life has again been disrupted. Widespread absences due to coronavirus infections have left hospitals in many countries understaffed, forced schoolchildren to return to remote learning, and limited global mobility. And even if a relatively small percentage of those infected require hospitalization, sky-high infection rates across large populations mean that many people will still face life-threatening disease and long-term disability. This is particularly true for the unvaccinated — a group that includes a large proportion of the world's population, especially children.

For those who had hoped that 2021 would be the year that put the pandemic in the past tense, it was a harsh reminder that it is still very much present. Rather than laying plans to return to the 'normal' life we knew before the pandemic, 2022 is the year the world must come to terms with the fact that SARS-CoV-2 is here to stay.

Countries must decide how they will live with COVID-19 — and living with COVID-19 does not mean ignoring it. Each region must work out how to balance the deaths, disability and disruption caused by the virus with the financial and societal costs of measures used to try to control the virus, such as mask mandates and business closures. This balance will vary from one place to another, and with time, as more therapies and vaccines become available — and as new variants emerge.

The emergence of the Omicron variant last November highlighted the ongoing challenges of life with SARS-CoV-2. Some countries were already facing surges in the highly transmissible Delta variant, but vaccines and prior infection conferred relatively high levels of protection against Delta, particularly against severe disease. Many researchers — and a fair few politicians — hoped that future waves would be less disruptive, thanks to the build-up of immunity in populations that would keep viral circulation in

check and protect most people from the severe manifestations of disease that drain health-care resources.

It was expected that mutations in the viral genome would slowly chip away at this immunity, particularly its ability to stop viral transmission. But Omicron dealt a swifter and more serious blow to immunity than predicted. It is now clear that SARS-CoV-2 reinfections are more common, and that some of the most widely used COVID-19 vaccines have faltered in the face of the variant. Existing vaccines, developed against an earlier variant, now require a booster to provide substantial levels of protection against infection.

But the news has not all been grim. Vaccines, particularly when boosted, still seem to provide substantial protection against severe disease and death. Early data from animal studies suggest that Omicron might generate a different pathology compared with previous variants, causing greater infection of the upper respiratory tract and [less infection in the lungs](#). Data from several countries suggest that the variant is associated with less severe disease, although whether this is due to the variant itself or widespread pre-existing immunity requires further study.

Countries have charted a variety of courses through the latest surge. Many with the resources have accelerated the distribution of vaccine boosters, but many others do not have this luxury. Some countries have reinstated lockdowns, whereas others are holding back, waiting to see the extent to which climbing infection rates affect hospitals.

With infection rates soaring around the globe and many countries still unable to access adequate vaccine supplies, more SARS-CoV-2 variants of concern will continue to emerge. And, as Omicron has illustrated, the ability to predict what course those variants will take becomes more difficult as the complexities of viral evolution and pre-existing immunity complicate the models that have previously been used to anticipate the course of the pandemic. Now modellers need to factor in the effects of vaccines, previous infections, waning immunity over time, booster shots and viral variants — and, as the year progresses, they will also have to consider the impact of emerging antiviral treatments.

But what is clear is that the hope that vaccines and prior infection could generate herd immunity to COVID-19 — an unlikely possibility from the start — has all but disappeared. It is widely thought that SARS-CoV-2 will become endemic rather than extinct, with vaccines providing protection from severe disease and death, but not eradicating the virus.

As Omicron and other variants have shown, this only adds to the urgency with which vaccines must be distributed to countries that currently lack supplies. Efforts are under way to bolster vaccine production in countries such as South Africa, which have not historically been centres for vaccine manufacturing. These and other efforts to boost global access to vaccines remain in the best interests of all countries: devastating variants are particularly likely to emerge and seed blazing outbreaks in regions with low vaccination rates, and their spread will be further exacerbated where levels of testing and genomic surveillance are also low.

Fortunately, 2022 is poised to add to our defences against the pandemic. New vaccines — such as protein-based vaccines, which might cost less and have less-stringent storage requirements than mRNA vaccines currently do — will become more widely available. In December, the World Health Organization approved the long-awaited protein vaccine made by Novavax in Gaithersburg, Maryland, for emergency use. Ongoing clinical trials will establish whether upcoming vaccine candidates that target specific coronavirus variants, or that can be inhaled or taken orally rather than injected, will also be useful. Several nasal candidates are in clinical testing, including one from CanSino in Tianjin, China, and another developed by AstraZeneca in Cambridge, UK.

Meanwhile, new antiviral drugs, formulated in tablets that can be easily administered early in the course of infection to reduce the chance of serious disease and death, offer another approach against COVID-19. In the past few months, some countries have authorized the use of two such drugs: molnupiravir, made by Merck in Kenilworth, New Jersey, and Ridgeback Biotherapeutics in Miami, Florida; and Paxlovid, made by Pfizer, based in New York. Data from pivotal clinical trials of other candidates are expected in the coming year.

All of these will expand the world's capacity to manage SARS-CoV-2 outbreaks. They are cause for hope and optimism, but with a hefty dose of realism: the virus will continue to circulate and change, and governments must continue to rely on the guidance and advice of scientists. We will not always be able to predict the virus's path, and we must be ready to adapt with it.

Nature **601**, 165 (2022)

doi: <https://doi.org/10.1038/d41586-022-00057-y>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00057-y>.

| [Section menu](#) | [Main menu](#) |

- EDITORIAL
- 11 January 2022

Research evaluation needs to change with the times

The focus on a narrow set of metrics leads to a lack of diversity in the types of leader and institution that win funding.



The European Commission in Brussels wants a Europe-wide agreement on research assessment that recognizes ethics and integrity alongside teamwork and a diversity of outputs. Credit: Bernal Revert/Alamy

Many researchers who are funded from public sources are required to participate in national evaluations of their work. Such assessments are popular with governments because they help to ensure a degree of

accountability for taxpayer cash. Funders like them, too, because they provide a useful benchmark for the standard of research being done. Universities also benefit financially when they write their research strategies around the requirements of assessments. By contrast, [researchers generally see assessments as unhelpful to their work](#). Evaluations can also be stressful and burdensome, and in some cases create tensions between colleagues in academic and administrative roles.

With a few exceptions, the principal components of assessment systems have stayed largely the same since the exercises began, in the 1980s. But some countries are contemplating reworking these systems to reflect how science is done today. Change has been a long time coming, precipitated by initiatives such as the 2013 [San Francisco Declaration on Research Assessment](#), the 2015 [Leiden Manifesto](#) for research metrics and the 2020 [Hong Kong Principles](#) for assessing researchers. Official research assessments are clearly behind the times and need to catch up.

Last November, the European Commission announced plans to put together a European Union-wide agreement on research assessment. It is proposing that assessment criteria reward ethics and integrity, teamwork and a diversity of outputs in addition to research quality and impact. The UK Future Research Assessment Programme, due to report by the end of this year, has also been tasked with proposing ways to ensure that assessments become more inclusive. These changes cannot come soon enough.

Measures of success

Research-assessment systems are the nearest thing that universities have to the performance metrics that are common in business. [Individual researchers are assessed on a range of measures](#), such as the number and quality of journal articles, books and monographs they have published; their research income; the number of their students who complete postgraduate degrees; and any non-academic impact from their work, such as its influence on society or policy. In the United Kingdom, for example, this information is compressed into a composite index and the results are used to allocate funding.

UK public funding goes preferentially to the university departments with the highest-performing researchers. But assessments that measure individual performance make it harder for institutions to recognize science conducted in teams — both within and between disciplines. Moreover, research assessments have tended to focus on final published results, whereas researchers are increasingly producing more diverse outputs, including data sets, reproducibility studies and registered reports, in which researchers publish study designs before starting experiments. Most current assessments do not value mentorship and struggle to recognize the needs of researchers from minority communities.

And then there's the question of costs. The 2014 iteration of the UK Research Excellence Framework — the exercise takes place roughly every seven years — [cost somewhere in the region of £246 million](#) (US\$334 million). The lion's share (£232 million) was borne by universities. It included the costs of academic staff who served on the review panels that assessed around 190,000 outputs in 36 subject areas; and the costs to institutions, which go to great lengths to prepare their staff, including running mock assessment exercises. Here, smaller institutions lack the resources to compete with better-funded ones.

Researchers who study assessment methods regularly put forward ideas for how evaluations could change for the better. Last August, a working group from the International Network of Research Management Societies fleshed out [a framework called SCOPE](#). This encourages funders to design evaluation systems around the ‘values’ they wish to assess. For example, rewarding competitive behaviour might require a different set of criteria from incentivizing collegiality. The SCOPE framework also proposes that funders collaborate with the people being evaluated to design the assessment, and urges them to work with experts in research evaluation — a defined research field.

The importance of co-design cannot be overstated: it will enable the views of different research stakeholders to be represented, and ensure that no single voice dominates. Large, research-intensive institutions often do well in conventional evaluations, because they focus their multi-year strategies on attracting and retaining researchers who meet the criteria of success at publishing results and bringing in income, among other things.

Smaller institutions cannot always compete on these grounds — but could gain if future assessments include new criteria, such as rewarding collaborations, or if assessments put less weight on ability to obtain research funding. A broader range of evaluation criteria could ensure that a greater diversity of institutions have opportunities to do well. And that has to be welcomed.

Larger institutions should not in any way feel threatened by these changes. It is often said — in this journal and elsewhere — that making research culture more welcoming requires systemic change. Research evaluation is core to the research system. If evaluation criteria can be made more representative of how research is done, that much-needed culture change will move one important step closer.

Nature **601**, 166 (2022)

doi: <https://doi.org/10.1038/d41586-022-00056-z>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00056-z>

| [Section menu](#) | [Main menu](#) |

- WORLD VIEW
- 11 January 2022

Stamp out fake clinical data by working together



How can we make sure that medical trials reported in the scientific literature are real? It is surprisingly hard — but not impossible.

- [Lisa Bero](#) 9

In mid-2021, a handful of meta-analyses looked at the use of ivermectin, a drug to treat people infected with parasitic worms, against COVID-19. The analyses included data suggesting it was effective, which came from clinical trials that almost certainly did not happen as described. (Problems detected in reported trials include copied data, results for patients who died before the trial began, and disputes about whether a trial occurred at all.)

A few weeks later, a meta-analysis by Cochrane, an international group that specializes in reviewing evidence in medicine and health, found that

ivermectin had not been shown to be effective ([M. Popp et al. Cochrane Database Syst. Rev. 7, CD015017; 2021](#)). That analysis took several steps to exclude fraudulent studies. But confusion and distrust continue.

In my view, most reported clinical trials are conducted properly, but fraudulent studies are still rife, and not confined to COVID-19. An analysis of trial manuscripts submitted to the journal *Anaesthesia* found that upwards of 40% probably contained false patient data ([J. B. Carlisle Anaesthesia 76, 472–479; 2021](#)).

I study bias in the design, conduct and publication of research. I have been part of a years-long initiative to exclude fraudulent studies from Cochrane's reviews. Detecting fake studies is difficult; so is implementing policies to do so. Solving the problem requires cooperation by publishers, editors, institutions and reviewers.

There is no universal telltale sign of fraud. Data might be fabricated completely or partly. Real data might be represented falsely. (Data collected on 9-year-olds might be reused in a study describing 3-year-olds.) Interventions might not be administered as stated. (Patients described as randomized into treatment groups could have been selected by condition or convenience. Or ‘treated’ patients might not have actually received medicine.) Implausible results might come from egregious misuse of statistics.

All of these practices, even if they are unintentional, produce untrustworthy results. So multiple checks are necessary. Our guide for reviewers at Cochrane includes a range of steps, such as checking registration and consulting databases including PubPeer, a platform for post-publication peer review. Cochrane also asks reviewers to consider the plausibility of claims such as the number of participants studied, given the time and the number of sites and investigators involved. We ask reviewers to look for inconsistencies across the whole article, for overlapping text and for improbable baseline and outcome data.

This is tough work. Some reviewers have suggested that, instead, broad classes of study should be excluded, such as those from certain countries or those that have not been prospectively registered. But that's problematic:

global representation is important, and registration is no guarantee that a study was done appropriately. Furthermore, observational studies, common in public health, are often unregistered.

There are also risks of assuming too readily that a study is fraudulent. Corporate interests sometimes use (false) accusations of fraud to discredit legitimate research that puts them in a bad light, as do many ideological factions. Accusations of fraud can damage reputations and put both accused and whistle-blowers at risk of losing their jobs.

Cochrane's research-integrity team is only two people, working part time — not enough to investigate every possible fraudulent study. That's why we provide tools to help reviewers detect potential fraud, and templates for asking journals for investigations and retractions. When reviewers find a problem, we advise them to get in touch with authors for more information; if there is no timely, reassuring response, we recommend contacting the journal. If systematic reviewers do not raise alarms, it's likely that no one will.

Many reviewers feel this is a waste of time. They assume that most journal editors would rather preserve their reputations than look into concerns, and that any investigation launched by an author's research institution will take years and be biased against finding problems. I understand this. That said, investigations prompted by Cochrane reviewers are under way. In one case, anomalies and unsatisfactory responses from an author group prompted reviewers to check more of the group's studies, exclude those with early warnings of fraud from reviews and present evidence to editors in a clear, structured way. Journals are investigating.

Reviewers and other 'data detectives' do laudable work, but their efforts alone cannot pull a fraudulent trial from the literature. Only a journal can act definitively to retract it.

Facing down fraud requires all involved in the publication pipeline to step up and coordinate. A journal editor has access to the manuscript and data; a publisher can look for problematic patterns across studies (such as [repetitive text or other indications of 'paper mills'](#)); an institution can check

for ethical approval, protocols and raw data that show a study was run as planned.

Universities, journals and publishers should implement data checks sooner rather than later. This means sharing information as well as technical resources, such as expertise in statistical and software tools to detect anomalies. They must also make fraud-detection tasks routine. Too often, investigations focus on pinning down blame, or sweeping misconduct under the rug. Only through a widespread community effort can we ferret out fraud.

Nature **601**, 167 (2022)

doi: <https://doi.org/10.1038/d41586-022-00025-6>

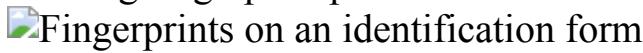
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00025-6>

| [Section menu](#) | [Main menu](#) |

- RESEARCH HIGHLIGHT
- 06 January 2022

The surprising genes behind a fingerprint's unique swirls

Genes that control limb development are looped into the process for generating fingerprint patterns.



The same genes that build an animal's limbs also encode the intricate patterns in fingerprints. Credit: Douglas Sacha/Getty

The arches, loops and whorls that make each person's fingerprints unique are created by some of the same genes that drive limb development¹.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 168 (2022)

doi: <https://doi.org/10.1038/d41586-021-03850-3>

References

1. 1.

Li, J. *et al.* *Cell* <https://doi.org/10.1016/j.cell.2021.12.008> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03850-3>

- RESEARCH HIGHLIGHT
- 06 January 2022

The tricky ‘hotspot’ volcanoes that belie the name

Seismic waves reveal unexpectedly cool temperatures for certain volcanoes classified as hot.



Lava erupts on the Island of Hawai‘i, which was formed by a hotspot volcano. Credit: Mario Tama/Getty

Some hotspot volcanoes aren’t so hot after all.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 168 (2022)

doi: <https://doi.org/10.1038/d41586-022-00002-z>

References

1. 1.

Bao, X., Lithgow-Bertelloni, C. R., Jackson, M. G. & Romanowicz, B. *Science* **375**, 57–61 (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00002-z>

- RESEARCH HIGHLIGHT
- 06 January 2022

A mighty antibody is the bane of multiple coronaviruses

Potent immune molecule could be used to treat COVID-19 and infections caused by viruses related to SARS-CoV-2.



Particles (pink; artificially coloured) of SARS-CoV-2 bud from a cell. The virus and related pathogens are inhibited by an antibody culled from a person who had COVID-19. Credit: Steve Gschmeissner/Science Photo Library

A newly isolated antibody that blocks SARS-CoV-2 from infecting cells could one day be used to treat infections caused by current and future variants of the virus that causes COVID-19, and even infections from related viruses¹.

David Veesler and his colleagues at the University of Washington, Seattle, searched the blood of an infected person for antibodies that bind the SARS-CoV-2 spike protein, which lets the virus enter human cells. They uncovered one particularly potent antibody, called S2K146, that protected cells from infection with the original strain of SARS-CoV-2, as well as the Alpha, Beta, Delta and Kappa variants. (The authors have now shown in a separate study² that S2K146 protects cells from the Omicron variant, too.)

S2K146 also binds spike protein from the two related viruses: SARS-CoV, which caused an outbreak of respiratory disease in 2002–2003, and WIV-1, which infects bats and has the potential to infect humans. Administering

S2K146 to hamsters infected with SARS-CoV-2 greatly reduced or eliminated viral replication.

The team found that mutations that prevented S2K146 from binding to spike protein also rendered SARS-CoV-2 much less effective at infecting cells. This suggests that the virus is unlikely to mutate its way out of S2K146's grasp.

Nature **601**, 168 (2022)

doi: <https://doi.org/10.1038/d41586-022-00001-0>

References

1. 1.

Park, Y-J. *et al. Science* <https://doi.org/10.1126/science.abm8143> (2022).

2. 2.

Cameroni, E. *et al. Nature* <https://doi.org/10.1038/d41586-021-03825-4> (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00001-0>

| [Section menu](#) | [Main menu](#) |

- RESEARCH HIGHLIGHT
- 04 January 2022

A cloud named Maggie: space streamer could birth new stars

Giant hydrogen filament is one of the longest features of its type in the Galaxy.

 False-colour image with a red dotted line showing the 'Maggie' filament

The dashed line shows the location of Maggie, a vast filament (artificially coloured) of hydrogen atoms that stretches some 1,200 parsecs, or nearly 4,000 light years. Credit: J. Syed/MPIA

A long filament-like cloud of hydrogen atoms lurking on the far side of the Milky Way is among the largest such structures in the Galaxy — and offers a rare glimpse into one of the earliest stages of star formation¹.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 168 (2022)

doi: [*https://doi.org/10.1038/d41586-021-03848-x*](https://doi.org/10.1038/d41586-021-03848-x)

References

1. 1.

Syed, J. *et al. Astron. Astrophys.* **657**, A1 (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03848-x>

| [Section menu](#) | [Main menu](#) |

- RESEARCH HIGHLIGHT
- 04 January 2022

Tomb reveals warrior women who roamed the ancient Caucasus

Dagger and arrows interred in a grave 3,000 years ago provide evidence of female warriors.

 Drawings of the women graves of the 13/12-9 centuries BC, with weapons

An artist's impression of two women, buried in ancient times, whose skeletons suggest that they took part in military activity. Credit: A. Y. Khudaverdyan *et al./Int. J. Osteoarchaeol.*

The skeletons of two women who lived some 3,000 years ago in what is now Armenia suggest that they were involved in military battles — probably as horse-riding, arrow-shooting warriors¹.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 169 (2022)

doi: <https://doi.org/10.1038/d41586-021-03828-1>

References

1. 1.

Khudaverdyan, A. Y., Yengibaryan, A. A., Khachatryan, H. H. & Hovhanesyan, A. A. *Int. J. Osteoarchaeol.*
<https://doi.org/10.1002/oa.3077> (2021).

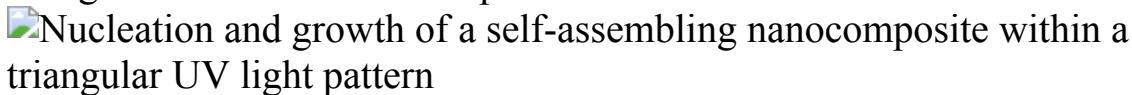
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03828-1>

| [Section menu](#) | [Main menu](#) |

- RESEARCH HIGHLIGHT
- 24 December 2021

A ‘chisel’ of light carves solid shapes out of a liquid

Ultraviolet light controls the emergence of solid particles from a fluid, creating coral- and blossom-shaped structures.



Light is projected onto a triangular section (yellow dashed line) of a solution, triggering the formation of a composite material inside the area. (Video is faster than the actual speed.) Credit: M. H. Bistervels *et al./Adv. Mater.*

An ultraviolet-light beam directs the formation of solid mineral nanoparticles in a carefully tailored solution, sculpting ‘flowers’ and other intricate shapes¹.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 169 (2022)

doi: <https://doi.org/10.1038/d41586-021-03842-3>

References

1. 1.

Bistervels, M. H., Kamp, M., Schoenmakers, H., Brouwer, A. M. & Noorduin, W. L. *Adv. Mater.* <https://doi.org/10.1002/adma.202107843> (2021).

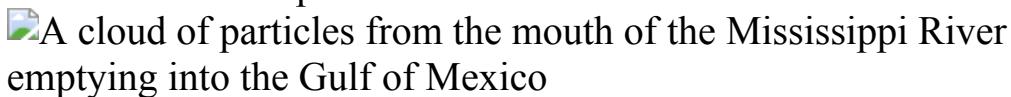
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03842-3>

| [Section menu](#) | [Main menu](#) |

- RESEARCH HIGHLIGHT
- 04 January 2022

Rivers buffer a vast sea against climate-change impacts

As global temperatures rise, parts of the Gulf of Mexico are undergoing less acidification than expected — for now.



A plume of particles marks where the Mississippi River empties into the Gulf of Mexico. The river's alkaline water has slowed the gulf's acidification. Credit: NASA/Alamy

Many parts of the ocean are becoming more acidic as a result of climate change. But in the northern Gulf of Mexico, that effect is nearly cancelled out as increasingly alkaline water pours in from rivers¹.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 169 (2022)

doi: [*https://doi.org/10.1038/d41586-021-03843-2*](https://doi.org/10.1038/d41586-021-03843-2)

References

1. 1.

Gomez, F. A., Wanninkhof, R., Barbero, L. & Lee, S-K. *Geophys. Res. Lett.* **48**, e2021 (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03843-2>

| [Section menu](#) | [Main menu](#) |

News in Focus

- **[Webb telescope success, WHO chief and Omicron versus antibody therapies](#)** [12 January 2022]
News Round-Up • The latest science news, in brief.
- **[Elizabeth Holmes verdict: researchers share lessons for science](#)** [04 January 2022]
News • Theranos case highlights the importance of peer review for biotech entrepreneurs, scientists say.
- **[Researchers fear growing COVID vaccine hesitancy in developing nations](#)** [23 December 2021]
News • Scientists worry that pools of unvaccinated people could be a source of new variants, such as Omicron.
- **[Fatal lab explosion in China highlights wider safety fears](#)**
[22 December 2021]
News • Deaths caused by university lab blasts have some scientists in China concerned about a lack of oversight and standardized safety protocols, especially in teaching labs.
- **[Omicron's feeble attack on the lungs could make it less dangerous](#)** [05 January 2022]
News • Mounting evidence from animal studies suggests that Omicron does not multiply readily in lung tissue, which can be badly damaged in people infected with other variants.
- **[Landmark Colombian bird study repeated to right colonial-era wrongs](#)** [11 January 2022]
News • A re-run of a 100-year-old, US-led bird survey will inform future conservation efforts — but be helmed by local researchers.
- **[The COVID generation: how is the pandemic affecting kids' brains?](#)** [12 January 2022]
News Feature • Child-development researchers are asking whether the pandemic is shaping brains and behaviour.
- **[The devastating mudslides that follow forest fires](#)** [12 January 2022]
News Feature • Regions that never used to burn are now suffering from forest fires — and that raises the risks of dangerous mudslides that are hard to forecast.

| [Next section](#) | [Main menu](#) | [Previous section](#) |

- NEWS ROUND-UP
- 12 January 2022

Webb telescope success, WHO chief and Omicron versus antibody therapies

The latest science news, in brief.



The James Webb Space Telescope, shown in this artist's illustration, successfully unfolded its mirrors and sunshield after launch. Credit: Adriana Manrique Gutierrez/NASA GSFC/CIL

Webb observatory is now officially a telescope

After several tense days of unfurling and clicking its various parts into place, the [biggest and most sophisticated space telescope ever launched is now complete.](#)

On 8 January, NASA's James Webb Space Telescope slowly swung the last 3 of its 18 hexagonal mirror segments into position, locking them together into one 6.5-metre-wide, gold-coated cosmic eye. The move capped an essentially flawless two weeks of engineering manoeuvres since the telescope's 25 December launch.

The US\$10-billion observatory still has many important tasks, such as calibrating its 4 scientific instruments. But it has finished the riskiest engineering moves, without which it would have been inoperable. Those include deploying a kite-shaped, tennis-court-sized sunshield to shade the telescope from the Sun's heat, and positioning its primary and secondary mirrors to capture light from stars, galaxies and other cosmic objects.

Photons are now bouncing between Webb's mirrors, making it an operational observatory. "This is unbelievable," said Bill Ochs, NASA's project manager for Webb, in a 5 January webcast from mission control at the Space Telescope Science Institute in Baltimore, Maryland. "We actually have a telescope."



Tedros Adhanom Ghebreyesus is the only candidate who will run for WHO director-general in May. Credit: Fabrice Coffrini/AFP/Getty

WHO chief Tedros looks guaranteed for re-election amid COVID pandemic

Tedros Adhanom Ghebreyesus, the director-general of the World Health Organization (WHO), is all but ensured to lead the organization for a second term, from 2022 to 2027, because he is the [only candidate in the race](#). As a matter of procedure, on 25 January, the WHO's executive board is expected to nominate him for re-election in May.

The probable extension of his leadership occurs against the backdrop of the ongoing COVID-19 pandemic, as the health agency continues to advise countries on how to curb global devastation.

This isn't the first time a WHO director-general has run for a second term unopposed. Typically, however, several countries propose candidates in the year before an election. Elections are held every five years at the World

Health Assembly in May, an annual meeting of delegates from WHO member states. This year, Tedros is the only candidate, with 28 member states backing him.

China, the United States and about 160 other countries did not nominate anyone. This could be a modest vote of confidence in Tedros, an acknowledgement that a competitor would not prevail or a matter of pandemic practicality, global-health researchers tell *Nature*. Amanda Glassman, executive vice-president at the Center for Global Development in Washington DC, says, “Generally, you don’t want to change leadership in a war.”

Tedros won the election for director-general in 2017, becoming the first African to lead the WHO since its founding in 1948. He quickly made his mark by creating the WHO’s first science division, and by expanding its programme to respond to infectious-disease outbreaks and other health emergencies.

Although he has won admirers since the COVID-19 pandemic began, it was not a given that Tedros would be embraced as a second-term leader. In the first year of the pandemic, several leaders faulted the WHO for not being tougher with China for its lack of transparency in reporting early COVID-19 cases.

But Tedros has also been applauded for helping to persuade the WHO’s 194 member states to move ahead with plans for a pandemic accord, which would dictate how countries respond to future outbreaks.



A nurse in Boston, Massachusetts, gives a dose of the monoclonal antibody bamlanivimab to a man with COVID-19. Credit: Craig F. Walker/The Boston Globe/Getty

Omicron overpowers antibody therapies

Preliminary experiments suggest that [most of the antibody treatments for COVID-19 are powerless against the Omicron variant.](#)

Doctors use artificial versions of natural antibodies to stave off severe COVID-19 in people who are infected with the coronavirus and have a high risk of severe disease or death. But a slew of publications posted on preprint servers report laboratory evidence that Omicron is totally or partially resistant to all currently available treatments based on these monoclonal antibodies ([E. Cameroni et al. Preprint at bioRxiv https://doi.org/hb7x \(2021\); A. Aggarwal et al. Preprint at medRxiv https://doi.org/hb73 \(2021\); D. Planas et al. Preprint at bioRxiv https://doi.org/hb74; 2021](#)). The publications have not yet been peer reviewed, but some of the companies

that manufacture antibody therapies already concede that their products have lower potency against Omicron than against other variants.

The preprints report that only two antibodies show strong evidence of retaining some ability to thwart the variant: sotrovimab, developed by Vir Biotechnology in San Francisco, California, and GSK, headquartered in London; and DXP-604, which is undergoing clinical trials in China and was developed by BeiGene and Singlomics, both based in Beijing.

Nature **601**, 171 (2022)

doi: <https://doi.org/10.1038/d41586-022-00026-5>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00026-5>

| [Section menu](#) | [Main menu](#) |

- NEWS
- 04 January 2022

Elizabeth Holmes verdict: researchers share lessons for science

Theranos case highlights the importance of peer review for biotech entrepreneurs, scientists say.

- [Emily Waltz](#)



Theranos founder Elizabeth Holmes and her lawyer leave the courthouse in San Jose, California, on 3 January.Credit: David Odisho/Getty

Elizabeth Holmes, the infamous biotech chief executive who promised to revolutionize blood testing, has been found guilty of fraud. The Theranos founder intentionally deceived investors, a US federal jury concluded yesterday after a nearly four-month trial. Holmes probably faces up to 20 years in prison and a hefty fine. She has not yet been sentenced.

The case will no doubt shape how biotechnology entrepreneurs approach investors, say researchers who spoke to *Nature* — and it makes abundantly clear the importance of validating early research through peer review.

“It’s a great teaching moment,” says Eleftherios Diamandis, head of clinical biochemistry at Mount Sinai Hospital in Toronto, Canada, who in 2015 [publicly called out Theranos](#) for exaggerating claims. “It’s an example of how a supposedly huge company with a US\$9-billion valuation went down the drain because of a series of mistakes.”

‘A fundamental flaw’

Holmes founded Theranos in 2003 at the age of 19, shortly before dropping out of Stanford University in California. Her goal was to build a company that would make blood tests available directly to consumers. She wanted to eliminate the large needles and tubes of blood required to operate standard diagnostic devices. To that end, she claimed to have developed a machine that could run more than 200 tests on just a few drops of blood taken from a finger prick.

Captivating and ambitious, Holmes attracted more media attention to laboratory diagnostics than ever before. She also brought in high-profile advisers and investors. Former US secretaries of state Henry Kissinger and George Shultz, and former US secretaries of defence James Mattis and William Perry, all joined Theranos’s board of directors. Investors included media mogul Rupert Murdoch and the family of former US education secretary Betsy DeVos.

The company, based in Palo Alto, California, raised about \$945 million and grew to more than 800 employees. It also inked deals with a couple of big retailers. In 2013, pharmacy chain Walgreens began putting Theranos

‘wellness centres’ in its stores in Arizona, eventually setting up 40 sites. The aim was to enable consumers to go to a local pharmacist at their convenience and order a panel of blood tests from a few drops of blood.

Investors and the public believed that Theranos was using its novel machines to analyse the blood samples it received. But in reality, the company could run only a few tests on its platform. The rest went through conventional blood-testing equipment developed by other companies. To meet the specifications of those instruments, the finger-prick samples had to be diluted to increase their volume, and [the results proved unreliable](#).

“There was a fundamental flaw in the idea of getting everything from a drop of blood, because there just aren’t enough molecules there” for detection, says Paul Yager, a diagnostics developer and researcher at the University of Washington in Seattle.

In 2015, Holmes’s ruse began to fall apart. After Diamandis called out Theranos, *The Wall Street Journal* reporter John Carreyrou [exposed the shortcomings of Theranos’s machines](#) in a splashy series of news stories. The US Centers for Medicare and Medicaid Services investigated, and banned Holmes from operating a medical lab for two years. Walgreens sued Theranos. And the US Securities and Exchange Commission charged Holmes and former Theranos president Sunny Balwani with massive fraud, and barred Holmes from serving as a director or officer of a public company for a decade.

On 3 January, a federal court in San Jose, California, found Holmes guilty on 4 of the 11 charges brought against her by prosecutors: 3 for wire fraud — a scheme intended to obtain money under false pretences using a method of remote communication such as e-mail — against investors and one for conspiracy to commit wire fraud against investors. The jury acquitted Holmes on four charges related to defrauding patients and could not reach a decision on three further charges related to defrauding investors.

A hearing next week will address how to handle the three unresolved charges, says Steven Clark, a lawyer in private practice in San Jose and a former district attorney in the region. Balwani will be tried separately in federal court proceedings scheduled to begin in February.

Cautionary tale

The Theranos scandal has provided salacious fodder for books, films and podcasts. But perhaps more importantly, the story has become a cautionary tale for blood-diagnostics companies and scientists with entrepreneurial interests. In particular, it reminds executives at start-up firms to share their data early on, and to participate in some kind of peer-review process, say diagnostics researchers interviewed by *Nature*, who have been observing the Theranos fallout.

“We are definitely going to see more pressure to produce technologies the right way,” says James Nichols, a clinical chemist at Vanderbilt University Medical Center in Nashville, Tennessee. Part of Holmes’s downfall was that she “held Theranos’s technology as proprietary, didn’t publish it and didn’t want to share it with the community”, he says.

Had Holmes participated in peer review, the problems with the technology could have been spotted before she defrauded investors, experts say. That might have forced Holmes to change direction or shut down her company, but it might also have kept her from committing a crime. It’s the self-correction of science that has “saved our butts” time and time again, Yager says.

For Holmes — whose scientific vision drove the company — having a more solid background in the scientific method might have helped, too, researchers suggest. The work required to get a degree in most sciences tends to instil in students the importance of vetting and publishing experiments, Yager says. “You learn that you have to line up your ducks and have credible data.”

The precedent set by the jury’s verdict in the Holmes case might ultimately compel biotech entrepreneurs to be more cautious — and honest — in their approach to investors. The four charges on which Holmes was found guilty centre on how she exaggerated the company’s involvement with potential partners. For example, Holmes added the logos of pharmaceutical giants Pfizer and Schering-Plough to lab reports that she used in her presentations to investors and Walgreens executives. The manoeuvre made it look as if the

pharmaceutical companies had validated Theranos's system, when they had not.

These lessons are likely to inform a generation of researchers, scientists say. "This is probably the biggest story in laboratory medicine, and it ended in disaster," Diamandis says. "The question is: what can we learn about it so that it doesn't happen again?"

Nature **601**, 173-174 (2022)

doi: <https://doi.org/10.1038/d41586-022-00006-9>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00006-9>

| [Section menu](#) | [Main menu](#) |

- NEWS
- 23 December 2021

Researchers fear growing COVID vaccine hesitancy in developing nations

Scientists worry that pools of unvaccinated people could be a source of new variants, such as Omicron.

- [Smriti Mallapaty](#) 0



A dose of vaccine is administered in Kathmandu, Nepal. Credit: Dipendra Rokka/SOPA Images/LightRocket/Getty

COVID-19 vaccine hesitancy has long been recognized as a problem in high- and middle-income nations. But in some of the world's poorest countries, lack of access to vaccines has been a much bigger hurdle. Now researchers say that as doses are trickling in, resistance to getting vaccinated is also emerging as a major issue in those nations.

Scientists fear that persistent pools of unvaccinated people around the world will present a greater risk for the emergence of new variants of concern, such as Omicron. "When you have a lot of community transmission, that's where variants will emerge," says Jeffrey Lazarus, a global health researcher at the Barcelona Institute for Global Health, Spain. Addressing people's hesitancy is therefore crucial, to curb viral spread and to avert hospitalizations and deaths, he says.

Scientists report that hesitancy might now be contributing to the slow uptake of vaccines in some nations where large proportions of the populations remain unvaccinated. These include South Africa — one of the nations where Omicron was first detected — Tanzania, the Democratic Republic of the Congo (DRC), Papua New Guinea and Nepal.

"We have more hesitant people in the global south than we ever thought we did," says Rupali Limaye, a behavioural scientist at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. Although in many countries, limited supply is still the main problem, say researchers.

Adequate doses?

Until late October, the problem in many African nations "was that we just didn't have enough doses", says Salim Abdool Karim, director of the Centre for the AIDS Programme of Research in South Africa (Cape Town) in Durban. "But we now have adequate amounts of vaccines in most countries," he says. According to the Africa Centres for Disease Control and Prevention, just 64% of the vaccines supplied to the continent have so far been administered.

In South Africa, for example, the number of doses administered each week has fallen to less than one-quarter of doses given at the peak of the

vaccination drive in September. This is despite only 44% of adults having been vaccinated with at least one dose.

The calls on social media for more doses from Western countries are perplexing, says Espoir Malembaka, an epidemiologist at the Johns Hopkins Bloomberg School of Public Health, who is based in Bukavu, DRC. Four types of vaccine are now available in eastern DRC, “but we don’t see people really in a rush to get the vaccine”, except for travellers getting ready to board flights, says Malembaka. He believes that the problem is not access to, but mistrust of, the vaccines.

Researchers say that countries might be struggling to get vaccines into arms for many reasons — some of which have nothing to do with hesitancy — including poorly funded health-care systems, the fact that doses are often close to their expiry dates and logistical issues getting vaccines to remote regions. But people’s delay or refusal to get vaccinated is also part of the puzzle.

Global hesitancy

Studies have attempted to estimate the extent of hesitancy worldwide. One [survey¹](#) of nearly 45,000 participants in 12 countries — conducted before COVID-19 vaccines started being rolled out, and published in July — found that hesitancy was lower in the 10 low- and middle-income nations than in Russia and the United States. But researchers say that the situation has changed throughout the pandemic. In Nepal, for example, where the study found acceptance was highest (97%), the pace of vaccination has slowed, despite only 40% of adults having received one dose.

Another survey² of almost 27,000 people in 32 countries conducted from October to December 2020 found that people’s intentions varied considerably, with high levels of hesitancy in some developing nations. At the extreme, 43% of respondents in Lebanon said they definitely would not be vaccinated.

Another survey of a few thousand people observed even [lower acceptance in Papua New Guinea](#), where only 3% of people have been vaccinated.

Researchers found that more than 80% were not planning on getting vaccinated or were uncertain.

Similar but different

Some reasons for hesitancy are shared globally, but there are also local differences. A major concern is safety, especially because the vaccines were developed and delivered rapidly and the recommendations for their use have often changed, say researchers.

Trust in governments is a related concern. The 32-country survey found that a belief that a government was handling the pandemic well was associated with higher acceptance of vaccines. Another analysis³ found that increased trust in medical and scientific authorities made people more likely to agree to vaccination.

The spread of inaccurate information has also hampered roll-outs in some developing nations. “Misinformation in many places is outpacing evidence-based information,” says Limaye.

But local differences also influence people’s views. In the eastern DRC, for example, decades of war and devastating Ebola outbreaks have fuelled distrust in the leadership and in products from the West, says Malembaka. He also found, in a recent unpublished survey, that COVID-19 vaccine hesitancy might have spilled over to affect people’s willingness to accept other vaccines.

Global vaccine inequity might have also contributed to hesitancy, because of “how we’re distributing vaccines to the global south”, says Limaye. “It’s sort of like — here’s our leftovers, they expire in a week.”

The long wait for doses could have encouraged hesitancy, adds Kaushik Ramaiya, a physician-researcher and chief executive of Shree Hindu Mandal Hospital in Dar es Salaam, Tanzania. People start to wonder if they even need to be vaccinated if they have so far avoided becoming infected, says Ramaiya.

Vaccine appeal

There are ways to overcome hesitancy, say researchers. Abdool Karim argues that South Africa has hit the point at which people need incentives or even government mandates to get vaccinated.

In a recent unpublished survey of people in a number of countries, Lazarus and his colleagues found that mandates — such as requiring vaccination to travel by air or attend a workplace — could help to sway decisions. It found that, among people who were hesitant about COVID-19 vaccines, one-third said they would get vaccinated if they had to do so that they could travel internationally.

Patrick Mdletshe, who heads the community programmes at CAPRISA, says the South African government should take lessons from HIV epidemic and engage directly with communities to persuade them, instead of investing in mass-media campaigns.

Opportunities to be vaccinated should also be integrated into existing services for the treatment and prevention of other infectious diseases, such as tuberculosis screening or distribution of HIV medications, which are accessible and familiar to people. “COVID-19 mustn’t be a stand-alone issue,” he says.

Nature **601**, 174–175 (2022)

doi: <https://doi.org/10.1038/d41586-021-03830-7>

References

1. 1.

Solís Arce, J. S. *et al. Nature Med.* **27**, 1385–1394 (2021).

2. 2.

de Figueiredo, A. & Larson, H. J. *Commun. Med.* **1**, 30 (2021).

3. 3.

Kerr, J. *et al.* *BMJ Open*. **11**, e048025 (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03830-7>

| [Section menu](#) | [Main menu](#) |

- NEWS
- 22 December 2021

Fatal lab explosion in China highlights wider safety fears

Deaths caused by university lab blasts have some scientists in China concerned about a lack of oversight and standardized safety protocols, especially in teaching labs.

- [Andrew Silver](#)



Firemen at the site of an explosion at Beijing Jiaotong University in 2018.Credit: Nicolas Asfouri/AFP/Getty

The deaths of two people following a laboratory explosion at a Chinese university in October have raised alarm among researchers. The full circumstances that led to the deaths at the Nanjing University of Aeronautics and Astronautics (NUAA), in Jiangsu province, are not yet known — but they come amid wider concerns about safety in university teaching labs in China.

The deaths are the latest in a series of fatalities caused by explosions in academic laboratories in China, often involving students in chemistry departments, that have been reported in recent years. Some researchers are optimistic that the situation is improving. But others say China's government needs to do more to improve safety.

In the recent incident, nine people were injured and two died as a result of an explosion just before 4 p.m. in the NUAA's College of Materials Science and Technology, according to a [24 October post by the university](#) on the social-media network Weibo. The NUAA said this week that an investigation was ongoing, but declined to answer *Nature*'s questions regarding the safety of its laboratories.

Earlier this year, on 31 March, a graduate student was killed following an explosion at the Institute of Chemistry of the Chinese Academy of Sciences in Beijing. Previous lab blasts led to deaths of three students conducting a sewage-treatment experiment at Beijing Jiaotong University in December 2018; one death [in the chemistry department at Tsinghua University](#) in Beijing in December 2015; and one death in a chemistry lab at the China University of Mining and Technology in Xuzhou in April 2015.

It's not possible to say what was the cause of any individual explosion and death without a full investigation report — none of which have been made public except for the incident at Beijing Jiaotong University — and some may not have been caused by negligence or lack of safety procedures.

‘Renewed safety concerns’

But a study¹ that focused on safety in graduate and undergraduate teaching labs, and was published online last month, claims that “the past two decades

have seen a rise in university laboratory accidents in China”.

Led by researchers at the China University of Petroleum in Qingdao, the team looked at 110 publicly reported lab accidents in China between 2000 and 2018, finding that they had led to 102 injuries and 10 fatalities. But researchers think that the true number of accidents is likely to be significantly under-reported, because it’s likely some incidents are never made public.

The authors of the study note that the number of graduate students enrolled in laboratory-related disciplines in China ballooned from 90,000 in 2000 to about 5.3 million in 2019 — and the number of labs grew along with them, which they say may be part of the reason accidents have increased in number.

After the Nanjing blast, Chinese state news outlet *Global Times* said the incident had triggered “renewed safety concerns”, adding that “Chinese chemists have been calling for improvements to lab safety at research institutions following previous incidents, which reflect systematic negligence of safety.”

Saad Javed, a systems scientist at Nanjing University of Information Science and Technology who has studied and worked at the NUAA, worries that little has been learnt from the string of accidents at Chinese universities, and that more needs to be done to improve university governance to prevent future incidents.



Nanjing University of Aeronautics and Astronautics was the site of a laboratory explosion in October. Credit: Yang Bo/China News Service/Getty

Nature asked the Chinese Academy of Sciences, Beijing Jiaotong University, Tsinghua University and the China University of Mining and Technology if problems with safety may have been linked to the incidents at their institutions and whether there is a wider problem with safety in university labs in China, but all either declined to answer or did not respond.

Very few countries keep detailed records on numbers of university lab accidents. But in many countries deaths seem to be rare. In the United States, for example, the Chemical Safety and Hazard Investigation Board listed only [one fatality](#) related to a chemical incident at a university between 2001 and mid-2018. According to the Laboratory Safety Institute, which maintains an informal [online memorial wall](#) for researchers killed in labs, Germany and France each had one fatality from a lab incident over the same period, and the United Kingdom had none.

Safety is an issue globally

However, China has far more students than many other countries, so it's difficult to say whether it has a larger number of lab-related deaths per capita, and researchers note that lab safety is an issue in many countries. Some also question the idea that incidents are becoming more frequent in China, and argue that serious accidents were probably happening 15–20 years ago as well, but they just weren't reported publicly before the rise of the Internet and social media.

Still other researchers in China who spoke to *Nature* say that some universities don't place enough emphasis on lab experience and training for students, and some senior staff may even not have had enough safety training. Some institutions "don't produce students with basic practical skills required to work in any research lab with hazards", says one experimental physicist in China, who didn't want to be named.

Yang Guang-Fu, a chemist who studies pesticides at Central China Normal University in Wuhan, agrees that there is a severe shortage of staff with safety expertise, that rules are poorly enforced and that the administrations of some Chinese universities don't attach enough importance to safety.

Jason Chruma, an organic chemist at the University of Virginia in Charlottesville, who was a professor and then an assistant dean at Sichuan University in Chengdu from 2012 to 2020, says that some government safety regulations are not clear, are interpreted differently from place to place and are difficult for the government to enforce.

Although he can't comment on current practices, Chruma says that during his time at Sichuan University he sometimes saw safety issues at first hand, such as students conducting chemical reactions in hallways because there weren't enough fume hoods in busy labs.

Sichuan University didn't respond to *Nature*'s questions regarding the safety of its laboratories.

According to the November study, the 2018 explosion at Beijing Jiaotong University — which led to the deaths of three students and destroyed an entire laboratory — was caused by the ignition of 66 kilograms of improperly stored magnesium dust. The authors report that this was the first

university lab accident in China for which a detailed investigation was conducted and the results were reported on a government website.



Security guards cordon off the building where a blast occurred in a chemistry lab at Tsinghua University, Beijing, in 2015. Credit: Chen Yehua/Xinhua/Alamy

Optimism on improvements

Some researchers are optimistic that the safety situation is improving, as it has in many countries in recent decades.

“Compared with 20 years ago, lab safety in China has definitely made significant progress,” says Samuel Yu, director of the Health, Safety and Environment Office at the Hong Kong University of Science and Technology. He says big improvements include the addition of safety features such as fume cupboards and emergency showers. This “is most obvious in the leading universities”, he says.

Denis Simon, who was executive vice-chancellor of Duke Kunshan University in Jiangsu province from 2015 to 2020, agrees that “this is a country that has come a tremendously long way in improving the protocols”. But he says that China needs more specialists with a career focus on lab safety, and that grant awards should require certification of lab safety protocols for funded projects.

However, Yu doubts there is a simple solution to the problem in China. “The fatal lab accidents that we are aware of usually have multiple faults adding up to the tragic outcome, so I don’t think there is a single magic bullet,” he says.

The concerns over university teaching labs in China follow a debate over whether the COVID-19 pandemic might have started in a lab in China. But Gigi Gronvall, a biosecurity specialist at the Bloomberg School of Public Health at Johns Hopkins University in Baltimore, Maryland, who has visited labs in China, says that there is no reason to think that the problems flagged in student chemistry labs in China reveal anything about the biological-research labs that handle dangerous pathogens.

“I wouldn’t necessarily assume that those same problems are occurring with biological organisms,” she says. Even the kind of non-teaching biological-research labs that do quite advanced experiments don’t use dangerous pathogenic organisms, which typically require extensive safety training and follow more stringent safety standards.

Nature **601**, 176-177 (2022)

doi: <https://doi.org/10.1038/d41586-021-03589-x>

Additional reporting by Smriti Mallapaty.

References

1. 1.

Bai, M. *et al. J. Loss Prev. Process Ind.* **74**, 104671 (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03589-x>

| [Section menu](#) | [Main menu](#) |

- NEWS
- 05 January 2022
- Correction [06 January 2022](#)

Omicron's feeble attack on the lungs could make it less dangerous

Mounting evidence from animal studies suggests that Omicron does not multiply readily in lung tissue, which can be badly damaged in people infected with other variants.

- [Max Kozlov](#)



A doctor in Uzhhorod, Ukraine, examines the lungs of a person with COVID-19. The Omicron variant might affect the lungs less than do

previously circulating variants.Credit: Serhii Hudak/Ukrinform/Barcroft Media via Getty

Early indications from South Africa and the United Kingdom signal that the fast-spreading Omicron variant of the coronavirus SARS-CoV-2 [is less dangerous](#) than its predecessor Delta. Now, a series of laboratory studies offers a tantalizing explanation for the difference: Omicron does not infect cells deep in the lung as readily as it does those in the upper airways.

“It’s a very attractive observation that might explain what we see in patients,” says Melanie Ott, a virologist at the Gladstone Institute of Virology in San Francisco, California, who was not involved in the research. But she adds that Omicron’s hyper-transmissibility means that hospitals are filling quickly — despite any decrease in the severity of the disease it causes.

Authorities in South Africa announced on 30 December that the country had passed its Omicron peak without a major spike in deaths. And a 31 December UK government [report](#) said that people in England who were infected with Omicron were about half as likely to require hospitalization or emergency care as were those infected with Delta.

But the number of people who have gained immune protection against COVID-19 through vaccination, infection or [both](#) has grown over time, making it difficult to determine whether Omicron intrinsically causes milder disease than earlier variants. For answers, researchers have turned to animals and to cells in laboratory dishes.

Michael Diamond, a virologist at Washington University in St. Louis, Missouri, and his colleagues infected hamsters and mice with Omicron and other variants to track disease progression. The differences were staggering: after a few days, the concentration of virus in the lungs of animals infected with Omicron was at least ten times lower than that in rodents infected with other variants¹. Other teams have also noted that compared with previous variants, Omicron is found at reduced levels in lung tissue^{2,3}.

Diamond says he was especially shocked to see that the Omicron-infected animals nearly maintained their body weight, whereas the others quickly lost

weight — a sign that their infections were causing severe disease. “Every strain of SARS-CoV-2 has infected hamsters very easily, to high levels,” he says, “and it’s clear that this one is different for hamsters.” The lungs are where the coronavirus does much of its damage, and lung infection can set off an inflammatory immune response that ravages infected and uninfected cells alike, leading to tissue scarring and oxygen deprivation. Fewer infected lung cells could mean milder illness.

Another group found that Omicron is much less successful than previous variants at infecting lung cells and miniature lung models called [organoids](#)⁴. These experiments also identified a plausible player in the difference: a protein called TMPRSS2, which protrudes from the surfaces of many cells in the lungs and other organs, but is notably absent from the surfaces of most nose and throat cells.

[Previous variants have exploited this protein to infect cells](#), but the researchers noticed that Omicron doesn’t bind to TMPRSS2 so well. Instead, it tends to enter cells when it is ingested by them^{5,6}.

Upper airway preferred

Difficulty entering lung cells could help to explain why Omicron does better in the upper airways than in the lungs, says Ravindra Gupta, a virologist at the University of Cambridge, UK, who co-authored one of the TMPRSS2 studies⁴. This theory could also explain why, [by some estimates](#), Omicron is nearly as transmissible as measles, which is the benchmark for high transmissibility, says Diamond. If the variant lingers in the upper airways, viral particles might find it easy to hitch a ride on material expelled from the nose and mouth, allowing the virus to find new hosts, says Gupta. Other data provide direct evidence that Omicron replicates more readily in the upper airways than in the lungs^{2,5}.

The latest results could mean that “the virus establishes a very local infection in the upper airways and has less chance to go and wreak havoc in the lungs”, Ott says. That would be welcome news — but [a host’s immune response](#) plays an important part in disease severity, and scientists need more

clinical data if they are to understand how Omicron's basic biology influences its disease progression in humans.

Omicron's course of infection could also have implications for [children](#), says Audrey John, a specialist in paediatric infectious disease at the Children's Hospital of Philadelphia in Pennsylvania. Young children have relatively small nasal passages, and babies breathe only through their noses. Such factors can make upper respiratory conditions more serious for children than for adults, John says. But she adds that she has not seen data suggesting an uptick in the numbers of young children hospitalized for croup and other conditions that could indicate a severe infection of the upper respiratory tract.

Although there is still much to learn about the new variant, Gupta says that [fears raised in late November](#) by the multitude of mutations in Omicron's genome have not been completely borne out. He says the initial alarm offers a cautionary tale: it's difficult to predict how a virus will infect organisms from its genetic sequence alone.

Nature **601**, 177 (2022)

doi: <https://doi.org/10.1038/d41586-022-00007-8>

Updates & Corrections

- **Correction 06 January 2022:** An earlier version of this story incorrectly described how Omicron enters cells. The story has now been corrected.

References

1. 1.

Diamond, M. *et al.* Preprint at Research Square
<https://doi.org/10.21203/rs.3.rs-1211792/v1> (2021).

2. 2.

McMahan, K. *et al.* Preprint at bioRxiv
<https://doi.org/10.1101/2022.01.02.474743> (2022).

3. 3.

Bentley, E. G. *et al.* Preprint at bioRxiv
<https://doi.org/10.1101/2021.12.26.474085> (2021).

4. 4.

Meng, B. *et al.* Preprint at bioRxiv
<https://doi.org/10.1101/2021.12.17.473248> (2021).

5. 5.

Peacock, T. P. *et al.* Preprint at bioRxiv
<https://doi.org/10.1101/2021.12.31.474653> (2022).

6. 6.

Willett, B. J. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2022.01.03.21268111> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00007-8>

| [Section menu](#) | [Main menu](#) |

- NEWS
- 11 January 2022

Landmark Colombian bird study repeated to right colonial-era wrongs

A re-run of a 100-year-old, US-led bird survey will inform future conservation efforts — but be helmed by local researchers.

- [Luke Taylor](#)



Ornithologist Andrés Cuervo takes a selfie of a team of Colombia Resurvey Project researchers during an expedition in Caquetá. Credit: Andrés M. Cuervo

Colombia has more animal and plant species per square kilometre than anywhere else in the world. Pioneering US bird scientist Frank Chapman once said that the country was so rich in biodiversity that when his research team explored the area in the early twentieth century, it could have studied a single mountain range for five years and still not have mapped all of its fauna.

More than 100 years later, Colombian researchers are redoing his legendary bird survey, which is a reference for ornithologists the world over. They are surveying the areas that Chapman catalogued between 1911 and 1915, to investigate how a century of war, global warming and industrialization has affected the landscape and its biodiversity.

But this project will not snatch birds and whisk them to a museum abroad — as Chapman's team did. Instead, local scientists will keep specimens in Colombia and engage with local communities during their expeditions, to include them in the momentous endeavour, improve the quality of the research and set an ethical standard for future fieldwork.

Chapman and at least 5 other collectors shot many of the nearly 16,000 birds that they hauled back to the American Museum of Natural History in New York City, offering local residents little explanation — or credit. “You wouldn’t like it if I came to your house, surveyed it without permission, took photos and then went back to Colombia without telling you what I had found,” says Nelsy Niño-Rodríguez, the Colombia Resurvey Project’s community-relations coordinator, who is an ornithologist at the Alexander von Humboldt Biological Resources Research Institute in Bogotá.

Without local guides knowledgeable about Colombia and its birds, Chapman couldn’t possibly have located and collected so many specimens, says Natalia Ocampo-Peñuela, a research partner on the resurvey project and a conservation ecologist at the University of California, Santa Cruz. Yet Chapman’s logs hardly mention guides; when they are discussed, it’s usually in racist or pejorative terms, she says.

“His interest was to feed his curiosity, his scientific intellect and the museum,” she adds, but not to inform the wider population — and definitely not the local populations.

A changed landscape

Colombian researchers have dreamt of re-running Chapman’s expeditions for decades. But it wasn’t possible until the past few years, because many areas were inaccessible owing to armed conflict. Following a landmark peace deal in 2016, remote regions that had been under the control of the Revolutionary Armed Forces of Colombia (FARC), a left-wing guerrilla group, [once again opened to exploration](#). That, and an infusion of funding from the Colombian government and international donors, meant researchers could attempt a resurvey.



Birds of Colombia: top, many-banded araçari (*Pteroglossus pluricinctus*); left, pileated finch (*Coryphospingus pileatus*); right, white-fringed antwren (*Formicivora grisea*). Credit: Andrés M. Cuervo

Chapman visited Colombia because he thought that its geography made it one of the most biodiverse places in the world. He theorized that the presence of the Andes Mountains, combined with the country's position bridging South and Central America, made it an evolutionary melting pot.

Although Colombia is still home to around 10% of the world's biodiversity, the forests once explored by Chapman have changed immensely. Pristine

jungles have been cleared to create uniform pastures resembling golf courses, says Andrés Cuervo, an ornithologist at the National University of Colombia in Bogotá who is one of the directors of the resurvey project. The dirt tracks that Chapman and his team traversed on mules are now roads. And climate change has pushed birds to higher elevations and altered their migratory patterns.

Seeking to understand the effects of these changes on biodiversity, researchers launched the Colombia Resurvey Project in 2019. The main objective is to gather bird specimens, including DNA and tissue samples, to compare the modern population with Chapman's collection. The team, which includes US researchers as well as local ones, has so far conducted 6 expeditions, visiting 14 of Chapman's original sites — leaving 60 to go.

A useful catalogue

The researchers are finding that they have to venture deep into the forest to find birds that were once a stone's throw from Chapman's campsites, Cuervo says. And some species are nowhere to be found, including the red-ruffed fruitcrow (*Pyroderus scutatus*) — almost certainly lost when the trees in its territory were cut down to grow avocados, he adds.



Resurvey project researchers Jessica Diaz (right) and Andrés Sierra (left) record data from a mist net, used to collect birds during expeditions. Credit: Andrés M. Cuervo

The team has also confirmed that birds dependent on unique ecosystems are being replaced by generalist species — which are more adaptable to fragmented forest and a disrupted diet — reducing the country's biodiversity¹. Larger species and fruit eaters seem to have been hit particularly hard over the past century, because they require vast expanses of forest to thrive.

The effects of climate and landscape changes on bird populations in the tropics are not well understood, so the project will inform future conservation efforts, researchers say. “It’s almost impossible to imagine all the ways in which this data can potentially be used down the road,” says John Bates, curator and head of life sciences at the Field Museum of Natural History in Chicago, Illinois.

Members of the resurvey project hope their catalogue will have as much impact as Chapman's. It will include resources such as a genomic map illustrating birds' evolution, generated from DNA samples.

"We are collecting the most complete set of specimens that one can imagine so that scientists from now and the future can answer questions that we haven't thought of," Ocampo says.

Taking charge

The Colombia Resurvey Project team especially hopes that its anti-colonial approach will resonate with the scientific community. The researchers run workshops before each excursion to inform local communities about why they are planning to kill some birds, and how this is important for conservation and science. They are storing the specimens at the National University of Colombia, where the birds will be digitally catalogued, so that people can view them online, listen to audio of their song and scroll through interactive maps of the expeditions. And the team is creating birdwatching tours at the expedition sites to boost tourism.

Involving communities leads to better results, Niño-Rodríguez says. For instance, even if some Indigenous people do not know the scientific names for birds, they might be able to identify them on sight and know where they are most likely to be found. And community knowledge of how the forests have changed has passed from generation to generation, so local residents are able to fill gaps when satellite data and research logs aren't available.

It's equally important to the researchers that those leading the project are from Colombia. They say it's common for local experts to help visiting foreign researchers to find new species and make discoveries, but be excluded from the scientific process and the credit. "We don't want to be the guys with the permits or the guys who facilitate the logistics of someone else's research," Cuervo says. "We want to do our own high-quality research, and we want it to be available for people to use."

This time around, the American Museum of Natural History is a partner on the project, rather than its lead. "Although the Chapman expedition was

conducted with help and permissions from the Colombian government, today's expeditions appropriately look much different than they did in Chapman's time," says a museum spokesperson, adding that the museum "is proud of the very active relationship it maintains with Colombia's scientific institutions through education and research".

Meanwhile, project researchers are training curious members of local communities in how to identify birds scientifically, so they can continue to log species with their cameras and mobile phones once the researchers leave the forest. Areas previously ruled by FARC guerrillas are now falling under the control of other armed groups, which might not let outsiders in, so local residents could soon be the only people who have access to some of Colombia's most biodiverse jungles and the birds that inhabit them.

"Hopefully we won't have to wait another hundred years for scientists to return to these sites and assess their bird fauna," Cuervo says. "Communities can do it with empowerment and interest in their biodiversity and surroundings."

Nature **601**, 178-179 (2022)

doi: <https://doi.org/10.1038/d41586-021-03527-x>

References

1. 1.

Gómez, C., Tenorio, E. A. & Cadena, C. D. *Conserv. Biol.* **35**, 1552–1563 (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03527-x>

- NEWS FEATURE
- 12 January 2022

The COVID generation: how is the pandemic affecting kids' brains?

Child-development researchers are asking whether the pandemic is shaping brains and behaviour.

- [Melinda Wenner Moyer](#) ⁰



A teacher in a biosecurity suit gives a lesson to a girl in her home in Cali, Colombia. Credit: Luis Robayo/AFP/Getty

Like many paediatricians, Dani Dumitriu braced herself for the impact of the SARS-CoV-2 coronavirus when it first surged in her wards. She was relieved when most newborn babies at her hospital who had been exposed to COVID-19 seemed to do just fine. Knowledge of the effects of Zika and other viruses that can cause birth defects meant that doctors were looking out for problems.

But hints of a more subtle and insidious trend followed close behind. Dumitriu and her team at the New York–Presbyterian Morgan Stanley Children's Hospital in New York City had more than two years of data on infant development — since late 2017, they had been analysing the communication and motor skills of babies up to six months old. Dumitriu thought it would be interesting to compare the results from babies born before and during the pandemic. She asked her colleague Morgan Firestein, a postdoctoral researcher at Columbia University in New York City, to assess whether there were neurodevelopmental differences between the two groups.

A few days later, Firestein called Dumitriu in a panic. “She was like, ‘We’re in a crisis, I don’t know what to do, because we not only have an effect of a pandemic, but it’s a significant one,’” Dumitriu recalled. She was up most of that night, poring over the data. The infants born during the pandemic scored lower, on average, on tests of gross motor, fine motor and communication skills compared with those born before it (both groups were assessed by their parents using an established questionnaire)¹. It didn’t matter whether their birth parent had been infected with the virus or not; there seemed to be something about the environment of the pandemic itself.

Dumitriu was stunned. “We were like, oh, my God,” she recalled. “We’re talking about hundreds of millions of babies.”

Although children have generally fared well when infected with SARS-CoV-2, preliminary research suggests that pandemic-related stress during pregnancy could be negatively affecting fetal brain development in some children. Moreover, frazzled parents and carers might be interacting differently or less with their young children in ways that could affect a child’s physical and mental abilities.

Lockdowns — which have been crucial for controlling the spread of the coronavirus — have isolated many young families, robbing them of playtime and social interactions. Stressed out and stretched thin, many carers also haven’t been able to provide the one-to-one time that babies and toddlers need.

“Everyone wants to document how this is impacting child development, and parent-child relationships and peer relationships,” says James Griffin, chief of the Child Development and Behavior Branch at the Eunice Kennedy Shriver National Institute of Child Health and Human Development in Bethesda, Maryland. “Everyone has concerns.”

Some of the teams looking into these issues around the world are starting to publish their findings. New studies have begun. Firm answers are hard to come by, not least because many child-development research laboratories shut down during the pandemic.

Some babies born during the past two years might be experiencing developmental delays, whereas others might have thrived, if carers were at home for extended periods and there were more opportunities for siblings to interact. As with many aspects of health during the pandemic, social and economic disparities have a clear role in who is affected the most. Early data suggest that the use of masks has not negatively affected children's emotional development. But prenatal stress might contribute to some changes in brain connectivity. The picture is evolving and many studies have not yet been peer reviewed.

Some researchers propose that many of the children falling behind in development will be able to catch up without lasting effects. "I do not expect that we're going to find that there's a generation that has been injured by this pandemic," says Moriah Thomason, a child and adolescent psychologist at the New York University Grossman School of Medicine.

A precipitous drop in play

One lab that managed to stay open during the COVID-19 pandemic was Brown University's Advanced Baby Imaging Lab in Providence, Rhode Island. In it, Sean Deoni, a medical biophysicist, and his colleagues use magnetic resonance imaging (MRI) and other techniques to study how environmental factors shape brain development in infants.

Although the pandemic changed how they conducted their research — fewer visitors and more cleaning — they continued inviting babies to their lab, to track motor, visual and language skills as part of a seven-year National Institutes of Health study on early childhood development and its effects on later health.

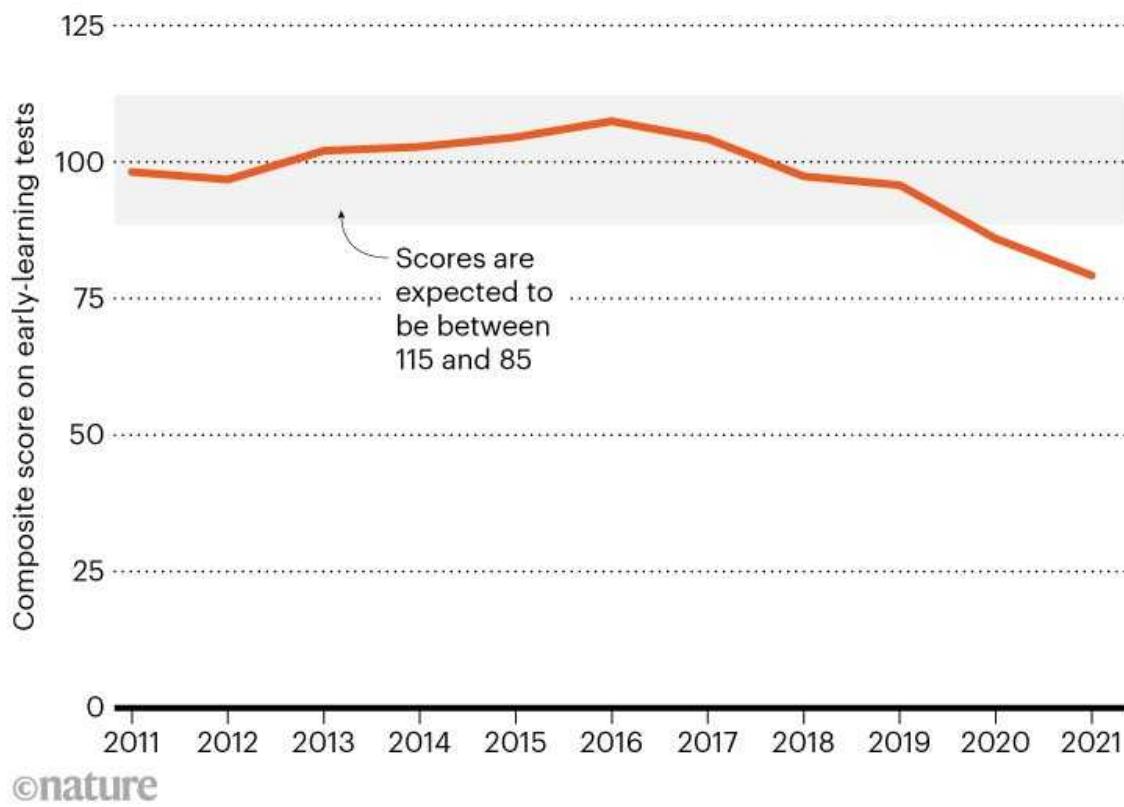
However, as the pandemic progressed, Deoni began hearing worrying comments from his colleagues. "What our staff began to tell me, anecdotally, was 'Man, it's taking these kids a lot longer to get through these assessments,'" Deoni recalled.

He was mystified, so asked his researchers to plot and compare the yearly averages and variances from the infants' neurodevelopmental scores. That's

when they discovered that the scores during the pandemic were much worse than those from previous years (see ‘Development dip’). “Things just began sort of falling off a rock the tail end of last year and the beginning part of this year,” he said in late 2021. When they compared results across participants, the pandemic-born babies scored almost two standard deviations lower than those born before it on a suite of tests that measure development in a similar way to IQ tests. They also found that babies from low-income families experienced the largest drops, that boys were more affected than girls² and that gross motor skills were affected the most.

DEVELOPMENT DIP

Researchers tested the cognitive performance of more than 600 children aged 3 months to 3 years, including 39 babies born during the COVID-19 pandemic. On average, those assessed in 2020 and 2021 scored lower on tests of early learning, including language, puzzle-solving and motor skills such as standing and walking.



Source: Ref. 2

At first, Deoni assumed that selection bias was at play: perhaps the families who made the effort to come in for testing during the pandemic were those whose children were at risk of developmental problems or were already showing them. But, over time, he grew convinced that selection bias wasn't explaining the findings, because the children coming in did not have different backgrounds, birth outcomes or socio-economic statuses compared with previous participants.

These effects appeared drastic, but some researchers argue that they are not necessarily predictive of long-term problems. "IQ, as babies, doesn't predict much," says Marion van den Heuvel, a developmental neuropsychologist at Tilburg University in the Netherlands. "It's really hard to say anything about what that will mean for their future." She points to a study³ showing that Romanian girls who started life in orphanages but were then adopted by foster families before 2.5 years of age were less likely to have psychiatric problems at 4.5 years of age than were girls who remained in institutional care. That situation is different from a pandemic, but suggests that babies could make up for hardship once restrictions are lifted.

Worryingly, however, Deoni has found that the longer the pandemic has continued, the more deficits children have accumulated. "The magnitude is massive — it's just astonishing," Deoni says of the findings, which are now under revision in *JAMA Pediatrics*.

When Deoni first posted his results on a preprint server², there was a flurry of worrying media coverage — and backlash from the research community. There was "a real concern about the fact that these results were being put out without proper peer review," Griffin says.

But, assuming the findings do have merit, why might babies born during the COVID-19 pandemic be experiencing significant cognitive — and especially motor — deficits? Deoni suspects that the problems stem from a lack of human-to-human interactions. In follow-up research that has not yet been published, he and his team have recorded parent-child interactions at home, finding that the number of words spoken by parents to their children, and vice versa, in the past two years has been lower than in previous years. He also suspects that babies and toddlers are not getting as much gross motor practice as usual because they aren't regularly playing with other

children or going to playgrounds. “And the unfortunate thing is that those skills kind of lay the foundation for all the other skills,” he says.

Other recent research supports the idea that lack of peer interactions could be holding some kids back. In a study published earlier this year, researchers in the United Kingdom surveyed 189 parents of children between the ages of 8 months and 3 years, asking whether their children received daycare or attended preschool during the pandemic, and assessing language and executive functioning skills. The authors found that the children’s skills were stronger if they had received group care during the pandemic, and that these benefits were more pronounced among children from lower-income backgrounds⁴.

Those most at risk seem to be children of colour or those from low-income families. For instance, a growing body of research suggests that among school-aged children, remote learning might be widening the already-large learning and development gaps between children from affluent and low-income backgrounds and between white kids and children of colour. In the Netherlands, researchers found that kids did worse on national assessments in 2020 — compared with the three previous years — and that learning losses were up to 60% larger for children from less-educated families⁵.

In parts of sub-Saharan Africa — including Ethiopia, Kenya, Liberia, Tanzania and Uganda — research suggests that some children have lost as much as a full year of learning⁶. And in the United States, after the first lockdown, a report by the consultancy firm McKinsey suggested that students of colour began school in autumn three to five months behind in learning, whereas white students were only one to three months behind (go.nature.com/3fauntp).

Masked effects

Children who have attended school or other group settings during the pandemic have typically been interacting with others who wore face masks. One important question is whether masks, which obscure parts of the face important for expressing emotions and speech, might also be affecting kids’ emotional and language development.



A mother with mask interacts with her daughter while waiting for her PCR results in Houston, Texas.Credit: Brandon Bell/Getty

Edward Tronick, a psychologist at the University of Massachusetts Boston, has been bombarded with e-mails from parents and paediatricians concerned about the potential developmental effects of masking. Tronick is famous for his 1975 ‘Still Face’ experiment, which showed that when birth parents suddenly remained straight-faced when interacting with their infants, their kids at first tried to get their attention, and then slowly withdrew and grew increasingly upset and wary⁷.

Tronick decided to see whether masks had a similar effect. With his colleague, psychologist Nancy Snidman, he conducted an experiment (which has not yet been peer reviewed) in which parents used smartphones to record interactions with their babies before, during and after they put on face masks. Although babies noticed when their parents put on masks — they would briefly change their facial expression, look away or point at the mask — they would then continue interacting with their parents as they had before⁸. The mask is blocking only one channel of communication, Tronick

says. “The parent wearing a mask is still saying, ‘I’m interacting with you, I’m still here for you, I’m still connecting to you.’”

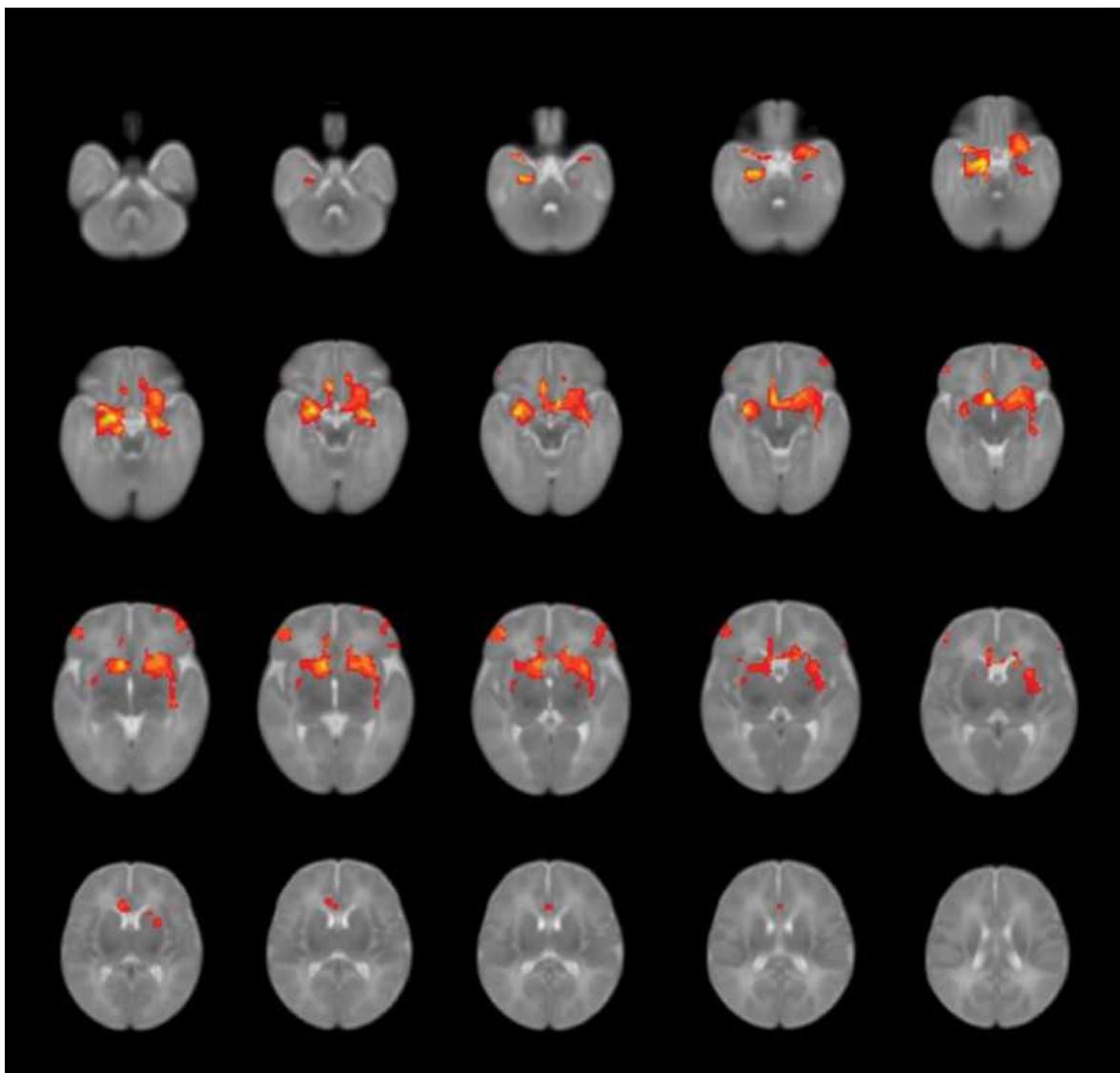
Face masks don’t seem to interfere all that much with emotional or language perception, either. A study published in May reported that two-year-olds were still able to understand words spoken by adults in opaque face masks⁹. Children “compensate for information deficits more readily than we think”, says study lead author Leher Singh, a psychologist at the National University of Singapore. Researchers in the United States found that, although face masks made it harder for school-age children to perceive adults’ emotions — about as difficult as when adults were wearing sunglasses — the kids were still, for the most part, able to make accurate inferences¹⁰.

“There’s a lot of other cues that kids can use to parse apart how other people are feeling, like vocal expressions, body expressions, context,” says study author Ashley Ruba, a postdoctoral fellow at the University of Wisconsin–Madison.

Pregnant and stressed

Other researchers are keen to know whether the pandemic could be affecting children’s development before they are born. Catherine Lebel, a psychologist who runs the Developmental Neuroimaging Lab at the University of Calgary in Canada, and her colleagues surveyed more than 8,000 pregnant people during the pandemic. Nearly half reported experiencing symptoms of anxiety, while one-third had symptoms of depression — a much higher percentage than in pre-pandemic years. How was this stress affecting babies in the womb?

To find out, the researchers used MRI imaging to scan the brains of 75 of the babies 3 months after birth. In a preprint posted in October, they found that babies born to people who reported more prenatal distress — more anxiety or depression symptoms — showed different structural connections between their amygdala, a brain region involved in emotional processing, and their prefrontal cortex, an area responsible for executive functioning skills¹¹.



Brain scans showing average connectivity patterns between the amygdala and other regions in infants. Pandemic-related stress during pregnancy weakened connections in some babies.Credit: Kathryn Manning

In a previous, small study, Lebel and her team had made the link between prenatal depression and brain connectivity differences in those same areas, and had suggested that in boys, these brain changes correlated with aggressive and hyperactive behaviour at preschool age¹². Other teams have found that changes in connectivity between these areas in adults are risk factors for depression and anxiety¹³. “Those are the areas that are involved in emotion processing, and lots of different behaviours,” Lebel says.

Other research has found similar associations between prenatal pandemic stress and child development. Livio Provenzi, a psychologist at the IRCCS Mondino Foundation in Pavia, Italy, and his colleagues observed that three-month-old babies of people who reported experiencing more stress and anxiety during pregnancy had more problems regulating their emotions and attention — they were less able to maintain their attention on social stimuli, for instance, and were less easily soothed — than were babies of people who were less stressed and anxious during pregnancy^{[14](#)}.

Thomason is running her own study to assess the effects of maternal stressors on children's brains and behaviour. She notes that, although there is a lot of concern about how prenatal stress might affect pandemic babies, early findings such as these do not mean that children are going to struggle for the rest of their lives. "Children are so adaptive, and elastic. And we do expect that things are going to improve and that they should be able to be resilient to a lot of what's happened," she says.

Indeed, research on historical disasters suggests that, although stress in the womb can be harmful to babies, it doesn't always have lasting effects. Children born to people who experienced considerable stress as a result of the 2011 floods in Queensland, Australia, showed deficits in problem-solving and social skills at six months of age, compared with children born to people who experienced less stress^{[15](#)}. However, by 30 months, these outcomes were no longer correlated with stress, and the more responsive that parents were to their babies' and toddlers' needs after birth, the better the toddlers did^{[16](#)}.

Caution and action

The research on pandemic babies presents a mixed picture, and scientists say it's too early to draw meaningful interpretations. For one thing, some of these early, often unpublished findings might not reflect reality, says Catherine Monk, a medical psychologist who works with Dimitriu at New York–Presbyterian.

The parents who chose to participate in some early studies, for instance, might not be a representative sample, Monk says. Perhaps they were already

worried about their kids on the basis of the behaviours they are seeing. Furthermore, she says, the results of in-person studies such as Deoni's could be affected by the wearing of face masks — perhaps not a lot, but enough to skew results.

As Thomason wrote last year in a commentary¹⁷ in *JAMA Pediatrics*, the incentive to publish interesting findings might also be shaping these early studies. “Scientists are quick to go look for a harmful difference. It’s the thing that’s going to get the attention of the media; it’s the thing that’s going to get published in a high-impact journal,” she says.

Researchers and funders are launching large studies and collaborations that could help to build a clearer picture. The US National Institute on Drug Abuse is funding a handful of studies through its Healthy Brain and Child Development Study. These will look at how maternal stress and substance use during the pandemic affect child development. In addition, alliances and conferences have been formed to bring researchers together and share emerging data. In March 2020, Thomason launched the international COVID Generation Research Alliance, which brings together researchers from 14 countries studying families with young children during the pandemic. The alliance, which held a research summit in November 2021, includes researchers in North and South America, Europe, Australia, Asia, the Middle East and Africa.

Even if kids’ brains are truly being affected by the pandemic, there is still time to steer them back on course, Dumitriu notes. “We can totally get ahead of this becoming a public-health emergency,” she says. “The brains of six-month-olds are very plastic, and we can get in there, and we can change their trajectory.”

Parents can make headway by playing and talking with their young children regularly, and giving them opportunities to play with others in safe settings. Policy changes aimed at supporting families and children could make a difference, too. Lebel’s research¹¹ found that meaningful social support, such as from a partner or close friend, during pregnancy resulted in much less prenatal distress. “We could do so much more of that in the prenatal care ecosystem,” says Monk. Researchers also argue for interventions that support families immediately after birth. Provenzi’s research¹⁴ has found

that people who had just given birth and were visited at home by nurses and neonatologists experienced less stress and anxiety than those who did not receive these visits.

Overall, researchers maintain that most children will probably be OK — but more than usual might currently be struggling. And if we want to support those who are falling behind, we should ideally intervene soon. “Kids are certainly very resilient,” Deoni says. “But at the same time, we also recognize the importance of the first 1,000 days of a child’s life as being the crucial early foundations.” The first pandemic babies, born in March 2020 are, at this point, more than 650 days old.

Children “are a product of their environment”, Deoni says. “The more that we can stimulate them and play with them and read to them and love them — that’s what it’s going to take.”

Nature **601**, 180-183 (2022)

doi: <https://doi.org/10.1038/d41586-022-00027-4>

References

1. 1.

Shuffrey, L. C. *et al. JAMA Pediatr.*
<https://doi.org/10.1001/jamapediatrics.2021.5563> (2022).

2. 2.

Deoni, S. C. L., Beauchemin, J., Volpe, A., D’Sa, V. & the RESONANCE Consortium. Preprint at medRxiv
<https://doi.org/10.1101/2021.08.10.21261846> (2021).

3. 3.

Zeanah, C. H. *et al. Am. J. Psychiatry* **166**, 777–785 (2009).

4. 4.

Davies, C. *et al.* *Infant Child Dev.* **30**, e2241 (2021).

5. 5.

Engzell, P., Frey, A. & Verhagen, M. D. *Proc. Natl Acad. Sci. USA* **118**, e2022376118 (2021).

6. 6.

Angrist, N. *et al.* *Int. J. Edu. Dev.* **84**, 102397 (2021).

7. 7.

Tronick, E., Als, H., Adamson, L., Wise, S., Brazleton, T. B. *J. Am. Acad. Child Psychiatry* **17**, 1–13 (1978).

8. 8.

Tronick, E. & Snidman, N. Preprint at SSRN
<https://doi.org/10.2139/ssrn.3899140> (2021).

9. 9.

Singh, L., Tan, A. & Quinn, P. C. *Dev. Sci.* **24**, e13117 (2021).

10. 10.

Ruba, A. L. & Pollak, S. D. *PLoS ONE* **15**, e0243708 (2020).

11. 11.

Manning, K. Y. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.10.04.21264536> (2021).

12. 12.

Hay, R. E. *et al.* *J. Neurosci.* **40**, 6969–6977 (2020).

13. 13.

Holmes, A. J. *et al.* *J. Neurosci.* **32**, 18087–18100 (2012).

14. 14.

Provenzi, L. *et al.* *Dev. Psychopathol.*
<https://doi.org/10.1017/S0954579421000766> (2021).

15. 15.

Simcock, G. *et al.* *Infancy* **22**, 282–302 (2017).

16. 16.

Austin, M.-P. *et al.* *Infant Behav. Dev.* **49**, 296–309 (2017).

17. 17.

Thomason, M. E. *JAMA Pediatr.*
<https://doi.org/10.1001/jamapediatrics.2021.5168> (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00027-4>

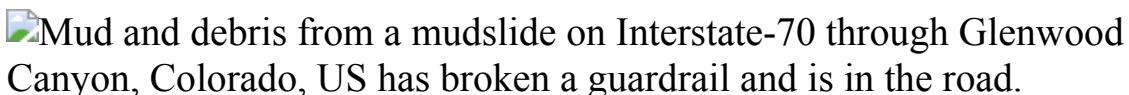
| [Section menu](#) | [Main menu](#) |

- NEWS FEATURE
- 12 January 2022

The devastating mudslides that follow forest fires

Regions that never used to burn are now suffering from forest fires — and that raises the risks of dangerous mudslides that are hard to forecast.

- [Jane Palmer](#) ⁰



Intense rain triggered mudslides in Glenwood Canyon, Colorado, during June and July last year — one year after a fire scorched the area. Credit: AP/Shutterstock

The summer of 2021 brought ideal fire weather to southern British Columbia in Canada. A dome of hot, high-pressure air settled over the area, sending temperatures soaring into record territory after months of drought.

In early July, an average of 40 fires sprang up each day in the province — in what would become one of the region's worst wildfire seasons ever recorded. Flames scorched the steep, tree-covered slopes above the valleys that are home to the Trans-Canada Highway and national railway lines, as well as oil and gas pipelines. The blaze raised concern about another imminent threat: landslides from the destabilized hills.

In mid-November, a massive storm known as an atmospheric river dumped a month's worth of rain on the region in just two days. When the downpour hit the burnt, scarred slopes, it set off giant surges of mud and debris that swept across the highway and railway lines. "They severed several key transportation corridors in western Canada, which then meant critical

supply-chain interruptions,” says Matthias Jakob, a geoscientist at BGC Engineering in Vancouver, Canada, who worked on assessing the potential for landslides in the region.

Although they might not be as deadly as hurricanes or heatwaves, landslides such as this can cost hundreds of millions of dollars in economic damage, Jakob says. And more are expected this winter as atmospheric rivers pummel the northwest of the United States and British Columbia, following a year of intense and widespread fires.

This one-two punch of fire and flood, is just a taste of what’s to come there, and in many other regions. Climate change is increasing the frequency and intensity of fires around the world, and it is also leading to more extreme bouts of precipitation¹. When rain hits recently burnt hillsides it can trigger a particularly deadly type of landslide called a debris flow — a water-laden mass of soil, rocks and other matter that can surge downslope with devastating force.

In regions prone to these types of hazard, scientists have developed models that can feed into warning systems and save lives. But wildfires are now claiming hill slopes that have rarely been burnt before. This is happening around the world, from the boreal forests of Alaska and Canada to the mountains of Austria, where fires are burning more frequently and more intensely than ever before.

“There will be an increased likelihood of both fire and post-fire debris flows in areas where they are currently only a theoretical possibility,” says Bruce Malamud, a geophysicist at King’s College London.

The central region of British Columbia has always had wildfires but now the province is even seeing blazes in coastal areas. The models used to forecast inland debris flows simply wouldn’t work for these regions, where the soils and vegetation differ, Jakob says. It’s a similar scenario in the United States, where fires in the past few years have scorched areas of northern California, Oregon and Washington that rarely burn.

Evidence suggests that the debris-flow behaviour is going to be different in these wetter, more vegetated regions than in arid southwestern states, where

fires and landslides are more common, says Jason Kean, a hydrologist at the United States Geological Survey (USGS) National Landslide Hazards Program in Golden, Colorado.

“While it’s kind of clear what’s happening in the drier areas, the picture gets murkier as we move north,” Kean says. “So right now, we are scrambling to collect data to figure out how well our current model works and how to make a better one.”

Surging slopes

When rainfall hits an unburnt slope, it gets caught on trees or shrubs and then slowly trickles to the ground, where it can filter into the soil. But fires make it harder for hillsides to soak up water. When flames burn away vegetation, that allows rain to hit the surface with its full force. And fires burn off the waxy compounds that coat leaves and needles, some of which then evaporate and condense on cooler soil particles just below the surface. That can form a water-repellent layer that stops the soil from absorbing water².

“That takes the soil from acting like a sponge that just soaks up the rain, to acting like a giant plastic sheet that water runs off,” says Gary Sheridan, a soil scientist at the University of Melbourne, Australia.



A fire burns hillsides in Glenwood Canyon in August 2020.Credit: Helen H. Richardson/MediaNews Group/The Denver Post via Getty

Rain runs off this water-repellent surface until it reaches a crack and then can soak into the soil. But if the rain is falling at an intense rate, only some of it seeps into crevices and the rest forms rivulets that flow downhill. These streams pick up soil and rocks, which gain speed, and surge downhill creating a debris flow.

“It can be a small amount of rainfall, but it needs to be intense,” says Stefan Doerr, a wildfire scientist at Swansea University, UK. “Particularly soon after a fire, because the stuff is just sitting there, ready to be moved.”

When an extreme storm, known as a bomb cyclone, blasted California and the Pacific Northwest in October 2021, it broke single-day rainfall records for several regions that didn’t burn and caused flash flooding, but didn’t set off any significant mudslides in the burnt areas. “It is not about the storm total rainfall, but the intensity of the rain,” says Nina Oakley, a research meteorologist at the Scripps Institute of Oceanography in San Diego, California, who studies the connection between extreme rainfall

events and landslides. “That’s why you really need 15-minute rainfall information to forecast post-fire debris flows.”

Mudslides and similar types of slope failure can cause tremendous damage. In the United States, rough estimates put the direct economic costs of landslides in the range of several billion dollars each year, says Jonathan Godt, the coordinator of the Landslide Hazards Program at the USGS. But the indirect costs, which include losses in commerce and related factors, are probably several times that estimate, Godt says.

In January 2021, Congress passed the National Landslide Preparedness Act, which authorized US\$37 million annually from 2021 to 2024 for federal agencies to broaden their efforts to reduce the hazards from landslides. One goal of the bill is to expand existing early-warning systems for post-wildfire burnt landscapes in the United States.

After a fire, teams of US federal and academic soil scientists, hydrologists, biologists and landslide experts typically flock to the area immediately, even before the flames are completely out, to assess the damage. US Forest Service researchers evaluate the severity of the burn on vegetation from the ground and from the skies, by comparing satellite imagery from before and after a fire. They pass on the information to USGS landslide researchers, who feed the data into their models and create hazard maps, designed to indicate the potential for debris flows across the burnt region.

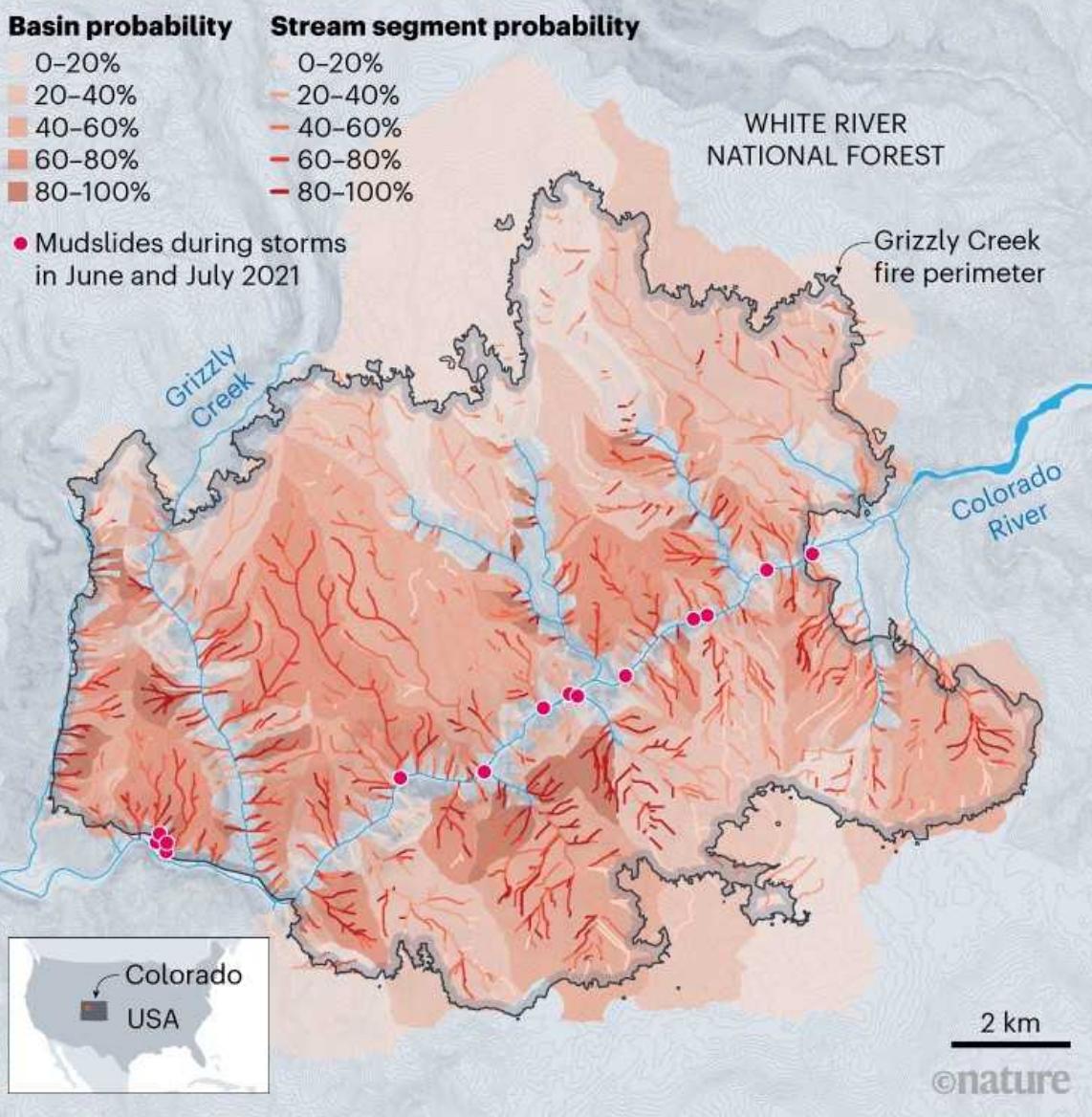
The models take into account the size and steepness of the affected area, how severely the fire scorched it and the nature of the soil, as well as a multitude of other variables. With all of that information, the models indicate how probable debris flows will be depending on the intensity of rainfall. It’s a statistical association based on data from previous post-fire mudslides in the southwest of the United States.

Although relatively simple, the models have proved effective at informing early-warning systems in that part of the United States, where the slopes share similar vegetation and soils. After a fire tore through Glenwood Canyon, Colorado, in August 2020, scientists were quick to assess the damage and feed the data into a model, which helped emergency-response managers to assess the risks of debris flows when rainstorms approached

(see ‘Unstable slopes’). In mid-2021, that warning system prompted authorities to close Interstate 70 — a highway that runs through the canyon — a number of times. Several debris flows surged across the major road. No lives were lost, but it will cost more than \$100 million to repair the damage to the highway.

UNSTABLE SLOPES

After the Grizzly Creek fire burnt hillsides in Glenwood Canyon, Colorado, in 2020, researchers at the United States Geological Survey used a model to estimate the likelihood of debris flows during rainstorms. The map shows risk estimates for drainage basins and streams, based on a storm with a maximum rainfall intensity of 24 millimetres per hour for a 15-minute interval. Storms in July 2021 significantly exceeded that threshold.



Source: USGS

When the local authorities in Oregon ask the USGS scientists to make hazard maps after fires in the state, however, the model output has limited

value because the environment is so different from conditions in the arid southwest. “You can’t take the probability numbers at face value because the model doesn’t apply to that area,” Kean says.

Delayed disasters

That became clear after a fire tore across hillsides in Oregon’s Columbia River Gorge in September 2017. In the steep, forested slopes, fir trees typically tower over a dense, lush undergrowth of ferns and bushes. The fire that year incinerated the vegetation in a 20,000 hectare area. But despite some intense storms in the following three years, there were no catastrophes. Then, in January 2021, heavy rains triggered a deadly mudslide. The torrent of mud swept away a car driving through the gorge, killing its driver.

The delayed catastrophe didn’t surprise geologist Joshua Roering, at the University of Oregon in Eugene, who has been studying landslides in the Pacific Northwest for nearly two decades. Roering’s group has been investigating how the Pacific Northwest soils can behave differently from those in drier locations, after a fire.

So far, his group has found that after flames scorch vegetation, a water-repellent layer coats the soil, but it is fragmented and not as continuous as the coatings formed farther south. So water can soak into the ground during extreme rainfall in the Pacific Northwest, even after a fire. The fact that several significant storms haven’t triggered debris flows in the region, suggests that the hydrologic system behaves differently from those observed in southern California, Roering says.

That doesn’t mean the post-fire danger disappears, Roering cautions. Instead, the risks of debris flows are delayed, a pattern researchers have seen after timber harvesting on slopes. When a tree is cut, its roots decay over several years. And as their hold on the soil loosens, the chances grow for a slope to slide when rain hits³. “You cut down a tree, or you burn a tree,” Roering says. “It doesn’t matter how you kill a tree, those root systems in the upper metre of the soil are losing about 90% of their strength after three, four or five years.”

If this theory bears out, scientists and emergency managers will need to prepare for storms several years after wildfires in the heavily wooded Pacific Northwest, Roering says. “If that’s the case, then we’re sort of sitting on this ticking time bomb in terms of badly burned areas, with steep slopes, perched above major travel corridors and freeways,” he says.

Researchers don’t expect the same kind of delay for the central region of British Columbia, which is more arid and has different soils from coastal parts of the Pacific Northwest. “We really need to create our own model for British Columbia, ideally one for the coast and one for the drier interior,” Jakob says. “But for that we need much more data.”

In Australia, researchers are also grappling with changing patterns of fires and debris flows. Wildfires are common in southeastern Australia, where Sheridan works, but now they’re starting to spring up in the wet, alpine areas of Tasmania, which have rarely seen flames before. Rather than create a whole new landslide model, Sheridan’s group is attempting to refine it, one fire at a time, and to test its effectiveness with each debris flow. But it’s a tricky and drawn-out process, Sheridan says.

After a fire, there’s such a diverse range of outcomes, from almost no change to the landscape to catastrophic debris flows and landslides, and everything in between, Sheridan says. Landslide researchers around the world are discussing a common set of factors that can be fed into a universal forecasting model, but such a model is still a long way off, because the landscape, climate and soil behaviour varies in each location, he says. “There’s such a high variability in risk from place to place, and we still don’t know why,” Sheridan says.

Debris-flow models can only go so far. They help to save lives, but it’s more difficult to keep buildings or transport networks safe. In British Columbia, there simply isn’t the money to protect roads, railway lines or pipelines from every single slope that could yield a debris flow, Jakob says. “Early-warning systems can prevent harm to people, but infrastructures will still be damaged or obliterated.”

Moving forwards, it will be even more important to use models to identify roads or pipelines at risk, so that authorities can then prioritize where to

shore up defences, Jakob says. Homeowners, too, need to know the risks to their property, and lives.

“With the dramatic and rapid change in climate, many people may not know that they are at risk from post-fire landslides,” Jakob says. Still, such efforts are not without controversy. Politicians and developers have been known to oppose efforts to create landslide hazard maps because they see them as hindering growth and income.

Only time, more storms and more data will help researchers learn how to better forecast when or where landslides might strike after fires. In anticipation of the winter storm season in the western United States, researchers have installed instruments on some of the slopes that burnt last year in California blazes, including the giant Caldor fire and the Dixie fire — the largest individual blaze in state history.

The combined network of instruments will tell scientists the intensity of rainfall that does, or doesn’t, trigger debris flows; how much water seeps into the ground; how much runs off; and other crucial variables that will help to improve current models or build region-specific ones, Kean says.

The biggest challenge to building a clearer picture is the lack of resources in relation to the scale of the wildfire devastation, Kean says. The Dixie fire alone burnt nearly 400,000 hectares and the scientists have a restricted number of instruments, which limits their ability to build up knowledge regarding debris flow behaviour in new regions. “We don’t have a big-data problem. We have a little-data problem,” Kean says.

The lack of resources and information about landslides is universal for researchers across the world and it hinders the community’s ability to advance its understanding. “You’ve got a tiny little bit of data,” says Sheridan, “and it’s like you’re trying to read the tea leaves and work out what’s going on.”

Nature **601**, 184-186 (2022)

doi: <https://doi.org/10.1038/d41586-022-00028-3>

References

1. 1.

Tabari, H. *Sci. Rep.* **10**, 13768 (2020).

2. 2.

Mataix-Solera, J. *et al.* *Catena* **108**, 6–13 (2013).

3. 3.

Montgomery, D. R. *et al.* *Geology* **28**, 311–314 (2000).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00028-3>

| [Section menu](#) | [Main menu](#) |

Opinion

- **[Unblock research bottlenecks with non-profit start-ups](#)** [11 January 2022]
Comment • ‘Focused research organizations’ can take on mid-scale projects that don’t get tackled by academia, venture capitalists or government labs.
- **[Wind power versus wildlife: root mitigation in evidence](#)** [11 January 2022]
Correspondence •
- **[EU Nature Restoration Law needs ambitious and binding targets](#)** [11 January 2022]
Correspondence •
- **[Two million species catalogued by 500 experts](#)** [11 January 2022]
Correspondence •
- **[Abel-winning geometry pioneer, remembered](#)** [11 January 2022]
Correspondence •

- COMMENT
- 11 January 2022

Unblock research bottlenecks with non-profit start-ups

‘Focused research organizations’ can take on mid-scale projects that don’t get tackled by academia, venture capitalists or government labs.

- [Adam Marblestone](#) ⁰,
- [Anastasia Gamick](#) ¹,
- [Tom Kalil](#) ²,
- [Cheryl Martin](#) ³,
- [Milan Cvitkovic](#) ⁴ &
- [Samuel G. Rodrigues](#) ⁵



Mouse neurons imaged using a method similar to one being developed as a high-throughput tool by a focused research organization. Credit: P. W. Tillberg *et al.*/*Nature Biotechnol.*

It takes more than a great idea to accomplish a great project. Our research and experience have convinced us that many worthy projects wither or are never launched because neither academic laboratories, start-up firms nor government facilities can support them.

This applies particularly to projects that would produce public goods, such as data sets or tools, that could make research faster and easier. Few research-enabling projects will be commercially viable enough to attract venture capital. Nor is academia a suitable incubator. Academics can rarely muster the time, focus and workforce coordination needed to turn a proof-

of-principle technology into a robust, scalable technique or to transform a research project into a platform. These engineering improvements do not fulfil teaching requirements or provide the papers or pizzazz that both senior academics and their trainees need to propel their careers.

A type of non-profit start-up could be a better way to support projects that enable research. These would have full-time scientists, engineers and executives, and total funding of about US\$20 million to \$100 million that would last around 5 years — longer than most grants or venture-capital funding rounds allow. And they would be set up to pursue predefined milestones, such as improving the resolution of a measurement system by tenfold, or gathering a pre-specified amount of data. We call them focused research organizations (FROs).

Fledgling FROs

FROs are conceptually similar to grand projects established to support ambitious and charismatic initiatives. The Human Genome Project, the Large Hadron Collider at CERN (Europe's particle-physics laboratory near Geneva, Switzerland) and the Hubble Space Telescope come to mind. Other examples include large data-collection projects such as the Human Cell Atlas, ENCODE (the Encyclopedia of DNA Elements) and the BRAIN Initiative Cell Census Network. Each of these produced research resources that are now considered essential for routine work.

But such grand projects are one-off, once-a-decade efforts that require feats of coordination and consensus, not least because most work is usually done by academics who are employed not by the project, but by their own research institutions. The Human Genome Project, for instance, cost \$5.4 billion in today's dollars and took almost 13 years. The FRO model would support a steady stream of smaller-scale projects that could be launched with much less effort.

How did we arrive at this idea? Starting in 2014, two of us (S.G.R. and A.M.) began working on efforts to establish non-academic projects in brain-mapping technology and realized that no suitable organizations existed. We then interviewed more than 100 people who had developed research-

enabling tools (administrators and researchers at university, government, corporate and military labs, as well as entrepreneurs). What hampered their work? What did they think of the FRO idea? In May 2021, a [workshop on national networks of research institutes](#), sponsored by the US National Science Foundation, vetted the concept and explored new models of institutional innovation that could lead to the next generation of science institutes and FROs. (C.M. was a co-organizer of that event.)

In October 2021, after months of ideation, review and design work, three of us (A.M., T.K. and A.G.) launched an incubator to support FROs, called Convergent Research (where M.C. also works) in Arlington, Massachusetts. It is part of the Schmidt Futures Network, which aims to help scale and diversify the funding sources of promising early-stage initiatives supported by Schmidt Futures, a philanthropic organization headquartered in New York City. We also helped to launch two philanthropic FROs with the support of Schmidt Futures, and a third effort supported by the non-profit Astera Institute in Berkeley, California. The proposed ‘products’ include high-throughput brain-mapping techniques, tools to engineer non-model microorganisms and comprehensive analysis of ageing interventions in mice. The projects have begun to hire core technical staff and recruit advisory boards. Each has secured its own lab space and purchased initial equipment. Experiments will begin early this year.

In a few years, we’ll learn whether these FROs can accelerate research in neuroscience, synthetic biology and longevity. We will also test how well FROs support the overall goal of enabling research.

Industry innovation

In the second half of the twentieth century, Bell Labs, Xerox PARC and other large US corporate labs famously merged aspects of fundamental research with large-scale product development and manufacturing (see ‘Innovation invented’). They are credited with introducing laser printers, photovoltaic cells in solar panels, the programming language C++, transistors and more. Such innovation continues, as shown by the success of Alphabet’s subsidiary Google DeepMind in developing its [AlphaFold algorithm for predicting protein folding](#) just last year. Overall, however,

most industry labs today lack the freedom to pursue projects that are divorced from nearer-term commercial objectives, and the resulting knowledge is often kept proprietary.

Innovation invented

Selected launches of institutions that support research.

1925: Bell Telephone Laboratories set up in the United States for fundamental research in communications.

1933: US Rockefeller Foundation funds molecular-biology initiative.

1942: US government launches Manhattan Project to develop nuclear weapons.

1958: Advanced Research Projects Agency (ARPA) founded by US government to advance military technologies.

1970: US laboratory Xerox PARC emerges from early computer-research communities convened by ARPA.

1972: ARPA renamed Defense Advanced Research Projects Agency (DARPA).

1976: US biotechnology firm Genentech founded with venture capitalists.

1990: Human Genome Project launches. Governments invest billions.

1998: Europe's particle-physics laboratory CERN begins construction of the Large Hadron Collider near Geneva, Switzerland.

2006: HHMI Janelia Research Campus opens in Ashburn, Virginia, to pursue neuroscience in a collaborative way.

2008: Large Hadron Collider starts up.

2009: Advanced Research Projects Agency-Energy (ARPA-E) founded to focus on energy challenges using the DARPA model.

2016: Francis Crick Institute opens in London. It has strong links with universities and provides internal funding.

2018: European Innovation Council pilot launches to support commercialization of high-risk, high-impact technologies.

2021: Institutional experimentation in science, including launch of the first focused research organizations (FROs).

The Defense Advanced Research Projects Agency (DARPA) model is widely, and reasonably, considered an exemplar of institutional innovation. It identifies highly specific technological needs and charges research groups with fulfilling well-defined tasks¹. The United States has extended this model to other Advanced Research Projects Agencies: ARPA-E for energy and IARPA for intelligence-related projects².

Several other DARPA-like agencies have been created in the past year or are under discussion. These include ARPA-C (climate) and ARPA-H (health) in the United States³; the Advanced Research and Invention Agency in the United Kingdom; and Wellcome Leap (focusing on global human health). Others are Actuate, a US non-profit organization that will “create breakthroughs” for complex societal problems, led by former DARPA director Arati Prabhakar, and PARPA, a ‘private ARPA’ effort focused on space and Earth technologies. Launched by the technologist Ben Reinhardt, PARPA will support work that is “too researchy for a start-up and too engineering-heavy for academia” (see go.nature.com/3h6idjy).

ARPA-like programmes do many things that FROs should also do. They focus on technical milestones that are challenging. Programme managers have the power to start up, and shut down, high-risk projects. And DARPA deliberately plans for projects to end in a finite time and for resulting technologies to move to broader applications. But the ARPA model still relies on researchers who are employed elsewhere. This means that, within companies, projects can fail when they are not aligned with the company’s main focus. ARPA projects in academia can struggle when they are not aligned with a lab’s need to generate publications.

Some organizations are set up to take on tightly specified government- or industry-sponsored applied-engineering projects. Research institutions with this purpose include the non-profit SRI International in Menlo Park, California, the Fraunhofer Society institutes in Germany and the Battelle Memorial Institute in Columbus, Ohio. The European Union has a mechanism called the European Innovation Council Accelerator that makes mid-scale investments in ‘deep tech’ start-ups. Germany’s Helmholtz Association enables large partnerships between a Helmholtz Centre and a university, bringing more-scalable resources to basic science, as do similar institutes elsewhere.

Some permanent institutes, such as the Allen Institute for Brain Science in Seattle, Washington, or the Howard Hughes Medical Institute’s Janelia Research Campus in Ashburn, Virginia, are distinct enough from academia to conduct large-scale data collection and to develop broadly used tools. For example, the Allen Brain Atlas established a technique that can be used on an industrial scale to map gene expression across the mouse brain, standardized such that data from many experiments are easily unified. Likewise, a Janelia team developed a systematic, industrialized approach to produce a fluorescent calcium indicator protein that allows neuroscience researchers worldwide to record when neurons fire in the brain. This came from a project called [GENIE \(Genetically Encoded Neuronal Indicators and Effectors\)](#).

Janelia separates its large-scale projects, which are professionally staffed and managed, from its academic research projects. The latter are staffed by students and postdocs in small labs led by principal investigators. But such institutes are few and expensive, and are difficult to create. Furthermore, their nimbleness is hard to sustain. Janelia’s first director, Gerald Rubin, wrote in 2019 that “without an opposing force provided by management, there is a slow, steady drift toward a more conventional environment increasingly focused on maintaining successful programs and documenting individual achievement at the expense of risk-taking and collaborative, interdisciplinary work”⁴.

Expanded playbook

Our goal is to create a model to support an ecosystem of small- to mid-scale projects that fall between the cracks of what start-ups, academia and other organizations do. Start-up companies have a standard playbook involving business agreements and pitch meetings. Academic funding has standard requirements such as CVs and project proposals. We hope to develop a similar playbook as we monitor our FROs. This will make future launches easier.

Projects that can be achieved through start-ups or in an academic lab should be done in those ways — the existing infrastructure is vast and fit for purpose. The mission of each FRO should be to get technologies or data sets deployed quickly so they can be used effectively across the research community.

That means FROs should move beyond academic proof-of-concept into standardized systems that don't rely on graduate students to fix glitches with tweezers and tape. For example, an FRO might have a milestone both to develop a technology and to demonstrate that independent labs can implement it. It should have time-bound milestones unrelated to academic publishing, and strong project management to help achieve them. Training students or aiming to become a permanent institution must not be part of the mission.

There are known and unknown hurdles to the FRO model. One key question is how to maintain strong relationships and interactions with existing academic efforts that might have planted the seeds for an FRO. Another is how to get logistics in place and a lab running quickly, as well as smoothly establishing partnerships with other organizations, where needed. And finally, crucially, we need to learn how to ensure that resources developed in FROs are applied to meet real needs.

Perhaps semi-permanent teams of project managers and administrators can be matched with scientific staff from the start, essentially serving as hosts for a focused scientific leadership team. Other questions involve career progression, including what will lure talent away from academia or the potential financial returns of a start-up, and how to enable strong, post-FRO career options for all staff.

These are questions we will have in mind as we track the FROs that were launched last October, and as we iterate, launch more FROs and expand the base of philanthropic support. We and others hope to develop the model to a point at which governments could set goals to fund a certain number of FROs each year, confident that, although some will fail, others will make research more powerful and efficient.

The US National Institutes of Health, DARPA, the National Science Foundation, venture capitalists, US National Laboratories, the modern research university, and every other institution that advances science and technology are human inventions that can be refined with experience and experimentation. We believe the same will be true of FROs.

Nature **601**, 188-190 (2022)

doi: <https://doi.org/10.1038/d41586-022-00018-5>

References

1. 1.

Dugan, R. E. & Gabriel, K. J. *Harv. Bus. Rev.* **91**, 74–84 (2013).

2. 2.

Prabhakar, A. *How to Unlock the Potential of the Advanced Research Projects Agency Model* (Day One Project, 2021).

3. 3.

Collins, F. S., Schwetz, T. A., Tabak, L. A. & Lander, E. S. *Science* **373**, 165–167 (2021).

4. 4.

Rubin, G. M. & O’Shea, E. K. *eLife* **8**, e44826 (2019).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00018-5>

| [Section menu](#) | [Main menu](#) |

- CORRESPONDENCE
- 11 January 2022

Wind power versus wildlife: root mitigation in evidence

- [Tim Schmoll](#)⁰ &
- [Frank M. Schurr](#)¹

Germany's new government plans to dramatically expand the production of onshore wind power. It intends to deploy "innovative technical mitigation measures such as anti-collision systems" for turbines to avoid large-scale killing of birds and bats and undermining conservation goals. We argue that it should also use this roll-out to systematically evaluate such systems.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 191 (2022)

doi: <https://doi.org/10.1038/d41586-022-00012-x>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00012-x>

- CORRESPONDENCE
- 11 January 2022

EU Nature Restoration Law needs ambitious and binding targets

- [Kris Decleer](#) ORCID: <http://orcid.org/0000-0001-9621-8925>⁰,
- [Jordi Cortina-Segarra](#) ORCID: <http://orcid.org/0000-0002-8231-3793>¹ &
- [Aveliina Helm](#) ORCID: <http://orcid.org/0000-0003-2338-4564>²

Initiatives by the European Commission to restore the continent's degraded areas ([J. Cortina-Segarra et al. *Nature* 535, 231; 2016](#)) have proved disappointing. As the United Nations Decade on Ecosystem Restoration gathers momentum, the commission is preparing a law that has legally binding targets. To underscore the urgency, some 1,400 European scientists and 30 expert networks and institutions have signed a declaration by the Society for Ecological Restoration Europe (see [go.nature.com/3st6k88](#)).

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 191 (2022)

doi: <https://doi.org/10.1038/d41586-022-00011-y>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00011-y>.

- CORRESPONDENCE
- 11 January 2022

Two million species catalogued by 500 experts

- [Mark John Costello](#)⁰,
- [R. Edward DeWalt](#)¹,
- [Thomas M. Orrell](#)² &
- [Olaf Banki](#)³

More than two million accepted species are now listed in the open-access Catalogue of Life (go.nature.com/3ym3h2g). This achievement addresses a major impediment to the management of biodiversity data by presenting an almost complete index of accepted names and known synonyms.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

Subscribe

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

Buy

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 191 (2022)

doi: <https://doi.org/10.1038/d41586-022-00010-z>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00010-z>

- CORRESPONDENCE
- 11 January 2022

Abel-winning geometry pioneer, remembered

- [Jean-Pierre Bourguignon](#)⁰,
- [Michel Broué](#)¹,
- [Alain Connes](#)² &
- [Étienne Ghys](#)³

The Belgium-born French mathematician Jacques Tits, an honorary professor at Collège de France in Paris who transformed geometry in the twentieth century, died last month, aged 91. A child prodigy, he also shared the 1993 Wolf Prize in Mathematics, won the German Mathematical Society's Cantor medal in 1996 and shared the 2008 Abel Prize for “profound achievements in algebra and in particular for shaping modern group theory”.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 191 (2022)

doi: <https://doi.org/10.1038/d41586-022-00009-6>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00009-6>

Work

- **Weaving Indigenous knowledge into the scientific method**

[11 January 2022]

Career Feature • Scientists and funders with close links to local communities outline how Western teams can collaborate fairly and effectively with those groups.

- **Pleistocene dreams: recreating ancient grasslands to save the planet** [10 January 2022]

Where I Work • Sergey Zimov is rewilling a remote part of northeastern Siberia to help fight climate change.

- CAREER FEATURE
- 11 January 2022

Weaving Indigenous knowledge into the scientific method

Scientists and funders with close links to local communities outline how Western teams can collaborate fairly and effectively with those groups.

- [Saima May Sidik](#)

[Find a new job](#)



Dominique David-Chavez works with Randal Alicea, an Indigenous farmer, in his tobacco-drying shed in Cidra, Borikén (Puerto Rico). Credit: Norma Ortiz

Many scientists rely on Indigenous people to guide their work — by helping them to find wildlife, navigate rugged terrain or understand changing weather trends, for example. But these relationships have often felt colonial, extractive and unequal. Researchers drop into communities, gather data and leave — never contacting the locals again, and excluding them from the publication process.

Today, many scientists acknowledge the troubling attitudes that have long plagued research projects in Indigenous communities. But finding a path to better relationships has proved challenging. Tensions surfaced last year, for example, when seven University of Auckland academics argued that planned changes to New Zealand’s secondary school curriculum, to “ensure parity between mātauranga Māori”, or Maori knowledge, and “other bodies of knowledge”, could undermine trust in science.

Last month, the University of Auckland’s vice-chancellor, Dawn Freshwater, announced a symposium to be held early this year, at which different viewpoints can be discussed. In 2016, the US National Science Foundation (NSF) launched Navigating the New Arctic — a programme that encouraged scientists to explore the wide-reaching consequences of climate change in the north. A key sentence in the programme description reflected a shift in perspective: “Given the deep knowledge held by local and Indigenous residents in the Arctic, NSF encourages scientists and Arctic residents to collaborate on Arctic research projects.” The Natural Sciences and Engineering Research Council of Canada and New Zealand’s Ministry of Business, Innovation and Employment have made similar statements. So, too, have the United Nations cultural organization UNESCO and the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

But some Indigenous groups feel that despite such well-intentioned initiatives, their inclusion in research is only a token gesture to satisfy a funding agency.

There's no road map out of science's painful past. *Nature* asked three researchers who belong to Indigenous communities in the Americas and New Zealand, plus two funders who work closely with Alaskan Natives, how far we've come toward decolonizing science — and how researchers can work more respectfully with Indigenous groups.

DANIEL HIKUROA: Weave folklore into modern science

Daniel Hikuroa is an Earth systems and environmental humanities researcher at Te Wānanga o Waipapa, University of Auckland, New Zealand, and a member of the Māori community.

We all have a world view. Pūrākau, or traditional stories, are a part of Māori culture with great potential for informing science. But what you need to understand is that they're codified according to an Indigenous world view.

For example, in Māori tradition, we have these things called taniwha that are like water serpents. When you think of taniwha, you think, danger, risk, be on your guard! Taniwha as physical entities do not exist. Taniwha are a mechanism for describing how rivers behave and change through time. For example, pūrākau say that taniwha live in a certain part of the Waikato River, New Zealand's longest, running for 425 kilometres through the North Island. That's the part of the river that tends to flood. Fortunately, officials took knowledge of taniwha into account when they were designing a road near the Waikato river in 2002. Because of this, we've averted disasters.

Sometimes, it takes a bit of explanation to convince non-Indigenous scientists that pūrākau are a variation on the scientific method. They're built on observations and interpretations of the natural world, and they allow us to predict how the world will function in the future. They're repeatable, reliable, they have rigour, and they're accurate. Once scientists see this, they have that 'Aha!' moment where they realize how well Western science and pūrākau complement each other.

We're very lucky in New Zealand because our funding agencies help us to disseminate this idea. In 2005, the Ministry of Research, Science and

Technology (which has since been incorporated into the Ministry of Business, Innovation and Employment) developed a framework called Vision Mātauranga. Mātauranga is the Māori word for knowledge, but it also includes the culture, values and world view of Māori people. Whenever a scientist applies for funding, they're asked whether their proposal addresses a Māori need or can draw on Māori knowledge. The intent of Vision Mātauranga is to broaden the science sector by unlocking the potential of Māori mātauranga.

In the early days of Vision Mātauranga, some Indigenous groups found themselves inundated with last-minute requests from researchers who just wanted Indigenous people to sign off on their proposals to make their grant applications more competitive. It was enormously frustrating. These days, most researchers are using the policy with a higher degree of sophistication.

Vision Mātauranga is at its best when researchers develop long-term relationships with Indigenous groups so that they know about those groups' dreams and aspirations and challenges, and also about their skill sets. Then the conversation can coalesce around where those things overlap with the researchers' own goals. The University of Waikato in Hamilton has done a great job with this, establishing a chief-to-chief relationship in which the university's senior management meets maybe twice a year with the chiefs of the Indigenous groups in the surrounding area. This ongoing relationship lets the university and the Indigenous groups have high-level discussions that build trust and can inform projects led by individual labs.

We've made great progress towards bridging Māori culture and scientific culture, but attitudes are still evolving — including my own. In 2011, I published my first foray into using Māori knowledge in science, and I used the word 'integrate' to describe the process of combining the two. I no longer use that word, because I think weaving is a more apt description. When you weave two strands together, the integrity of the individual components can remain, but you end up with something that's ultimately stronger than what you started with.

DOMINIQUE DAVID-CHAVEZ: Listen and learn with humility

Dominique David-Chavez is an Indigenous land and data stewardship researcher at Colorado State University in Fort Collins, and a member of the Arawak Taino community.

People often ask how can we integrate Indigenous knowledge into Western science. But framing the question in this way upholds the unhealthy power dynamic between Western and Indigenous scientists. It makes it sound as though there are two singular bodies of knowledge, when in fact Indigenous knowledge — unlike Western science — is drawn from thousands of different communities, each with its own knowledge systems.

At school, I was taught this myth that it was European and American white men who discovered all these different physical systems on Earth — on land, in the skies and in the water. But Indigenous people have been observing those same systems for hundreds or thousands of years. When Western scientists claim credit for discoveries that Indigenous people made first, they're stealing Indigenous people's contributions to science. This theft made me angry, but it also drove me. I decided to undertake graduate studies so that I could look critically at how we validate who creates knowledge, who creates science and who are the scientists.

To avoid perpetuating harmful power dynamics, researchers who want to work in an Indigenous people's homeland should first introduce themselves to the community, explain their skills and convey how their research could serve the community. And they should begin the work only if the community invites them to. That invitation might take time to come! The researchers should also build in time to spend in the community to listen, be humbled and learn.

If you don't have that built-in relational accountability, then maybe you're better off in a supporting role.

Overall, my advice to Western researchers is this: always be questioning your assumptions about where science came from, where it's going and what part you should be playing in its development.

MARY TURNIPSEED: Fund relationship building and follow-ups

Mary Turnipseed is an ecologist and grantmaker at the Gordon and Betty Moore Foundation, Palo Alto, California.

I've been awarding grants in the Arctic since 2015, when I became the Gordon and Betty Moore Foundation's marine-conservation programme officer. A lesson I learnt early on about knowledge co-production — the term used for collaborations between academics and non-academics — is to listen. In the non-Indigenous parts of North America, we're used to talking, but flipping that on its end helps us to work better with Indigenous communities.

Listening to our Indigenous Alaskan Native partners is often how I know whether a collaboration is working well or not. If the community is supportive of a particular effort, that means they've been able to develop a healthy relationship with the researchers. We have quarterly check-ins with our partners about how projects are going; and, in non-pandemic times, I frequently travelled to Alaska to talk directly with our partners.

One way in which we help to spur productive relationships is by giving research teams a year of preliminary funding — before they even start their research — so that they can work with Indigenous groups to identify the questions their research will address and decide how they're going to tackle them. We really need more funding agencies to set aside money for this type of early relationship-building, so that everyone goes into a project with the same expectations, and with a level of trust for one another.



Members of the Ikaagvik Sikukun collaboration at the Native Village of Kotzebue, Alaska. Credit: Sarah Betcher/Farthest North Films

Developing relationships takes time, so it's easiest when Indigenous communities have a research coordinator, such as Alex Whiting (environmental programme director for the Native Village of Kotzebue), to handle all their collaborations. I think the number of such positions could easily be increased tenfold, and I'd love to see the US federal government offer more funding for these types of position.

Funding agencies should provide incentives for researchers to go back to the communities that they've worked with and share what they've found. There's always talk among Indigenous groups about researchers who come in, collect data, get their PhDs and never show up again. Every time that happens, it hurts the community, and it hurts the next researchers to come. I think it's essential for funding agencies to prevent this from happening.

ALEX WHITING: Develop a toolkit to decolonize relationships

Alex Whiting is an environmental specialist in Kotzebue, Alaska, and a formally adopted member of the Qikiktagrukmiut community.

A lot of the time, researchers who operate in a colonial way aren't aware of the harm they're doing. But many people are realizing that taking knowledge without involving local people is not only unethical, but inefficient. In 1997, the Native Village of Kotzebue — a federally recognized seat of tribal government representing the Qikiktagrukmiut, northwest Alaska's original inhabitants — hired me as its environmental programme director. I helped the community to develop a research protocol that lays out our expectations of scientists who work in our community, and an accompanying questionnaire.

By filling in the one-page questionnaire, researchers give us a quick overview of what they plan to do; its relevance and potential benefit to our community; the need for local involvement; and how we'll be compensated financially. This provides us with a tool through which to develop relationships with researchers, make sure that our priorities and rights are addressed, and hold researchers accountable. Making scientists think about how they'll engage with us has helped to make research a more equitable, less extractive activity.

We cannot force scientists to deal with us. It's a free country. But the Qikiktagrukmiut are skilled at activities such as boating, travelling on snow and capturing animals — and those skills are extremely useful for fieldwork, as is our deep historical knowledge of the local environment. It's a lot harder for scientists to accomplish their work without our involvement. Many scientists realize this, so these days we get 6–12 research proposals per year. We say yes to most of them.

The NSF's Navigating the New Arctic programme has definitely increased the number of last-minute proposals that communities such as ours get swamped with a couple of weeks before the application deadline. Throwing an Indigenous component into a research proposal at the last minute is definitely not an ideal way to go about things, because it doesn't give us time to fully consider the research before deciding whether we want to participate. But at least the NSF has recognized that working with Indigenous people is a thing! They're just in the growing-pains phase.

Not all Indigenous groups have had as much success as we have, and some are still experiencing the extractive side of science. But incorporating Indigenous knowledge into science can create rapid growths in understanding, and we're happy we've helped some researchers do this in a respectful way.

NATAN OBED: Fund research on Indigenous priorities

Natan Obed is president of Inuit Tapiriit Kanatami, and a member of the Inuit community.

Every year, funding agencies devote hundreds of millions of dollars to work that occurs in the Inuit homeland in northern Canada. Until very recently, almost none of those agencies considered Inuit peoples' priorities.

These Indigenous communities face massive social and economic challenges. More than 60% of Inuit households are food insecure, meaning they don't always have enough food to maintain an active, healthy life. On average, one-quarter as many doctors serve Inuit communities as serve urban Canadian communities. Our life expectancy is ten years less than the average non-Indigenous Canadian's. The list goes on. And yet, very little research is devoted to addressing these inequities.

Last year, the Inuit advocacy organization Inuit Tapiriit Kanatami (the name means 'Inuit are united in Canada') collaborated with the research network ArcticNet to start its own funding programme, which is called the Inuit Nunangat Research Program (INRP). Funding decisions are led entirely by Inuit people to ensure that all grants support research on Inuit priorities. Even in the programme's first year, we got more requests than we could fund. We selected 11 proposals that all relate directly to the day-to-day lives of Inuit people. For example, one study that we're funding aims to characterize a type of goose that has newly arrived in northern Labrador; another focuses on how social interactions spread disease in Inuit communities.

Our goal with the INRP is twofold: first, we want to generate knowledge that addresses Inuit concerns, and second, we want to create an example of how other granting agencies can change so that they respect the priorities of all groups. We've been moderately successful in getting some of the main Canadian granting agencies, such as the Canadian Institutes of Health Research, to allocate more resources to things that matter to Inuit people. I'd like to think that the INRP gives them a model for how to become even more inclusive.

We hope that, over the next ten years, it will become normal for granting agencies to consider the needs of Indigenous communities. But we also know that institutions change slowly. Looking back at where we've been, we have a lot to be proud of, but we still have a huge task ahead of us.

Nature **601**, 285-287 (2022)

doi: <https://doi.org/10.1038/d41586-022-00029-2>

These interviews have been edited for length and clarity.

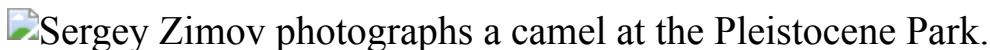
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00029-2>

- WHERE I WORK
- 10 January 2022

Pleistocene dreams: recreating ancient grasslands to save the planet

Sergey Zimov is rewilding a remote part of northeastern Siberia to help fight climate change.

- [Olga Dobrovidova](#) 0



Sergey Zimov is founding director of the Northeast Science Station in Cherskii, Russia. Credit: Maxim Shemetov/Reuters

I am a geophysicist and founding director of the Northeast Science Station in Cherskii, Russia — a settlement of about 2,500 people in the delta of the Kolyma River in northeastern Siberia, beyond the Arctic Circle. Annual temperatures here average around -10°C . For the past 30 years, my team has been trying to recreate the grassland ecosystem that existed here tens of thousands of years ago by repopulating the area with large herbivores, such as this camel. The area we're rewilding covers 20 square kilometres, and we call it Pleistocene Park.

Our aim is to investigate how we can affect the climate through ecosystem reconstruction. The project requires us to deal with a lot of societal and political issues. Our civilization is not able or willing to share living space with wild ecosystems: we tend to destroy them and then declare their pitiful remains ‘natural reserves’. If you treat nature this way, it’s hard to get it to help us save the climate.

We work with teams at the University of Alaska Fairbanks in the United States and the Max Planck Institute for Biogeochemistry in Jena, Germany, to track gas fluxes and other climate data from the ecosystem. In that way, we can see just how much the planet is warming. The winter here in Cherskii has already warmed by 6 or 7 degrees on average; we've even had winter rains.

This winter I'm spending my time writing papers and popular-science articles, as well as doing weekly visits to the park. The long polar night is upon us, and right now, as far as the eye can see — and that's about 40 kilometres — there are no lights.

I took this photograph in warmer times, in September last year. We want the animals here to be wild, not buddy-buddy with humans, and we don't normally name them. But this camel is very tame because my son, Nikita, fed the herd on the long journey here. We run another park, 300 kilometres south of Moscow, and I wouldn't dare get this close to the camels we have there.

Nature **601**, 290 (2022)

doi: <https://doi.org/10.1038/d41586-022-00030-9>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00030-9>

Research

- **[Extreme rainfall slows the global economy](#)** [12 January 2022]
News & Views • Excessive rainfall can cause catastrophic socio-economic losses to a community or nation. An analysis of changes in gross regional product identifies ways in which extreme precipitation affects global economic productivity.
- **[Transporter-protein structures show how salt gets a sweet ride into cells](#)** [08 December 2021]
News & Views • Proteins spanning the membranes of cells of the intestine and kidney use sodium-ion gradients to take up glucose, enabling water absorption, too. The structures of these transporter proteins have now been observed in detail.
- **[A microbe that uses crude oil to make methane](#)** [22 December 2021]
News & Views • A microorganism that dwells in an underground oil reservoir has been found to degrade various petroleum compounds and use them to produce methane through a previously unreported biochemical pathway.
- **[A strange metal emerges from a failed superconductor](#)** [12 January 2022]
News & Views • The curious electrical resistance that gives strange metals their name has been seen in a failed superconductor, in which disorder interferes with the material's ability to achieve zero resistance below a critical temperature.
- **[Beads reveal long-distance connections in early Africa](#)** [20 December 2021]
News & Views • Beads made from ostrich eggshells, produced by people over the past 50,000 years, provide evidence for a long period of social connection between eastern and southern Africa, followed by isolation and then reconnection.
- **[A WC/WO star exploding within an expanding carbon–oxygen–neon nebula](#)** [12 January 2022]
Article • Observations of the supernova SN 2019hgp, identified about a day after its explosion, show that it occurred within a nebula of carbon, oxygen and neon, and was probably the explosion of a massive WC/WO star.
- **[Signatures of a strange metal in a bosonic system](#)** [12 January 2022]
Article • Strange metallicity—in particular, resistance that is linear in temperature and magnetic field—is observed in a nanopatterned $\text{YBa}_2\text{Cu}_3\text{O}_7-\delta$ bosonic system.

- **A crossbar array of magnetoresistive memory devices for in-memory computing** [12 January 2022]

Article • A crossbar array of magnetic memory to execute analogue in-memory computing has been developed, and performs image classification and facial detection at low power.
- **Elastomeric electrolytes for high-energy solid-state lithium batteries** [12 January 2022]

Article • An elastomeric solid-state electrolyte shows desirable mechanical properties and high electrochemical stability, and is used to demonstrate a high-energy solid-state lithium battery at ambient temperature.
- **The effect of rainfall changes on economic production** [12 January 2022]

Article • A global assessment shows that increases in the number of wet days and extreme daily rainfall adversely affect economic growth, particularly in high-income nations and via the services and manufacturing sectors.
- **Air pollution exposure disparities across US population and income groups** [12 January 2022]

Article • Different racial/ethnic populations and income groups are found to have been exposed to different levels of air pollution in the USA during the years 2000 to 2016.
- **Ostrich eggshell beads reveal 50,000-year-old social network in Africa** [20 December 2021]

Article • By tracing the changing size of ostrich eggshell beads, climate is shown to have an important role in influencing when and where regional African populations interacted.
- **Adaptive stimulus selection for consolidation in the hippocampus** [08 December 2021]

Article • In memory consolidation, the hippocampus has a unique way to preferentially amplify behaviour-relevant information that entails ‘replaying’ this information during periods of rest.
- **NLRs guard metabolism to coordinate pattern- and effector-triggered immunity** [15 December 2021]

Article • The deubiquitinase PIC11 is identified as part of an immunity hub that coordinates pattern- and effector-triggered immunity and is involved in conferring broad-spectrum resistance to blast across different subspecies of rice.
- **Towards the biogeography of prokaryotic genes** [15 December 2021]

Article • A survey of species-level genes from 13,174 publicly available metagenomes shows that most species-level genes are specific to a single habitat, encode a small number of protein families and are under low positive (adaptive) pressure.

- **Non-syntrophic methanogenic hydrocarbon degradation by an archaeal species** [22 December 2021]

Article • ‘Candidatus Methanoliparum’ overexpresses genes encoding alkyl-coenzyme M and methyl-coenzyme M reductases—markers of archaeal multicarbon alkane and methane metabolism—and thrives on a variety of long-chain alkanes and n-alkylcyclohexanes, and n-alkylbenzenes with long n-alkyl ($C \geq 13$) moieties.

- **Cancer risk across mammals** [22 December 2021]

Article • An analysis of cancer mortality data for zoo mammals highlights marked differences across mammalian orders and an influence of diet, and shows that mortality risk is largely independent of body mass and life expectancy across species.

- **HELQ is a dual-function DSB repair enzyme modulated by RPA and RAD51** [22 December 2021]

Article • HELQ is differentially regulated by RAD51, which stimulates helicase activity, and RPA, which inhibits helicase activity and stimulates annealing.

- **Structure and mechanism of the SGLT family of glucose transporters** [08 December 2021]

Article • Cryo-electron microscopy structures of the sodium–glucose cotransporter SGLT1 and a related transporter SMCT1 define the architecture of this protein family and provide insights into substrate binding and transport function.

- **Structural basis of inhibition of the human SGLT2–MAP17 glucose transporter** [08 December 2021]

Article • Using cryogenic electron microscopy, the structure of the human SGLT2–MAP17 complex captured in the empagliflozin-bound state reveals the inhibitory mechanism of these anti-diabetic drugs.

- NEWS AND VIEWS
- 12 January 2022

Extreme rainfall slows the global economy

Excessive rainfall can cause catastrophic socio-economic losses to a community or nation. An analysis of changes in gross regional product identifies ways in which extreme precipitation affects global economic productivity.

- [Xin-Zhong Liang](#) 

In July 2021, record-breaking rainfall brought severe floods to Europe, where 200,000 properties lost electrical power. In the same month, torrential rain with a maximum intensity of 201.9 millimetres in a single hour led to devastating floods (Fig. 1) in Henan province, China, forcing more than one million people to relocate. These flooding events each caused roughly US\$12 billion in property damage. Such losses, incurred during or shortly after extreme events, represent a direct negative impact on the economy. But how does excessive precipitation affect macroeconomics indirectly in the longer term? [Writing in Nature](#), Kotz *et al.*¹ report a comprehensive assessment of changes in gross regional product (GRP) relating to excessive precipitation, and conclude that increases in the numbers of wet days and in extreme daily rainfall dramatically reduce worldwide macroeconomic growth rates.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 193-194 (2022)

doi: <https://doi.org/10.1038/d41586-021-03783-x>

References

1. 1.
Kotz, M., Levermann, A. & Wenz, L. *Nature* **601**, 223–227 (2022).
2. 2.
Botzen, W. J. W., Deschenes, O. & Sanders, M. *Rev. Environ. Econ. Policy* **13**, 167–188 (2019).
3. 3.
Li, Y., Guan, K., Schnitkey, G. D., DeLucia, E. & Peng, B. *Glob. Change Biol.* **25**, 2325–2337 (2019).
4. 4.
Vogel, E. *et al.* *Environ. Res. Lett.* **14**, 054010 (2019).
5. 5.
Liang, X.-Z. *et al.* *Proc. Natl Acad. Sci. USA* **114**, E2285–E2292 (2017).
6. 6.
Ortiz-Bobea, A., Ault, T. R., Carrillo, C. M., Chambers, R. G. & Lobell, D. B. *Nature Clim. Change* **11**, 306–312 (2021).
7. 7.
Boers, N. *et al.* *Nature* **566**, 373–377 (2019).
8. 8.
IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the*

Intergovernmental Panel on Climate Change (eds Masson-Delmotte, V. et al.) (Cambridge Univ. Press, in the press).

9. 9.

Sun, C. & Liang, X.-Z. *Clim. Dyn.* **55**, 1325–1352 (2020).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03783-x>

| [Section menu](#) | [Main menu](#) |

- NEWS AND VIEWS
- 08 December 2021

Transporter-protein structures show how salt gets a sweet ride into cells

Proteins spanning the membranes of cells of the intestine and kidney use sodium-ion gradients to take up glucose, enabling water absorption, too. The structures of these transporter proteins have now been observed in detail.

- [David Drew](#) ⁰

I was close — oh, so close — when I collapsed just outside the stadium with less than one kilometre of the marathon to go. As a biochemistry student, I should have known better, but I had made the mistake of selecting water instead of the electrolyte drinks offered, and so ended up inside an ambulance dehydrated and embarrassed. The discovery that the transport of glucose and sodium ions from the digestive tract into the body is coupled and facilitates the absorption of water was a breakthrough of the twentieth century¹. A simple salt–glucose mixture has since saved millions of lives, and kept runners from collapsing — well, mostly. Writing in *Nature*, [Han et al.](#)² and [Niu et al.](#)³ present structures of the membrane-spanning transporter proteins that carry sodium and glucose across cell membranes in humans.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 194-196 (2022)

doi: <https://doi.org/10.1038/d41586-021-03555-7>

References

1. 1.
Lancet **312**, 300–301 (1978).
2. 2.
Han, L. *et al.* *Nature* **601**, 274–279 (2022).
3. 3.
Niu, Y. *et al.* *Nature* **601**, 280–284 (2022).
4. 4.
Hediger, M., Coady, M. J., Ikeda, T. S. & Wright, E. M. *Nature* **330**, 379–381 (1987).
5. 5.
Ikeda, T. S. *et al.* *J. Membr. Biol.* **110**, 87–95 (1989).
6. 6.
Wright, E. M., Ghezzi, C. & Loo, D. D. F. *Physiology* **32**, 435–443 (2017).
7. 7.
Wright, E. M. *Kidney360* <https://doi.org/10.34067/KID.0002772021> (2021).
8. 8.
Loo, D. F., Zeuthen, T., Chandy, G. & Wright, E. M. *Proc. Natl Acad. Sci. USA* **93**, 13367–13370 (1996).
9. 9.

Coady, M. J. *et al.* *J. Am. Soc. Nephrol.* **28**, 85–93 (2017).

10. 10.

Faham, S. *et al.* *Science* **321**, 810–814 (2008).

11. 11.

Drew, D. & Boudker, O. *Annu. Rev. Biochem.* **85**, 543–572 (2016).

12. 12.

Miyauchi, S., Gopal, E., Fei, Y.-J. & Ganapathy, V. *J. Biol. Chem.* **279**, 13293–13296 (2004).

13. 13.

Coleman, J. A., Green, E. M. & Gouaux, E. *Nature* **532**, 334–339 (2016).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03555-7>

| [Section menu](#) | [Main menu](#) |

- NEWS AND VIEWS
- 22 December 2021

A microbe that uses crude oil to make methane

A microorganism that dwells in an underground oil reservoir has been found to degrade various petroleum compounds and use them to produce methane through a previously unreported biochemical pathway.

- [Guillaume Borrel](#) [ORCID: http://orcid.org/0000-0003-4893-8180](#) ⁰

Microbial communities tend to use the most energy-rich and most easily metabolized compounds that they have at their disposal. This leads to a progressive enrichment of compounds that are difficult to break down and that provide little energy, particularly in the absence of oxygen or other inorganic electron acceptors. Under these conditions, the use of hydrocarbons — molecules consisting of carbon and hydrogen, such as alkanes — has been thought to rely entirely on a collaboration (known as syntrophy) between bacteria that break down these compounds into acetate and molecular hydrogen (H_2), and microorganisms called methanogenic archaea that use the molecules to produce methane (CH_4), the simplest hydrocarbon^{1–3}. [Writing in Nature](#), Zhou *et al.*⁴ overturn this long-standing account of a division of labour in the methanogenic degradation of hydrocarbons by reporting that a single type of microorganism can degrade various large hydrocarbons into methane (Fig. 1).

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 196-197 (2022)

doi: <https://doi.org/10.1038/d41586-021-03729-3>

References

1. 1.

Zengler, K., Richnow, H. H., Rosselló-Mora, R., Michaelis, W. & Widdel, F. *Nature* **401**, 266–269 (1999).

2. 2.

Schink, B. *Microbiol. Mol. Biol. Rev.* **61**, 262–280 (1997).

3. 3.

Jones, D. M. *et al.* *Nature* **451**, 176–180 (2008).

4. 4.

Zhou, Z. *et al.* *Nature* **601**, 257–262 (2022).

5. 5.

Borrel, G. *et al.* *Nature Microbiol.* **4**, 603–613 (2019).

6. 6.

Nobu, M. K., Narihiro, T., Kuroda, K., Mei, R. & Liu, W.-T. *ISME J.* **10**, 2478–2487 (2016).

7. 7.

Vanwonterghem, I. *et al.* *Nature Microbiol.* **1**, 16170 (2016).

8. 8.

Sorokin, D. Y. *et al.* *Nature Microbiol.* **2**, 17081 (2017).

9. 9.

Söllinger, A. & Urich, T. *Biochem. Soc. Trans.* **47**, 1895–1907 (2019).

10. 10.

Laso-Pérez, R. *et al.* *mBio* **10**, e01814-19 (2019).

11. 11.

Laso-Pérez, R. *et al.* *Nature* **539**, 396–401 (2016).

12. 12.

Chen, S.-C. *et al.* *Nature* **568**, 108–111 (2019).

13. 13.

Hahn, C. J. *et al.* *Science* **373**, 118–121 (2021).

14. 14.

Shima, S. *et al.* *Nature* **481**, 98–101 (2012).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03729-3>

| [Section menu](#) | [Main menu](#) |

- NEWS AND VIEWS
- 12 January 2022

A strange metal emerges from a failed superconductor

The curious electrical resistance that gives strange metals their name has been seen in a failed superconductor, in which disorder interferes with the material's ability to achieve zero resistance below a critical temperature.

- [Nicholas P. Breznay](#) ⁰

Above a certain temperature, the electrical resistance of copper–oxygen ceramic materials (called cuprates) increases in direct proportion to their temperature, reaching values far above the reasonable limits of electrical conduction¹. This behaviour cannot be attributed to the way electrons usually interact with one another, or with atomic nuclei, and it has no widely accepted explanation — a trait that has given these materials the name of bad, or strange, metals. [Writing in Nature](#), Yang *et al.*² report this strange-metallic behaviour in a thin-film cuprate containing yttrium, barium, copper and oxygen ($\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$), in which nanopatterning enhances the disorder present in the material. Moreover, they find that this behaviour occurs at lower temperatures than expected, redrafting the accepted strange-metal blueprint.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 198-199 (2022)

doi: <https://doi.org/10.1038/d41586-021-03831-6>

References

1. 1.

Gurvitch, M. & Fiory, A. T. *Phys. Rev. Lett.* **59**, 1337–1340 (1987).

2. 2.

Yang, C. *et al.* *Nature* **601**, 205–210 (2022).

3. 3.

Bednorz, J. G. & Müller, K. A. *Z. Phys. B* **64**, 189–193 (1986).

4. 4.

Kasahara, S. *et al.* *Phys. Rev. B* **81**, 184519 (2010).

5. 5.

Bruin, J. A. N., Sakai, H., Perry, R. S. & Mackenzie, A. P. *Science* **339**, 804–807 (2013).

6. 6.

Cao, Y. *et al.* *Phys. Rev. Lett.* **124**, 076801 (2020).

7. 7.

Keimer, B., Kivelson, S. A., Norman, M. R., Uchida, S. & Zaanen, J. *Nature* **518**, 179–186 (2015).

8. 8.

Hayes, I. M. *et al.* *Nature Phys.* **12**, 916–919 (2016).

9. 9.

Goldman, A. M. & Marković, N. *Phys. Today* **51**, 39–44 (1998).

10. 10.

Varma, C. M., Littlewood, P. B., Schmitt-Rink, S., Abrahams, E. & Ruckenstein, A. E. *Phys. Rev. Lett.* **63**, 1996–1999 (1989).

11. 11.

Sachdev, S. & Ye, J. *Phys. Rev. Lett.* **70**, 3339–3342 (1993).

12. 12.

Stewart, M. D. Jr, Yin, A., Xu, J. M. & Valles, J. M. Jr *Science* **318**, 1273–1275 (2007).

13. 13.

McGreevy, J. *Physics* **3**, 83 (2010).

14. 14.

Hartnoll, S. A. *Nature Phys.* **11**, 54–61 (2015).

15. 15.

Kapitulnik, A., Kivelson, S. A. & Spivak, B. *Rev. Mod. Phys.* **91**, 011002 (2019).

16. 16.

Yang, C. *et al.* *Science* **366**, 1505–1509 (2019).

17. 17.

Li, Y. *et al.* Preprint at <https://arxiv.org/abs/2111.15488> (2021).

- NEWS AND VIEWS
- 20 December 2021

Beads reveal long-distance connections in early Africa

Beads made from ostrich eggshells, produced by people over the past 50,000 years, provide evidence for a long period of social connection between eastern and southern Africa, followed by isolation and then reconnection.

- [Benjamin R. Collins](#) ORCID: <http://orcid.org/0000-0002-2416-3184>
⁰ &
- [Amy Hatton](#) ORCID: <http://orcid.org/0000-0002-2900-3681>¹

“Let’s keep in touch!” is a phrase commonly used to indicate a social bond formed between people. Ideally, such bonds are maintained even across long distances and over the passage of time. Today, keeping in touch is easy, whether through a call, a text message, social media or the now omnipresent virtual meeting. However, what traces might remain of signs of social relationships from 50,000 years ago, and why is investigating these ancient relationships useful? [Writing in Nature](#), Miller and Wang¹ present data suggesting that small beads made of ostrich eggshell (OES) and fashioned into jewellery were exchanged between groups across eastern and southern Africa as part of long-distance social connections over the past 50,000 years. These relationships, across immense distances, then broke down about 33,000 years ago — around the same time as major climate changes occurred — and were renewed only about 2,000 years ago, the authors suggest.

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Nature **601**, 199–200 (2022)

doi: <https://doi.org/10.1038/d41586-021-03681-2>

References

1. 1.

Miller, J. M. & Wang, Y. V. *Nature* **601**, 234–239 (2022).

2. 2.

Hitchcock, R. K. *Botsw. Notes Rec.* **44**, 93–105 (2012).

3. 3.

Wiessner, P. W. *Hxaro: A Regional System of Reciprocity for Reducing Risk among the !Kung San* Vol. 1 (Univ. Michigan Press, 1977).

4. 4.

Wiessner, P. in *Politics and History in Band Societies* 61–84 (Cambridge Univ. Press, 1982).

5. 5.

Miller, J. M. & Sawchuk, E. A. *PloS ONE* **14**, e0225143 (2019).

6. 6.

Collins, B. in *Oxford Research Encyclopedia of Anthropology* (Oxford Univ. Press, 2021).

7. 7.

Stewart, B. A. *et al. Proc. Natl Acad. Sci. USA* **117**, 6453–6462 (2020).

8. 8.

van der Lubbe, H. J. L., Frank, M., Tjallingii, R. & Schneider, R. R. *Geochem. Geophys. Geosyst.* **17**, 181–198 (2016).

9. 9.

Kivisto, S. *Evaluating Paleoenvironmental and Landscape Mobility Dynamics: Stable Isotope and Strontium Isotope Analyses of Ostrich Eggshell at Spitzkloof Rockshelter, South Africa*. Master's thesis, Univ. Toronto (2016).

10. 10.

Miller, J. M. *Variability in Ostrich Eggshell Beads from the Middle and Later Stone Age of Africa*. PhD thesis, Univ. Alberta (2019).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03681-2>

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 12 January 2022](#)

A WC/WO star exploding within an expanding carbon–oxygen–neon nebula

- [A. Gal-Yam](#) [ORCID: orcid.org/0000-0002-3653-5598](#)¹,
- [R. Bruch](#)¹,
- [S. Schulze](#) [ORCID: orcid.org/0000-0001-6797-1889](#)^{1,2},
- [Y. Yang](#)^{1,3},
- [D. A. Perley](#)⁴,
- [I. Irani](#)¹,
- [J. Sollerman](#) [ORCID: orcid.org/0000-0003-1546-6615](#)²,
- [E. C. Kool](#) [ORCID: orcid.org/0000-0002-7252-3877](#)²,
- [M. T. Soumagnac](#) [ORCID: orcid.org/0000-0001-6753-1488](#)^{1,5},
- [O. Yaron](#)¹,
- [N. L. Strotjohann](#)¹,
- [E. Zimmerman](#)¹,
- [C. Barbarino](#)²,
- [S. R. Kulkarni](#)⁶,
- [M. M. Kasliwal](#)⁶,
- [K. De](#)⁶,
- [Y. Yao](#) [ORCID: orcid.org/0000-0001-6747-8509](#)⁶,
- [C. Fremling](#) [ORCID: orcid.org/0000-0002-4223-103X](#)⁶,
- [L. Yan](#)⁶,
- [E. O. Ofek](#)¹,
- [C. Fransson](#)²,
- [A. V. Filippenko](#)^{3,7},
- [W. Zheng](#)³,

- [T. G. Brink³](#),
- [C. M. Copperwheat⁴](#),
- [R. J. Foley⁸](#),
- [J. Brown⁸](#),
- [M. Siebert⁸](#),
- [G. Leloudas](#) ORCID: [orcid.org/0000-0002-8597-0756⁹](https://orcid.org/0000-0002-8597-0756),
- [A. L. Cabrera-Lavers¹⁰](#),
- [D. Garcia-Alvarez¹⁰](#),
- [A. Marante-Barreto¹⁰](#),
- [S. Frederick¹¹](#),
- [T. Hung](#) ORCID: [orcid.org/0000-0002-9878-7889⁸](https://orcid.org/0000-0002-9878-7889),
- [J. C. Wheeler¹²](#),
- [J. Vinkó^{12,13,14,15}](#),
- [B. P. Thomas¹²](#),
- [M. J. Graham](#) ORCID: [orcid.org/0000-0002-3168-0139⁶](https://orcid.org/0000-0002-3168-0139),
- [D. A. Duev](#) ORCID: [orcid.org/0000-0001-5060-8733⁶](https://orcid.org/0000-0001-5060-8733),
- [A. J. Drake⁶](#),
- [R. Dekany](#) ORCID: [orcid.org/0000-0002-5884-7867⁶](https://orcid.org/0000-0002-5884-7867),
- [E. C. Bellm](#) ORCID: [orcid.org/0000-0001-8018-5348¹⁶](https://orcid.org/0000-0001-8018-5348),
- [B. Rusholme](#) ORCID: [orcid.org/0000-0001-7648-4142¹⁷](https://orcid.org/0000-0001-7648-4142),
- [D. L. Shupe](#) ORCID: [orcid.org/0000-0003-4401-0430¹⁷](https://orcid.org/0000-0003-4401-0430),
- [I. Andreoni⁶](#),
- [Y. Sharma⁶](#),
- [R. Riddle](#) ORCID: [orcid.org/0000-0002-0387-370X⁶](https://orcid.org/0000-0002-0387-370X),
- [J. van Roestel](#) ORCID: [orcid.org/0000-0002-2626-2872⁶](https://orcid.org/0000-0002-2626-2872) &
- [N. Knezevic¹⁸](#)

Nature volume **601**, pages 201–204 (2022)

- 886 Accesses
- 108 Altmetric
- [Metrics details](#)

Subjects

- [Time-domain astronomy](#)
- [Transient astrophysical phenomena](#)

Abstract

The final fate of massive stars, and the nature of the compact remnants they leave behind (black holes and neutron stars), are open questions in astrophysics. Many massive stars are stripped of their outer hydrogen envelopes as they evolve. Such Wolf–Rayet stars¹ emit strong and rapidly expanding winds with speeds greater than 1,000 kilometres per second. A fraction of this population is also helium-depleted, with spectra dominated by highly ionized emission lines of carbon and oxygen (types WC/WO). Evidence indicates that the most commonly observed supernova explosions that lack hydrogen and helium (types Ib/Ic) cannot result from massive WC/WO stars^{2,3}, leading some to suggest that most such stars collapse directly into black holes without a visible supernova explosion⁴. Here we report observations of SN 2019hgp, beginning about a day after the explosion. Its short rise time and rapid decline place it among an emerging population of rapidly evolving transients^{5,6,7,8}. Spectroscopy reveals a rich set of emission lines indicating that the explosion occurred within a nebula composed of carbon, oxygen and neon. Narrow absorption features show that this material is expanding at high velocities (greater than 1,500 kilometres per second), requiring a compact progenitor. Our observations are consistent with an explosion of a massive WC/WO star, and suggest that massive Wolf–Rayet stars may be the progenitors of some rapidly evolving transients.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Fig. 1: Spectra of SN 2019hgp are dominated by carbon, oxygen and neon.

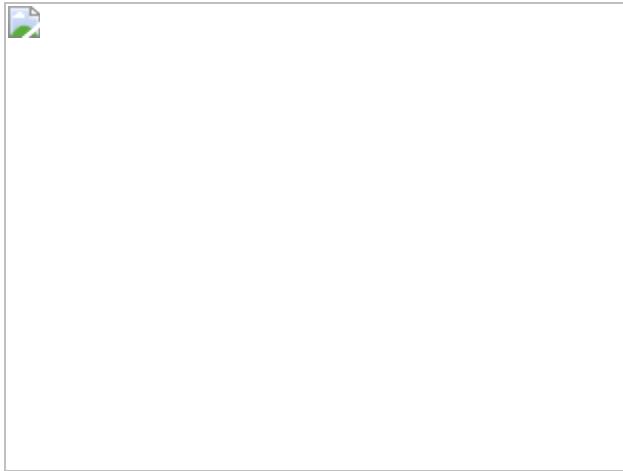


Fig. 2: SN 2019hgp is a rapidly rising, fairly luminous transient.

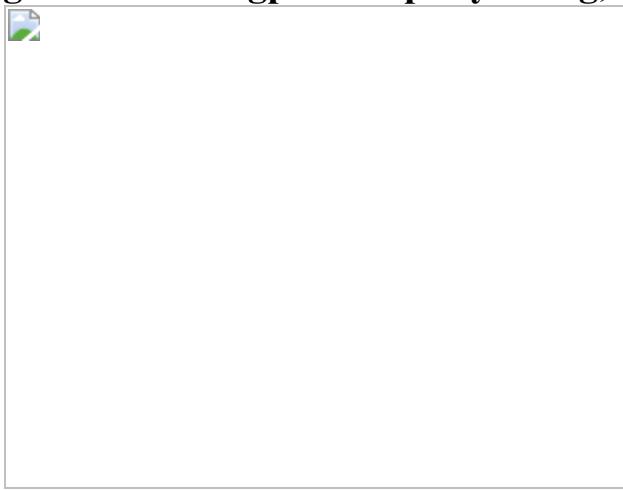


Fig. 3: The bolometric evolution of SN 2019hgp shows rapid cooling from an initial hot phase.

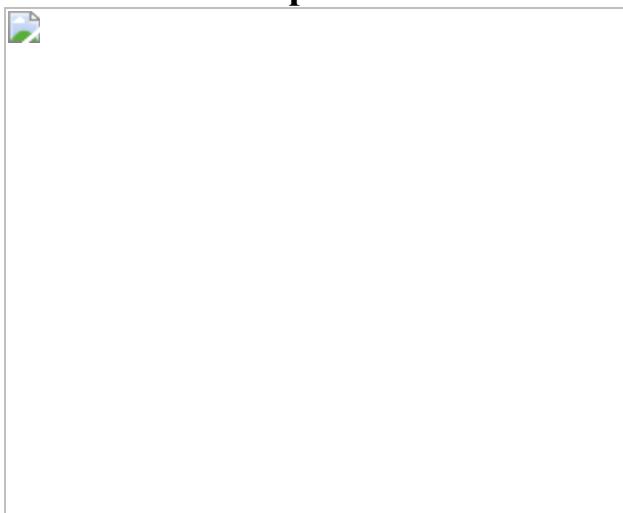
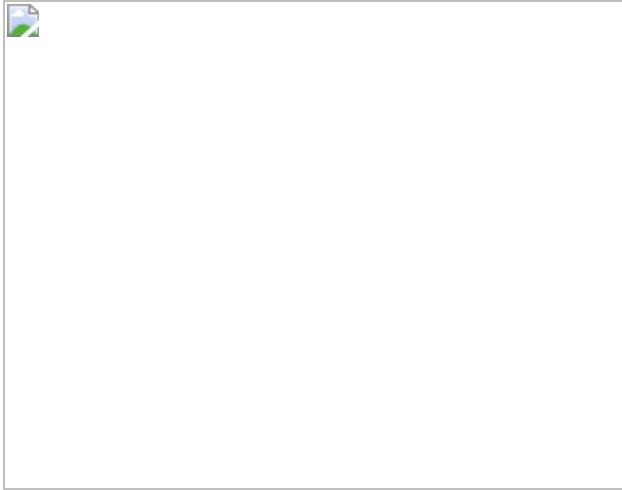


Fig. 4: SN 2019hgp is embedded in a rapidly expanding nebula.



Data availability

The photometry of SN 2019hgp is available in Supplementary Table 1, and all the observations (photometry and spectra) are available from WISeREP⁵² (<http://wiserep.weizmann.ac.il/>). Matlab scripts that generate most of the plots within this paper are available from the corresponding author upon request. Opticon observations were obtained under programme ID OPT/2019A/024, PI A.G.-Y.

Code availability

Relevant software sources have been provided in the text, web locations provided as references, and are publicly available.

References

1. 1.

Crowther, P. A. Physical properties of Wolf–Rayet stars. *Annu. Rev. Astron. Astrophys.* **45**, 177–219 (2007).

2. 2.

Smith, N., Li, W., Filippenko, A. V. & Chornock, R. Observed fractions of core-collapse supernova types and initial masses of their single and binary progenitor stars. *Mon. Not. R. Astron. Soc.* **412**, 1522–1538 (2011).

3. 3.

Taddia, F. et al. The Carnegie Supernova Project I: analysis of stripped-envelope supernova light curves. *Astron. Astrophys.* **609**, A136 (2018).

4. 4.

Dessart, L. et al. Core-collapse explosions of Wolf–Rayet stars and the connection to type IIb/Ib/Ic supernovae. *Mon. Not. R. Astron. Soc.* **414**, 2985–3005 (2011).

5. 5.

Drout, M. R. et al. Rapidly evolving and luminous transients from Pan-STARRS1. *Astrophys. J.* **794**, 23 (2014).

6. 6.

Arcavi, I. et al. Rapidly rising transients in the supernova–superluminous supernova gap. *Astrophys. J.* **819**, 35 (2016).

7. 7.

Pursiainen, M. et al. Rapidly evolving transients in the Dark Energy Survey. *Mon. Not. R. Astron. Soc.* **481**, 894–971 (2018).

8. 8.

Perley, D. A. et al. The Zwicky Transient Facility Bright Transient Survey. II. A public statistical sample for exploring supernova demographics. *Astrophys. J.* **904**, 35 (2020).

9. 9.

Bellm, E. C. et al. The Zwicky Transient Facility: system overview, performance, and first results. *Publ. Astron. Soc. Pac.* **131**, 018002 (2019).

10. 10.

Gal-Yam, A. et al. Real-time detection and rapid multiwavelength follow-up observations of a highly subluminous type II-P supernova from the Palomar Transient Factory Survey. *Astrophys. J.* **736**, 159 (2011).

11. 11.

Kool, E. C. et al. SN 2020bqj: a type Ibn supernova with a long-lasting peak plateau. *Astron. Astrophys.* **652**, A136 (2021).

12. 12.

Arnett, W. D. Type I supernovae. I—Analytic solutions for the early part of the light curve. *Astrophys. J.* **253**, 785–797 (1982).

13. 13.

Drout, M. R. et al. The first systematic study of type Ibc supernova multi-band light curves. *Astrophys. J.* **741**, 97 (2011); erratum *Astrophys. J.* **753**, 180 (2012).

14. 14.

Prentice, S. J. et al. Investigating the properties of stripped-envelope supernovae; what are the implications for their progenitors? *Mon. Not. R. Astron. Soc.* **485**, 1559–1578 (2019).

15. 15.

Chatzopoulos, E., Wheeler, J. C. & Vinko, J. Generalized semi-analytical models of supernova light curves. *Astrophys. J.* **746**, 121 (2012).

16. 16.

Rabinak, I. & Waxman, E. The early UV/optical emission from core-collapse supernovae. *Astrophys. J.* **728**, 63 (2011).

17. 17.

Janka, H.-T. Explosion mechanisms of core-collapse supernovae. *Annu. Rev. Nucl. Part. Sci.* **62**, 407–451 (2012).

18. 18.

Filippenko, A. V. Optical spectra of supernovae. *Annu. Rev. Astron. Astrophys.* **35**, 309–355 (1997).

19. 19.

Fremling, C. et al. The Zwicky Transient Facility Bright Transient Survey. I. Spectroscopic classification and the redshift completeness of local galaxy catalogs. *Astrophys. J.* **895**, 32 (2020).

20. 20.

Foley, R. J. et al. SN 2006jc: a Wolf–Rayet star exploding in a dense He-rich circumstellar medium. *Astrophys. J.* **657**, L105 (2007).

21. 21.

Pastorello, A. et al. Massive stars exploding in a He-rich circumstellar medium—I. Type Ibn (SN 2006jc-like) events. *Mon. Not. R. Astron. Soc.* **389**, 113–130 (2008).

22. 22.

Karamehmetoglu, E. et al. OGLE-2014-SN-131: a long-rising type Ibn supernova from a massive progenitor. *Astron. Astrophys.* **602**, A93 (2017).

23. 23.

Gal-Yam, A. Observational and physical classification of supernovae. In *Handbook of Supernovae* (eds Alsabti, A. W. & Murdin, P.) 195–237 (2017); https://doi.org/10.1007/978-3-319-21846-5_35.

24. 24.

Hosseinzadeh, G. et al. Type Ibn supernovae show photometric homogeneity and spectral diversity at maximum light. *Astrophys. J.* **836**, 158 (2017).

25. 25.

Pastorello, A. et al. Massive stars exploding in a He-rich circumstellar medium—IX. SN 2014av, and characterization of type Ibn SNe. *Mon. Not. R. Astron. Soc.* **456**, 853–869 (2016).

26. 26.

Perley, D. et al. ZTF/LT discovery of a fast, luminous blue transient with narrow carbon features. *Transient Name Server AstroNote* 2021-62 (2021); <https://www.wis-tns.org/astronotes/2021-62>.

27. 27.

Pastorello, A. et al. adH0cc spectroscopic classification of SN 2021ckj, an unusual “Type Icn” supernova. *Transient Name Server AstroNote* 2021-71 (2021); <https://www.wis-tns.org/astronotes/2021-71>.

28. 28.

Gal-Yam, A. et al. Introducing a new supernova classification type: SN Icn. *Transient Name Server AstroNote* 2021-76 (2021); <https://www.wis-tns.org/astronotes/2021-76>.

29. 29.

Gal-Yam, A. A simple analysis of type I superluminous supernova peak spectra: composition, expansion velocities, and dynamics. *Astrophys. J.* **882**, 102 (2019).

30. 30.

Sharon, A. & Kushnir, D. The γ -ray deposition histories of core-collapse supernovae. *Mon. Not. R. Astron. Soc.* **496**, 4517–4545 (2020).

31. 31.

Graham, M. J. et al. The Zwicky Transient Facility: science objectives. *Publ. Astron. Soc. Pac.* **131**, 078001 (2019).

32. 32.

Yaron, O., Gal-Yam, A., Ofek, E. & Sasse, A. The revised treatment of object coordinates (astrometric accuracies) is now active on the TNS. *Transient Name Server AstroNote* 2019-37 (2019); <https://www.wis-tns.org/astronotes/astronote/2019-37>.

33. 33.

Dekany, R. et al. The Zwicky Transient Facility: observing system. *Publ. Astron. Soc. Pac.* **132**, 038001 (2020).

34. 34.

Masci, F. J. et al. The Zwicky Transient Facility: data processing, products, and archive. *Publ. Astron. Soc. Pac.* **131**, 018003 (2019).

35. 35.

Zackay, B., Ofek, E. O. & Gal-Yam, A. Proper image subtraction—optimal transient detection, photometry, and hypothesis testing. *Astrophys. J.* **830**, 27 (2016).

36. 36.

Gal-Yam, A. Infant supernovae from ZTF. In *233rd American Astronomical Society Meeting* 131.06 (2019).

37. 37.

Kasliwal, M. M. et al. The GROWTH Marshal: a dynamic science portal for time-domain astronomy. *Publ. Astron. Soc. Pac.* **131**, 038003 (2019).

38. 38.

Bruch, R. et al. ZTF transient discovery report for 2019-06-10. *Transient Name Server Discovery Report* 2019-973 (2019); <https://www.wis-tns.org/ads/TNSTR-2019-973>.

39. 39.

Strotjohann, N. L. et al. Bright, months-long stellar outbursts announce the explosion of interaction-powered supernovae. *Astrophys. J.* **907**, 99 (2021).

40. 40.

Cenko, S. B. et al. The Automated Palomar 60 Inch Telescope. *Publ. Astron. Soc. Pac.* **118**, 1396–1406 (2006).

41. 41.

Blagorodnova, N. et al. The SED Machine: a robotic spectrograph for fast transient classification. *Publ. Astron. Soc. Pac.* **130**, 035003 (2018).

42. 42.

Fremling, C. et al. PTF12os and iPTF13bvn. Two stripped-envelope supernovae from low-mass progenitors in NGC 5806. *Astron. Astrophys.* **593**, A68 (2016).

43. 43.

Schulze, S. et al. Cosmic evolution and metal aversion in superluminous supernova host galaxies. *Mon. Not. R. Astron. Soc.* **473**,

1258–1285 (2018).

44. 44.

Aihara, H. et al. The eighth data release of the Sloan Digital Sky Survey: first data from SDSS-III. *Astrophys. J. Suppl. Ser.* **193**, 29 (2011).

45. 45.

Roming, P. W. A. et al. The Swift Ultra-Violet/Optical Telescope. *Space Sci. Rev.* **120**, 95–142 (2005).

46. 46.

Gehrels, N. et al. The Swift gamma-ray burst mission. *Astrophys. J.* **611**, 1005 (2004).

47. 47.

Schlafly, E. F. & Finkbeiner, D. P. Measuring reddening with Sloan Digital Sky Survey stellar spectra and recalibrating SFD. *Astrophys. J.* **737**, 103 (2011).

48. 48.

Komatsu, E. et al. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation. *Astrophys. J. Suppl. Ser.* **192**, 18 (2011).

49. 49.

Pastorello, A. et al. A giant outburst two years before the core-collapse of a massive star. *Nature* **447**, 829–832 (2007).

50. 50.

Ofek, E. O. et al. An outburst from a massive star 40 days before a supernova explosion. *Nature* **494**, 65–67 (2013).

51. 51.

Ofek, E. O. et al. Precursors prior to type IIn supernova explosions are common: precursor rates, properties, and correlations. *Astrophys. J.* **789**, 104 (2014).

52. 52.

Yaron, O. & Gal-Yam, A. WISeREP – an interactive supernova data repository. *Publ. Astron. Soc. Pac.* **124**, 668–681 (2012).

53. 53.

Ben-Ami, S. et al. The SED Machine: a dedicated transient IFU spectrograph. In *Proc. SPIE 8446: Ground-based and Airborne Instrumentation for Astronomy IV* (eds McLean, I. S. et al.) 844686 (SPIE, 2012).

54. 54.

Rigault, M. et al. Fully automated integral field spectrograph pipeline for the SEDMachine: pysedm. *Astron. Astrophys.* **627**, A115 (2019).

55. 55.

Hook, I. M. et al. The Gemini–North Multi-Object Spectrograph: performance in imaging, long-slit, and multi-object spectroscopic modes. *Publ. Astron. Soc. Pac.* **116**, 425–440 (2004).

56. 56.

Piasek, A. S. et al. SPRAT: Spectrograph for the Rapid Acquisition of Transients. *Proc. SPIE 9147: Ground-based and Airborne Instrumentation for Astronomy V* (eds Ramsay, S. K.) 91478H (2014).

57. 57.

Steele, I. A. et al. The Liverpool Telescope: performance and first results. *Proc. SPIE 5489: Ground-based Telescopes* (ed. Oschmann Jr,

J. M.) 679–692 (2004).

58. 58.

Chonis, T. S. et al. LRS2: design, assembly, testing, and commissioning of the second-generation low-resolution spectrograph for the Hobby–Eberly Telescope. *Proc. SPIE 9908: Ground-based and Airborne Instrumentation for Astronomy VI* (eds Evans, C. J. et al.) 99084C (2016).

59. 59.

Ramsey, L. W. et al. Early performance and present status of the Hobby–Eberly Telescope. *Proc. SPIE 3352: Advanced Technology Optical/IR Telescopes VI* (ed. Stepp, L. M.) 34–51 (1998).

60. 60.

Benn, C., Dee, K. & Agócs, T. ACAM: a new imager/spectrograph for the William Herschel Telescope. *Proc. SPIE 7104: Ground-based and Airborne Instrumentation for Astronomy II* (eds McLean, I. S. & Casali, M. M.) 70146X (2008).

61. 61.

Levine, S. E. et al. Status and performance of the Discovery Channel Telescope during commissioning. *Proc. SPIE 8444: Ground-based and Airborne Telescopes IV* (eds Stepp, L. M. et al.) 844419 (2012).

62. 62.

Levine, S. E., & DeGroff, W. T. Status and imaging performance of Lowell Observatory’s Discovery Channel Telescope in its first year of full science operations. *Proc. SPIE 9906: Ground-based and Airborne Telescopes VI* (eds Hall, H. J. et al.) 990621 (2016).

63. 63.

Oke, J. B. & Gunn, J. E. An efficient low- and moderate-resolution spectrograph for the Hale Telescope. *Publ. Astron. Soc. Pac.* **94**, 586–594 (1982).

64. 64.

Bellm, E. C. et al. pyraf-dbsp: reduction pipeline for the Palomar Double Beam Spectrograph. *Astrophysics Source Code Library* <https://www.ascl.net/1602.002> (2016).

65. 65.

Oke, J. B. et al. The Keck Low-Resolution Imaging Spectrometer. *Publ. Astron. Soc. Pac.* **107**, 375–385 (1995).

66. 66.

Perley, D. A. Fully automated reduction of longslit spectroscopy with the Low Resolution Imaging Spectrometer at the Keck Observatory. *Publ. Astron. Soc. Pac.* **131**, 084503 (2019).

67. 67.

Sander, A., Hamann, W.-R. & Todt, H. The Galactic WC stars: stellar parameters from spectral analyses indicate a new evolutionary sequence. *Astron. Astrophys.* **540**, A144 (2012).

68. 68.

Gal-Yam, A. et al. A Wolf-Rayet-like progenitor of SN 2013cu from spectral observations of a stellar wind. *Nature* **509**, 471–474 (2014).

69. 69.

Khazov, D. et al. Flash spectroscopy: emission lines from the ionized circumstellar material around <10-day-old type II supernovae. *Astrophys. J.* **818**, 3 (2016).

70. 70.

Bruch, R. J. et al. A large fraction of hydrogen-rich supernova progenitors experience elevated mass loss shortly prior to explosion. *Astrophys. J.* **912**, 46 (2021).

71. 71.

Yaron, O. et al. Confined dense circumstellar material surrounding a regular type II supernova. *Nat. Phys.* **13**, 510–517 (2017).

72. 72.

Gal-Yam, A. The most luminous supernovae. *Annu. Rev. Astron. Astrophys.* **57**, 305–333 (2019).

73. 73.

Valenti, S. et al. The carbon-rich type Ic SN 2007gr: the photospheric phase. *Astrophys. J.* **673**, L155–L158 (2008).

74. 74.

Branch, D. et al. Comparative direct analysis of type Ia supernova spectra. I. SN 1994D. *Publ. Astron. Soc. Pac.* **117**, 545–552 (2005).

75. 75.

Karamehmetoglu, E. et al. The luminous and rapidly evolving SN 2018bcc: clues toward the origin of type Ibn SNe from the Zwicky Transient Facility. *Astron. Astrophys.* **649**, A163 (2021).

76. 76.

Burrows, D. N. et al. The Swift X-Ray Telescope. *Space Sci. Rev.* **120**, 165–195 (2005).

77. 77.

Evans, P. A. et al. An online repository of Swift/XRT light curves of γ -ray bursts. *Astron. Astrophys.* **469**, 379–385 (2007).

78. 78.

Evans, P. A. et al. Methods and results of an automatic analysis of a complete sample of Swift–XRT observations of GRBs. *Mon. Not. R. Astron. Soc.* **397**, 1177–1201 (2009).

79. 79.

HI4PI Collaboration. HI4PI: A full-sky H i survey based on EBHIS and GASS. *Astron. Astrophys.* **594**, A116 (2016).

80. 80.

Ho, A. Y. Q. et al. The Koala: a fast blue optical transient with luminous radio emission from a starburst dwarf galaxy at $z = 0.27$. *Astrophys. J.* **895**, 49 (2020).

81. 81.

Coppejans, D. L. et al. A mildly relativistic outflow from the energetic, fast-rising blue optical transient CSS161010 in a dwarf galaxy. *Astrophys. J.* **895**, L23 (2020).

82. 82.

De, K. et al. A hot and fast ultra-stripped supernova that likely formed a compact neutron star binary. *Science* **362**, 201–206 (2018).

83. 83.

Yao, Y. et al. SN2019dge: a helium-rich ultra-stripped envelope supernova. *Astrophys. J.* **900**, 46 (2020).

84. 84.

Ho, A. Y. Q. et al. Evidence for late-stage eruptive mass loss in the progenitor to SN2018gep, a broad-lined Ic supernova: pre-explosion emission and a rapidly rising luminous transient. *Astrophys. J.* **887**, 169 (2019).

85. 85.

Perley, D. A. et al. Real-time discovery of AT2020xnd: a fast, luminous ultraviolet transient with minimal radioactive ejecta. *Mon. Not. R. Astron. Soc.* **887**, 5138–5147(2021).

86. 86.

Ofek, E. O. et al. SN 2010jl: optical to hard X-ray observations reveal an explosion embedded in a ten solar mass cocoon. *Astrophys. J.* **781**, 42 (2014).

87. 87.

Chandra, P., Chevalier, R. A., Chugai, N., Fransson, C. & Soderberg, A. M. X-Ray and radio emission from type IIn supernova SN 2010jl. *Astrophys. J.* **810**, 32 (2015).

88. 88.

Immler, S. et al. Swift and Chandra detections of supernova 2006jc: evidence for interaction of the supernova shock with a circumstellar shell. *Astrophys. J.* **674**, L85–L88 (2008).

89. 89.

Martin, C. & GALEX Team. The Galaxy Evolution Explorer – early data. In *Symposium – International Astronomical Union Vol. 216: Maps of the Cosmos* (eds Colless, M. et al.) 221–229 (2005).

90. 90.

Ahn, C. P. et al. The ninth data release of the Sloan Digital Sky Survey: first spectroscopic data from the SDSS-III Baryon Oscillation Spectroscopic Survey. *Astrophys. J. Suppl. Ser.* **203**, 21 (2012).

91. 91.

Wright, E. L. et al. The Wide-field Infrared Survey Explorer (WISE): mission description and initial on-orbit performance. *Astron. J.* **140**, 1868–1881 (2010).

92. 92.

Lang, D. unWISE: unblurred coadds of the WISE imaging. *Astron. J.* **147**, 108 (2014).

93. 93.

Mainzer, A. et al. Initial performance of the NEOWISE reactivation mission. *Astrophys. J.* **792**, 30 (2014).

94. 94.

Meisner, A. M. et al. Deep full-sky coadds from three years of WISE and NEOWISE observations. *Astron. J.* **154**, 161 (2017).

95. 95.

Breeveld, A. A. et al. An updated ultraviolet calibration for the Swift/UVOT. In *AIP Conf. Proc. Vol 1358: Gamma Ray Bursts 2010* (eds McEnery, J. E. et al.) 373–376 (2011).

96. 96.

Wright, A. H. et al. Galaxy and Mass Assembly: accurate panchromatic photometry from optical priors using LAMBDA. *Mon. Not. R. Astron. Soc.* **460**, 765–801 (2016).

97. 97.

Schulze, S. et al. The Palomar Transient Factory core-collapse supernova host-galaxy sample. I. Host-galaxy distribution functions and environment dependence of core-collapse supernovae. *Astrophys. J. Supp. Ser.* **255**, 29 (2021).

98. 98.

Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G. & Byler, N. Deriving physical properties from broadband photometry with Prospector: description of the model and a demonstration of its accuracy using 129 galaxies in the local Universe. *Astrophys. J.* **837**, 170 (2017).

99. 99.

Conroy, C., Gunn, J. E. & White, M. The propagation of uncertainties in stellar population synthesis modeling. I. The relevance of uncertain aspects of stellar evolution and the initial mass function to the derived physical properties of galaxies. *Astrophys. J.* **699**, 486–506 (2009).

100. 100.

Foreman-Mackey, D. et al. python-fsps: Python bindings to fsps (V0.1.1). *Zenodo* <https://zenodo.org/record/12157#.YbjtiUaxXdc> (2014).

101. 101.

Byler, N. et al. Nebular continuum and line emission in stellar population synthesis models. *Astrophys. J.* **840**, 44 (2017).

102. 102.

Chabrier, G. Galactic stellar and substellar initial mass function. *Publ. Astron. Soc. Pac.* **115**, 763–795 (2003).

103. 103.

Calzetti, D. et al. The dust content and opacity of actively star-forming galaxies. *Astrophys. J.* **533**, 682–695 (2000).

104. 104.

Speagle, J. S. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. R. Astron. Soc.* **493**, 3132–3158 (2020).

105. 105.

Marino, R. A. et al. The O3N2 and N2 abundance indicators revisited: improved calibrations based on CALIFA and T_e -based literature data. *Astron. Astrophys.* **559**, A114 (2013).

106. 106.

Asplund, M., Grevesse, N., Sauval, A. J. & Scott, P. The chemical composition of the Sun. *Annu. Rev. Astron. Astrophys.* **47**, 481–522 (2009).

107. 107.

Lunnan, R. et al. Hydrogen-poor superluminous supernovae and long-duration gamma-ray bursts have similar host galaxies. *Astrophys. J.* **787**, 138–156 (2014).

108. 108.

Leloudas, G. et al. Spectroscopy of superluminous supernova host galaxies. A preference of hydrogen-poor events for extreme emission line galaxies. *Mon. Not. R. Astron. Soc.* **449**, 917–932 (2015).

109. 109.

Perley, D. A. et al. Host-galaxy properties of 32 low-redshift superluminous supernovae from the Palomar Transient Factory. *Astrophys. J.* **830**, 13 (2016).

110. 110.

Kiewe, M. et al. Caltech Core-Collapse Project (CCCP) observations of type IIn supernovae: typical properties and implications for their progenitor stars. *Astrophys. J.* **744**, 10 (2012).

111. 111.

Rubin, A. et al. Type II supernova energetics and comparison of light curves to shock-cooling models. *Astrophys. J.* **820**, 33 (2016).

112. 112.

Yao, Y. et al. ZTF early observations of type Ia supernovae. I. properties of the 2018 sample. *Astrophys. J.* **886**, 152 (2019).

113. 113.

Miller, A. A. et al. ZTF early observations of type Ia supernovae. II. First light, the initial rise, and time to reach maximum brightness. *Astrophys. J.* **902**, 47 (2020).

114. 114.

Barbarino, C. et al. Type Ic supernovae from the (intermediate) Palomar Transient Factory. Preprint at <https://arxiv.org/abs/2010.08392> (2020).

115. 115.

Taddia, F. et al. Analysis of broad-lined Type Ic supernovae from the (intermediate) Palomar Transient Factory. *Astron. Astrophys.* **621**, A71 (2019).

116. 116.

Gal-Yam, A., Ofek, E. O. & Shemmer, O. Supernova 2002ap: the first month. *Mon. Not. R. Astron. Soc.* **332**, L73–L77 (2002).

117. 117.

Mazzali, P. A. et al. The type Ic hypernova SN 2002ap. *Astrophys. J.* **572**, L61–L65 (2002).

118. 118.

Galama, T. J. et al. An unusual supernova in the error box of the γ -ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).

119. 119.

Campana, S. et al. The association of GRB 060218 with a supernova and the evolution of the shock wave. *Nature* **442**, 1008–1010 (2006).

120. 120.

Bianco, F. B. et al. Multi-color optical and near-infrared light curves of 64 stripped-envelope core-collapse supernovae. *Astrophys. J. Suppl. Ser.* **213**, 19 (2014).

121. 121.

Richmond, M. W. et al. *UBVRI* photometry of the type Ic SN 1994I in M51. *Astron. J* **111**, 327–339 (1996).

122. 122.

Sauer, D. N. et al. The properties of the ‘standard’ type Ic supernova 1994I from spectral models. *Mon. Not. R. Astron. Soc.* **369**, 1939–1948 (2006).

123. 123.

Ben-Ami, S. et al. Discovery and early multi-wavelength measurements of the energetic type Ic supernova PTF12gzk: a massive-star explosion in a dwarf host galaxy. *Astrophys. J.* **760**, L33 (2012).

124. 124.

Cao, Y. et al. Discovery, progenitor and early evolution of a stripped envelope supernova iPTF13bvn. *Astrophys. J.* **775**, L7 (2013).

125. 125.

Stritzinger, M. et al. Optical photometry of the type Ia supernova 1999ee and the type Ib/c supernova 1999ex in IC 5179. *Astron. J* **124**, 2100–2117 (2002).

126. 126.

Mazzali, P. A. et al. The metamorphosis of supernova SN 2008D/XRF 080109: a link between supernovae and GRBs/hypernovae. *Science* **321**, 1185–1188 (2008).

127. 127.

Soderberg, A. M. et al. An extremely luminous X-ray outburst at the birth of a supernova. *Nature* **453**, 469–474 (2008).

128. 128.

Valenti, S. et al. SN 2009jf: a slow-evolving stripped-envelope core-collapse supernova. *Mon. Not. R. Astron. Soc.* **416**, 3138–3159 (2011).

129. 129.

Rest, A. et al. A fast-evolving luminous transient discovered by K2/Kepler. *Nat. Astron.* **2**, 307–311 (2018).

130. 130.

Whitesides, L. et al. iPTF 16asu: a luminous, rapidly evolving, and high-velocity supernova. *Astrophys. J.* **851**, 107 (2017).

131. 131.

Prentice, S. J. et al. The Cow: discovery of a luminous, hot, and rapidly evolving transient. *Astrophys. J.* **865**, L3 (2018).

132. 132.

Perley, D. A. et al. The fast, luminous ultraviolet transient AT2018cow: extreme supernova, or disruption of a star by an

intermediate-mass black hole? *Mon. Not. R. Astron. Soc.* **484**, 1031–1049 (2019).

133. 133.

Margutti, R. et al. An embedded X-ray source shines through the aspherical AT 2018cow: revealing the inner workings of the most luminous fast-evolving optical transients. *Astrophys. J.* **872**, 18 (2019).

134. 134.

Mazzali, P. A. et al. The Type Ic SN 2007gr: a census of the ejecta from late-time optical-infrared spectra. *Mon. Not. R. Astron. Soc.* **408**, 87–96 (2010).

135. 135.

Soumagnac, M. T. et al. Supernova PTF 12glz: a possible shock breakout driven through an aspherical wind. *Astrophys. J.* **872**, 141 (2019).

136. 136.

Smith, N. et al. Episodic mass loss in binary evolution to the Wolf-Rayet phase: Keck and HST proper motions of RY Scuti’s nebula. *Mon. Not. R. Astron. Soc.* **418**, 1959–1972 (2011).

137. 137.

Waxman, E. & Katz, B. Shock breakout theory. In *Handbook of Supernovae* (eds Alsabti, A. W. & Murdin, P.) 967–1015 (2017); https://doi.org/10.1007/978-3-319-21846-5_33.

138. 138.

Ganot, N. et al. The detection rate of early UV emission from supernovae: a dedicated GALEX/PTF survey and calibrated theoretical estimates. *Astrophys. J.* **820**, 57 (2016).

139. 139.

Ofek, E. O. et al. Supernova PTF 09UJ: a possible shock breakout from a dense circumstellar wind. *Astrophys. J.* **724**, 1396–1401 (2010).

140. 140.

Smith, N. et al. PTF11iqb: cool supergiant mass-loss that bridges the gap between Type IIn and normal supernovae. *Mon. Not. R. Astron. Soc.* **449**, 1876–1896 (2015).

141. 141.

Ben-Ami, S. et al. SN 2010mb: direct evidence for a supernova interacting with a large amount of hydrogen-free circumstellar material. *Astrophys. J.* **785**, 37 (2014).

142. 142.

Soumagnac, M. T. et al. SN 2018fif: the explosion of a large red supergiant discovered in its infancy by the Zwicky Transient Facility. *Astrophys. J.* **902**, 6 (2020).

Acknowledgements

This work is based on observations obtained with the Samuel Oschin 48-inch Telescope and the 60-inch Telescope at Palomar Observatory as part of the Zwicky Transient Facility project. ZTF is supported by the US National Science Foundation (NSF) under grant AST-1440341 and a collaboration including Caltech, IPAC, the Weizmann Institute for Science, the Oskar Klein Center at Stockholm University, the University of Maryland, the University of Washington, Deutsches Elektronen-Synchrotron and Humboldt University, Los Alamos National Laboratories, the TANGO Consortium of Taiwan, the University of Wisconsin at Milwaukee, and Lawrence Berkeley National Laboratories. Operations are conducted by COO, IPAC, and UW. This work includes observations made with the Nordic Optical Telescope (NOT), owned in collaboration by the University

of Turku and Aarhus University, and operated jointly by Aarhus University, the University of Turku and the University of Oslo (representing Denmark, Finland and Norway, respectively), the University of Iceland and Stockholm University, at the Observatorio del Roque de los Muchachos, La Palma, Spain, of the Instituto de Astrofísica de Canarias. These data were obtained with ALFOSC, which is provided by the Instituto de Astrofísica de Andalucía (IAA) under a joint agreement with the University of Copenhagen and NOT. This work includes observations made with the GTC telescope, in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias, under Director's Discretionary Time. Some of the data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California, and the National Aeronautics and Space Administration (NASA); the Observatory was made possible by the generous financial support of the W. M. Keck Foundation. This work includes observations obtained at the international Gemini Observatory, a programme of NSF's NOIRLab, which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the NSF on behalf of the Gemini Observatory partnership: the NSF (USA), National Research Council (Canada), Agencia Nacional de Investigación y Desarrollo (Chile), Ministerio de Ciencia, Tecnología e Innovación (Argentina), Ministério da Ciência, Tecnologia, Inovações e Comunicações (Brazil), and Korea Astronomy and Space Science Institute (Republic of Korea). We wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the Indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain. This work includes observations obtained at the Liverpool Telescope, which 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 Astrofísica de Canarias with financial support from the UK Science and Technology Facilities Council. Research at Lick Observatory is partially supported by a generous gift from Google. This work includes observations obtained with the Hobby-Eberly Telescope, which is a joint project of the University of Texas at Austin, the Pennsylvania State University, Ludwig-Maximilians-Universität München, and Georg-August-Universität Göttingen. These results made use of the

Lowell Discovery Telescope (LDT) at Lowell Observatory. Lowell is a private, non-profit institution dedicated to astrophysical research and public appreciation of astronomy and operates the LDT in partnership with Boston University, the University of Maryland, the University of Toledo, Northern Arizona University and Yale University. This work benefited from the OPTICON telescope access programme (<https://www.astro-opticon.org/index.html>), funded from the European Union’s Horizon 2020 research and innovation programme under grant agreement 730890. We made use of IRAF, which is distributed by the NSF NOIRLab. A.G.-Y. is supported by the EU via ERC grant no. 725161, the ISF GW excellence centre, an IMOS space infrastructure grant, BSF/Transformative and GIF grants, as well as by the Benoziyo Endowment Fund for the Advancement of Science, the Deloro Institute for Advanced Research in Space and Optics, The Veronika A. Rabl Physics Discretionary Fund, Minerva, Yeda-Sela and the Schwartz/Reisman Collaborative Science Program; A.G.-Y. is the incumbent of the The Arlyn Imberman Professorial Chair. M.M.K. acknowledges generous support from the David and Lucile Packard Foundation; the GROWTH project was funded by the NSF under grant AST-1545949. E.C.K., J.S. and S.S. acknowledge support from the G.R.E.A.T. research environment funded by Vetenskapsrådet, the Swedish Research Council, under project number 2016-06012; E.C.K. also received support from The Wenner-Gren Foundations. The Oskar Klein Center’s participation in ZTF was made available by the K.A.W. Foundation. G.L. is supported by a research grant (19054) from VILLUM FONDEN. J.C.W. and B.P.T. are supported by NSF grant AST-1813825. A.V.F.’s supernova group at UC Berkeley is supported by the TABASGo Foundation, the Christopher R. Redlich Fund, and the Miller Institute for Basic Research in Science (A.V.F. is a Senior Miller Fellow).

Author information

Affiliations

1. Department of Particle Physics and Astrophysics, Weizmann Institute of Science, Rehovot, Israel

A. Gal-Yam, R. Bruch, S. Schulze, Y. Yang, I. Irani, M. T. Soumagnac, O. Yaron, N. L. Strotjohann, E. Zimmerman & E. O. Ofek

2. The Oskar Klein Centre, Department of Astronomy and Department of Physics, AlbaNova, Stockholm University, Stockholm, Sweden

S. Schulze, J. Sollerman, E. C. Kool, C. Barbarino & C. Fransson

3. Department of Astronomy, University of California, Berkeley, Berkeley, CA, USA

Y. Yang, A. V. Filippenko, W. Zheng & T. G. Brink

4. Astrophysics Research Institute, Liverpool John Moores University, Liverpool, UK

D. A. Perley & C. M. Copperwheat

5. Lawrence Berkeley National Laboratory, Berkeley, CA, USA

M. T. Soumagnac

6. Division of Physics, Mathematics, and Astronomy, California Institute of Technology, Pasadena, CA, USA

S. R. Kulkarni, M. M. Kasliwal, K. De, Y. Yao, C. Fremling, L. Yan, M. J. Graham, D. A. Duev, A. J. Drake, R. Dekany, I. Andreoni, Y. Sharma, R. Riddle & J. van Roestel

7. Miller Institute for Basic Research in Science, University of California, Berkeley, Berkeley, CA, USA

A. V. Filippenko

8. Department of Astronomy and Astrophysics, University of California, Santa Cruz, Santa Cruz, CA, USA

R. J. Foley, J. Brown, M. Siebert & T. Hung

9. DTU Space, National Space Institute, Technical University of Denmark, Kongens Lyngby, Denmark

G. Leloudas

10. Grantecan, Centro de Astrofísica de La Palma, La Palma, Spain

A. L. Cabrera-Lavers, D. García-Alvarez & A. Marante-Barreto

11. Department of Astronomy, University of Maryland, College Park, MD, USA

S. Frederick

12. Department of Astronomy, University of Texas at Austin, Austin, TX, USA

J. C. Wheeler, J. Vinkó & B. P. Thomas

13. Konkoly Observatory, ELKH CSFK, Budapest, Hungary

J. Vinkó

14. Department of Optics and Quantum Electronics, University of Szeged, Szeged, Hungary

J. Vinkó

15. ELTE Institute of Physics, Eötvös Loránd University, Budapest, Hungary

J. Vinkó

16. DIRAC Institute, Department of Astronomy, University of Washington, Seattle, WA, USA

E. C. Bellm

17. IPAC, California Institute of Technology, Pasadena, CA, USA

B. Rusholme & D. L. Shupe

18. Department of Astronomy, Faculty of Mathematics, University of Belgrade, Belgrade, Serbia

N. Knezevic

Contributions

A.G.-Y. initiated the project, planned the observations, conducted spectroscopic and physical analysis, and wrote the manuscript. R.B. identified the transient, initiated follow-up observations, conducted photometric analysis, and contributed to the WIS infant SN programme. S.S. contributed to follow-up design and execution, conducted multiwavelength and host-galaxy analysis, and contributed to the manuscript. Y. Yang. contributed to follow-up design and execution, reduced the Gemini spectra, and contributed to the manuscript. D.A.P. conducted follow-up observations with the LT and contributed to the manuscript. I.I. conducted photometric and spectroscopic analysis and contributed to physical interpretation and manuscript writing. J.S. helped to plan and develop the manuscript and provided NOT data. E.C.K. analysed the bolometric light curve and provided physical modelling. M.T.S. contributed to photometric analysis. O.Y. conducted spectroscopic modelling. N.L.S. conducted a prediscovery variability search and contributed to the manuscript. E.Z. contributed to spectroscopic analysis. C.B. provided the sample of SNe Ic from PTF and reduced the NOT spectroscopy. S.R.K. is the ZTF PI. S.R.K., M.M.K., C. Fremling. and L.Y. provided Palomar and Keck data. K.D. and Y. Yao. reduced the Palomar and Keck data. E.O.O. and C. Fransson. contributed to physical interpretation. A.V.F., W.Z., T.G.B., R.J.F., J.B. and M.S. contributed Lick and Keck data; A.V.F. also contributed to the manuscript. C.M.C. contributed LT data. A.L.C.-L., D.G.-A. and A.M.-B. provided GTC observations. S.F. and T.H. provided LDT observations. J.C.W., B.P.T. and J.V. planned, obtained and reduced the HET observations. G.L. reduced the GTC spectra and contributed to the manuscript. M.J.G., D.A.D., A.J.D., R.D., E.C.B., B.R., D.L.S., I.A., Y.S., R.R. and J.v.R. are ZTF builders. N.K. contributed to the

spectroscopic analysis. Many authors provided comments on the manuscript, and all authors have approved it.

Corresponding author

Correspondence to [A. Gal-Yam](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

[Extended Data Fig. 1 Early spectroscopy shows strong emission lines of carbon, oxygen and neon.](#)

During the initial hot phase (1–5.6 d after explosion) these highly ionized lines (see Fig. 1 for detailed line identification) weaken with time, with pure emission lines evolving to P Cygni profiles, and then to absorption-dominated profiles. Major emission features are marked; the spectral area around 5,250 Å in restframe is impacted by imperfect subtraction of the strong atmospheric 5,577 Å skyline (grey shade). Five additional P60 and

LT spectra with lower signal-to-noise and spectral resolution obtained during this period are omitted for clarity.

Extended Data Fig. 2 Evolution toward lower ionization.

The spectroscopic series obtained during the intermediate phase (5–19 d after explosion) follows the weakening and disappearance of the CIII and OIII absorption features seen earlier, and the emergence of a set of low-ionization emission lines, initially of CII (red) and later OI (blue; 50% intensity lines extracted as in ref. ²⁹). Higher resolution spectra resolve the broad features in the blue into multiple narrow components better described by CIII and OIII at zero velocity than by OII blends sometimes seen in hot early phases of stripped SNe, including Type I SLSNe^{23,72}. By day 19 (bottom) broad features appear and the spectrum shows a marked blue excess. Seven additional spectra omitted for clarity.

Extended Data Fig. 3 Development of the photospheric spectrum.

The spectroscopic series obtained during the late phase (19–27 d after explosion) evolves as features of heavier elements (for example, Mg) begin to emerge, while broad absorption features develop. Initially, strong features (such as OI 7774Å and CII 6580Å) present both a narrow ($\sim 2,000 \text{ km s}^{-1}$ blue edge) absorption feature as well as a broader ($\sim 6,000 \text{ km s}^{-1}$ minimum) component. At 27 d after explosion, relatively broad absorption features have developed that are reminiscent of spectra of type Ic SNe, with features from Mg, Ca and Fe appearing in addition to C and O. Excess continuum in the blue is evident, probably arising from the Fe ii pseudo-continuum often seen in spectra of interacting SNe (types IIn and Ibn²³).

Extended Data Fig. 4 Blackbody SED fits calculated using PhotoFit142.

Our well sampled photometry extending from the Swift UV bands to the near-infrared (NIR) z' band is well fit by a blackbody curve during the first 12 days after explosion. From day 15 onwards, a clear blue excess develops

initially in the UV and extending into the blue part of the optical band from day 21 onward. The derived blackbody parameters (radius and temperature) are therefore less reliable from that date. Standard 1σ error bars marked.

Extended Data Fig. 5 Light curves of SN 2019hgp extending from the UV to the NIR.

Post-peak Swift B - and V -band photometry is inconsistent with data from other sources and probably unreliable. Five outlying P60 points ($1 u$, $2 g$ and $2 r$) are inconsistent with the rest of the data to well above their formal errors and have been removed. Standard 1σ error bars marked.

Extended Data Fig. 6 Low-order polynomial fits to the early r -band photometry indicate that the explosion occurred on 2019 June 7.1 ± 0.2 d.

Although a linear fit does not provide a good description of the data, low-order (degrees 2–4) polynomials fit the data well and converge on an estimated explosion time occurring ~ 1 d prior to discovery (stars denote extrapolated times of zero flux). Stacked pre-discovery data recover a detection during the prior night. All times in the paper are reported relative to this fiducial explosion time. The last 5σ non-detection is also marked. Standard 1σ error bars marked.

Extended Data Fig. 7 A comparison of spectra of interacting SNe.

Our spectrum of SN 2019hgp is overall quite similar to those of SNe Ibn (SN 2016jc²⁰ and SN 2018bcc⁷⁵), sharing in particular the unusual non-thermal continuum that is flat on the red side, and has a pronounced elevation bluewards of $\sim 5,500$ Å (dotted line); this emission probably arises from a quasi-continuum of multiple Fe ii emission lines (resolved in some cases, for example, the Type IIn SN 2005cl¹¹⁰, bottom). The hallmark strong He I emission lines common to SNe Ibn ($\lambda\lambda 5876, 6678, 7065, 7281$) are absent from the spectrum of SN 2019hgp. Remarkably, this object does, however, show broad absorption features that are missing from spectra of

Type Ibn and Type IIn, suggesting that strong shocks are not obscuring our line of sight at 27.4 d after explosion.

Extended Data Fig. 8 SN 2019hgp (marked by the crosshair) exploded in the outskirts of its host galaxy at a projected distance of 4.4 kpc (3.54").

The host shows elongated arms of diffuse emission which could suggest a spiral arm or a recent episode of galaxy interaction. In this image east is to the left and north up. The image size is 40" on the side.

Extended Data Fig. 9 Extinction fits to our first-epoch SED (+1.5 d) using various extinction laws.

a–c, Extinction fits using MW (**a**), LMC (**b**) and SMC (**c**) extinction laws. A fit with negligible host extinction (red) fits the data well. Values of extinction, extending up to $E_{B-V} = 0.15$ mag (requiring blackbody temperatures of ~ 100 kK) are allowed; higher extinction is ruled out regardless of extinction law parameters (MW (**d**) law shown, SMC and LMC are similar). χ^2 minimization is done using epochs well fit by blackbody curves (<15 d). Standard 1σ error bars marked.

Extended Data Fig. 10 Modelling of the emission complex around 4,660Å during the first two Gemini epochs.

1 day (**a, b**) and 3 days (**c, d**) after explosion. We fit a combination of a Lorentzian emission component of CIII $\lambda 4650\text{\AA}$ along with a blueshifted Gaussian absorption component. Including an additional Lorentzian emission from He ii $\lambda 4686\text{\AA}$ (**b, d**) is preferred by the data (in the χ^2 sense) even though this feature does not appear as a distinct emission peak. We conclude that the presence of He ii in these spectra cannot be ruled out.

Extended Data Fig. 11 A comparison of our +27.4 d Keck spectrum of SN 2019hgp to SYNOW models.

The spectrum can be well represented by a combination of common elements seen in supernovae (oxygen, sodium, magnesium, calcium and iron); the addition of neon, which is unique to this object, seems to improve the fit substantially around 6,200–7,000 Å (yellow). We compare models without (green) and with (red) He I; we find that the contribution of helium compromises the fit around 6,000–7,000 Å, owing to the expected but unobserved contribution of the P Cygni profile of He I λ 6678Å. Perhaps this could be reconciled by more sophisticated modelling, though we note that recent analysis⁷⁵ suggests that the emission component from this particular transition grows stronger with time in spectra of He-rich SNe Ibn.

Supplementary information

Supplementary Information

This file contains Supplementary Figs 1–5 and Supplementary Tables 1–3 (a full machine-readable file for Supplementary Table 1 is also provided separately).

Supplementary Table 1

Full photometry of SN 2019hgp.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Gal-Yam, A., Bruch, R., Schulze, S. *et al.* A WC/WO star exploding within an expanding carbon–oxygen–neon nebula. *Nature* **601**, 201–204 (2022). <https://doi.org/10.1038/s41586-021-04155-1>

- Received: 06 April 2021
- Accepted: 15 October 2021
- Published: 12 January 2022
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04155-1>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04155-1>

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 12 January 2022](#)

Signatures of a strange metal in a bosonic system

- [Chao Yang](#)¹ ✉nal,
- [Haiwen Liu](#)² ✉nal,
- [Yi Liu](#)³,
- [Jiandong Wang](#)¹,
- [Dong Qiu](#)¹,
- [Sishuang Wang](#)¹,
- [Yang Wang](#)¹,
- [Qianmei He](#)¹,
- [Xiuli Li](#)¹,
- [Peng Li](#)¹,
- [Yue Tang](#)³,
- [Jian Wang](#) ORCID: orcid.org/0000-0002-7212-0904^{3,4},
- [X. C. Xie](#)³,
- [James M. Valles Jr](#) ORCID: orcid.org/0000-0002-9786-0730⁵,
- [Jie Xiong](#) ORCID: orcid.org/0000-0003-3881-6948¹ &
- [Yanrong Li](#)^{1,6}

Nature volume **601**, pages 205–210 (2022)

- 3421 Accesses
- 107 Altmetric
- [Metrics details](#)

Subjects

- [Phase transitions and critical phenomena](#)
- [Superconducting properties and materials](#)

Abstract

Fermi liquid theory forms the basis for our understanding of the majority of metals: their resistivity arises from the scattering of well defined quasiparticles at a rate where, in the low-temperature limit, the inverse of the characteristic time scale is proportional to the square of the temperature. However, various quantum materials^{[1](#),[2](#),[3](#),[4](#),[5](#),[6](#),[7](#),[8](#),[9](#),[10](#),[11](#),[12](#),[13](#),[14](#),[15](#)}—notably high-temperature superconductors^{[1](#),[2](#),[3](#),[4](#),[5](#),[6](#),[7](#),[8](#),[9](#),[10](#)}—exhibit strange-metallic behaviour with a linear scattering rate in temperature, deviating from this central paradigm. Here we show the unexpected signatures of strange metallicity in a bosonic system for which the quasiparticle concept does not apply. Our nanopatterned $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (YBCO) film arrays reveal linear-in-temperature and linear-in-magnetic field resistance over extended temperature and magnetic field ranges. Notably, below the onset temperature at which Cooper pairs form, the low-field magnetoresistance oscillates with a period dictated by the superconducting flux quantum, $h/2e$ (e , electron charge; h , Planck's constant). Simultaneously, the Hall coefficient drops and vanishes within the measurement resolution with decreasing temperature, indicating that Cooper pairs instead of single electrons dominate the transport process. Moreover, the characteristic time scale τ in this bosonic system follows a scale-invariant relation without an intrinsic energy scale: $\hbar/\tau \approx a(k_{\text{B}}T + \gamma\mu_{\text{B}}B)$, where \hbar is the reduced Planck's constant, a is of order unity^{[7](#),[8](#),[11](#),[12](#)}, k_{B} is Boltzmann's constant, T is temperature, μ_{B} is the Bohr magneton and $\gamma \approx 2$. By extending the reach of strange-metal phenomenology to a bosonic system, our results suggest that there is a fundamental principle governing their transport that transcends particle statistics.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Fig. 1: Linear-in-temperature resistance near a bosonic anomalous metal–insulator transition in nanopatterned YBCO thin films.



Fig. 2: T -linear resistance and scale-invariant B -linear resistance in nanopatterned YBCO thin films under perpendicular magnetic field.

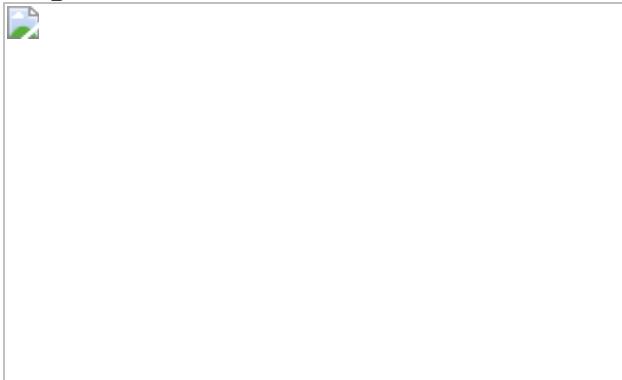


Fig. 3: $B-T$ scaling in nanopatterned YBCO thin films.



Fig. 4: Phase diagram of nanopatterned YBCO thin films.



Data availability

The data that support the plots within this paper are available from the Zenodo data repository, <https://doi.org/10.5281/zenodo.5603259>. [Source data](#) are provided with this paper.

References

1. 1.

Anderson, P. W. *The Theory of Superconductivity in the High- T_C Cuprates* (Princeton Univ. Press, 1998).

2. 2.

Keimer, B., Kivelson, S. A., Norman, M. R., Uchida, S. & Zaanen, J. From quantum matter to high-temperature superconductivity in copper oxides. *Nature* **518**, 179–186 (2015).

3. 3.

Zaanen, J. Why the temperature is high. *Nature* **430**, 512–513 (2004).

4. 4.

Jin, K., Butch, N. P., Kirshenbaum, K., Paglione, J. & Greene, R. L. Link between spin fluctuations and electron pairing in copper oxide superconductors. *Nature* **476**, 73–75 (2011).

5. 5.

Greene, R. L., Mandal, P. R., Poniatowski, N. R. & Sarkar, T. The strange metal state of the electron-doped cuprates. *Annu. Rev. Condens. Matter Phys.* **11**, 213–229 (2020).

6. 6.

Taillefer, L. Scattering and pairing in cuprate superconductors. *Annu. Rev. Condens. Matter Phys.* **1**, 51–70 (2010).

7. 7.

Giraldo-Gallo, P. et al. Scale-invariant magnetoresistance in a cuprate superconductor. *Science* **361**, 479–481 (2018).

8. 8.

Legros, A. et al. Universal T -linear resistivity and Planckian dissipation in overdoped cuprates. *Nat. Phys.* **15**, 142–147 (2019).

9. 9.

Varma, C. M., Littlewood, P. B., Schmitt-Rink, S., Abrahams, E. & Ruckenstein, A. E. Phenomenology of the normal state of Cu-O high-temperature superconductors. *Phys. Rev. Lett.* **63**, 1996–1999 (1989).

10. 10.

Varma, C. M. Linear in temperature resistivity and associated mysteries including high temperature superconductivity. *Rev. Mod. Phys.* **92**, 031001 (2020).

11. 11.

Patel, A. A. & Sachdev, S. Theory of a Planckian metal. *Phys. Rev. Lett.* **123**, 066601 (2019).

12. 12.

Bruin, J. A. N., Sakai, H., Perry, R. S. & Mackenzie, A. P. Similarity of scattering rates in metals showing T -linear resistivity. *Science* **339**, 804–807 (2013).

13. 13.

Hayes, I. M. et al. Scaling between magnetic field and temperature in the high-temperature superconductor $\text{BaFe}_2(\text{As}_{1-x}\text{P}_x)_2$. *Nat. Phys.* **12**, 916–919 (2016).

14. 14.

Doiron-Leyraud, N. et al. Correlation between linear resistivity and T_c in the Bechgaard salts and the pnictide superconductor $\text{Ba}(\text{Fe}_{1-x}\text{Co}_x)_2\text{As}_2$. *Phys. Rev. B* **80**, 214531 (2009).

15. 15.

Cao, Y. et al. Strange metal in magic-angle graphene with near Planckian dissipation. *Phys. Rev. Lett.* **124**, 076801 (2020).

16. 16.

Zaanen, J. Planckian dissipation, minimal viscosity and the transport in cuprate strange metals. *SciPost Phys.* **6**, 061 (2019).

17. 17.

Cao, C. et al. Universal quantum viscosity in a unitary Fermi gas. *Science* **331**, 58–61 (2011).

18. 18.

Kovtun, P. K., Son, D. T. & Starinets, A. O. Viscosity in strongly interacting quantum field theories from black hole physics. *Phys. Rev. Lett.* **94**, 111601 (2005).

19. 19.

Zaanen, J., Liu, Y., Sun, Y.-W. & Schalm, K. *Holographic Duality in Condensed Matter Physics* (Cambridge Univ. Press, 2015).

20. 20.

Hartnoll, S. A., Lucas, A. & Sachdev, S. Holographic quantum matter. Preprint at <https://arxiv.org/abs/1612.07324v1> (2016).

21. 21.

Kapitulnik, A., Kivelson, S. A. & Spivak, B. Anomalous metals: failed superconductors. *Rev. Mod. Phys.* **91**, 011002 (2019).

22. 22.

Phillips, P. & Dalidovich, D. The elusive Bose metal. *Science* **302**, 243–247 (2003).

23. 23.

Mason, N. & Kapitulnik, A. Dissipation effects on the superconductor–insulator transition in 2D superconductors. *Phys. Rev. Lett.* **82**, 5341–5344 (1999).

24. 24.

Saito, Y., Kasahara, Y., Ye, J., Iwasa, Y. & Nojima, T. Metallic ground state in an ion-gated two-dimensional superconductor. *Science* **350**, 409–413 (2015).

25. 25.

Breznay, N. P. & Kapitulnik, A. Particle–hole symmetry reveals failed superconductivity in the metallic phase of two-dimensional superconducting films. *Sci. Adv.* **3**, e1700612 (2017).

26. 26.

Yang, C. et al. Intermediate bosonic metallic state in the superconductor–insulator transition. *Science* **366**, 1505–1509 (2019).

27. 27.

Liu, Y. et al. Type-II Ising superconductivity and anomalous metallic state in macro-size ambient-stable ultrathin crystalline films. *Nano Lett.* **20**, 5728–5734 (2020).

28. 28.

Lindner, H. N. & Auerbach, A. Conductivity of hard core bosons: a paradigm of a bad metal. *Phys. Rev. B* **81**, 054512 (2010).

29. 29.

Stewart Jr, M. D., Yin, A., Xu, J. M. & Valles Jr, J. M. Superconducting pair correlations in an amorphous insulating nanohoneycomb film. *Science* **318**, 1273–1275 (2007).

30. 30.

Daou, R. et al. Linear temperature dependence of resistivity and change in the Fermi surface at the pseudogap critical point of a high- T_c superconductor. *Nat. Phys.* **5**, 31–34 (2009).

31. 31.

Hayes, I. M. et al. Superconductivity and quantum criticality linked by the Hall effect in a strange metal. *Nat. Phys.* **17**, 58–62 (2020).

32. 32.

Grissonnanche, G. et al. Direct measurement of the upper critical field in cuprate superconductors. *Nat. Commun.* **5**, 3280 (2014).

33. 33.

Fisher, M. P. A., Weichman, P. B., Grinstein, G. & Fisher, D. S. Boson localization and the superfluid–insulator transition. *Phys. Rev. B* **40**, 546–570 (1989).

34. 34.

Eckern, U., Schön, G., & Ambegaokar, V. Quantum dynamics of a superconducting tunnel junction. *Phys. Rev. B* **30**, 6419–6431 (1984).

35. 35.

Aji, V. & Varma, C. M. Theory of the quantum critical fluctuations in cuprate superconductors. *Phys. Rev. Lett.* **99**, 067003 (2007).

36. 36.

Chakravarty, S. *Understanding Quantum Phase Transitions* (CRC Press, 2010).

37. 37.

Zhu, L., Hou, C. & Varma, C. M. Quantum criticality in the two-dimensional dissipative quantum XY model. *Phys. Rev. B* **94**, 235156 (2016).

38. 38.

Wen, L., Xu, R., Mi, Y. & Lei, L. Multiple nanostructures based on anodized aluminium oxide templates. *Nat. Nanotechnol.* **12**, 244–250 (2017).

39. 39.

Chowdhury, D., Werman, Y., Berg, E. & Senthil, T. Translationally invariant non-Fermi-liquid metals with critical Fermi surfaces: solvable models. *Phys. Rev. X* **8**, 031024 (2018).

40. 40.

Patel, A. A., McGreevy, J., Arovas, D. P. & Sachdev, S. Magnetotransport in a model of a disordered strange metal. *Phys. Rev. X* **8**, 021049 (2018).

41. 41.

Segawa, K. & Ando, Y. Transport anomalies and the role of pseudogap in the 60-K phase of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. *Phys. Rev. Lett.* **86**, 4907–4910 (2001).

42. 42.

Tomita, T., Kuga, K., Uwatoko, Y., Coleman, P. & Nakatsuji, S. Strange metal without magnetic criticality. *Science* **349**, 506–509 (2015).

43. 43.

Shen, B. et al. Strange-metal behaviour in a pure ferromagnetic Kondo lattice. *Nature* **579**, 51–55 (2020).

44. 44.

Custers, J. et al. The break-up of heavy electrons at a quantum critical point. *Nature* **424**, 524–527 (2003).

45. 45.

Gegenwart, P., Si, Q. & Steglich, F. Quantum criticality in heavy-fermion metals. *Nat. Phys.* **4**, 186–197 (2008).

46. 46.

Prochaska, L. et al. Singular charge fluctuations at a magnetic quantum critical point. *Science* **367**, 285–288 (2020).

47. 47.

Hartnoll, S. A. & Hofman, D. M. Locally critical resistivities from umklapp scattering. *Phys. Rev. Lett.* **108**, 241601 (2012).

48. 48.

Hartnoll, S. A. Theory of universal incoherent metallic transport. *Nat. Phys.* **11**, 54–61 (2015).

49. 49.

Zaanen, J. Electrons go with the flow in exotic material systems. *Science* **351**, 1026–1027 (2016).

50. 50.

Faulkner, T., Iqbal, N., Liu, H., McGreevy, J. & Vegh, D. Strange metal transport realized by gauge/gravity duality. *Science* **329**, 4013–1047 (2010).

51. 51.

Davison, R. A., Schalm, K. & Zaanen, J. Holographic duality and the resistivity of strange metals. *Phys. Rev. B* **89**, 245116 (2014).

52. 52.

Doniach, S. & Das, D. The Bose metal: a commentary. *Braz. J. Phys.* **33**, 740–743 (2003).

53. 53.

Jaeger, H. M., Haviland, D. B., Orr, B. G. & Goldman, A. M. Onset of superconductivity in ultrathin granular metal films. *Phys. Rev. B* **40**, 182–196 (1989).

54. 54.

Garcia-Barriocanal, J. et al. Electronically driven superconductor–insulator transition in electrostatically doped $\text{La}_2\text{CuO}_{4+\delta}$ thin films. *Phys. Rev. B* **87**, 024509 (2013).

55. 55.

Han, Z. et al. Collapse of superconductivity in a hybrid tin–graphene Josephson junction array. *Nat. Phys.* **10**, 380–386 (2014).

56. 56.

Böttcher, C. G. L. et al. Superconducting, insulating and anomalous metallic regimes in a gated two-dimensional semiconductor–superconductor array. *Nat. Phys.* **14**, 1138–1144 (2018)

57. 57.

Glatz, A., Varlamov, A. & Vinokur, V. Fluctuation spectroscopy of disordered two-dimensional superconductors. *Phys. Rev. B* **84**, 104510 (2011).

58. 58.

Maksimovic, N. et al. Magnetoresistance scaling and the origin of H -linear resistivity in $\text{BaFe}_2(\text{As}_{1-x}\text{P}_x)_2$. *Phys. Rev. X* **10**, 041062 (2020).

59. 59.

Ayres, J. et al. Incoherent transport across the strange-metal regime of overdoped cuprates. *Nature* **595**, 661–666 (2021).

60. 60.

Boyd, C. & Phillips, P. W. Single-parameter scaling in the magnetoresistance of optimally doped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. B* **100**, 155139 (2019).

61. 61.

Rullier-Albenque, F., Alloul, H., Balakirev, F. & Proust, C. Disorder, metal–insulator crossover and phase diagram in high- T_c cuprates. *Europhys. Lett.* **81**, 37008 (2008).

62. 62.

Bollinger, A. T. et al. Superconductor–insulator transition in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ at the pair quantum resistance. *Nature* **472**, 458–460 (2011).

63. 63.

Wang, F., Biscaras, J., Erb, A. & Shukla, A. Superconductor–insulator transition in space charge doped one unit cell $\text{Bi}_{2.1}\text{Sr}_{1.9}\text{CaCu}_2\text{O}_{8+x}$. *Nat. Commun.* **12**, 2926 (2021).

64. 64.

Chakravarty, S., Kivelson, S., Zimanyi, G. T. & Halperin, B. I. Effect of quasiparticle tunneling on quantum-phase fluctuations and the onset of superconductivity in granular films. *Phys. Rev. B* **35**, 7256–7259 (1987).

65. 65.

Kapitulnik, A., Mason, N., Kivelson, S. A. & Chakravarty, S. Effects of dissipation on quantum phase transitions. *Phys. Rev. B* **63**, 125322 (2001).

66. 66.

Kampf, A. & Schön, G. Quantum effects and the dissipation by quasiparticle tunneling in arrays of Josephson junctions. *Phys. Rev. B* **36**, 3651–3660 (1987).

67. 67.

Tikhonov, K. S. & Feigel'man, M. V. Strange metal state near quantum superconductor–metal transition in thin films. *Ann. Phys.* **417**, 168138 (2020).

Acknowledgements

We thank C. M. Varma, H. Yao, A. Lucas and J. M. Kosterlitz for discussions. This work was supported by the National Natural Science Foundation of China (grants 51722204, 51972041, U20A20244, 11888101, 12022407, 11774008 and 12074056), the National Basic Research Program of China (grants 2021YFA0718800, 2017YFA0303300, 2018YFA0305604 and 2017YFA0304600), Beijing Natural Science Foundation (Z180010) and the China National Postdoctoral Program for Innovative Talents (BX2021054).

Author information

Author notes

1. These authors contributed equally: Chao Yang, Haiwen Liu

Affiliations

1. State Key Laboratory of Electronic Thin Films and Integrated Devices, University of Electronic Science and Technology of China, Chengdu, China

Chao Yang, Jiandong Wang, Dong Qiu, Sishuang Wang, Yang Wang, Qianmei He, Xiuli Li, Peng Li, Jie Xiong & Yanrong Li

2. Center for Advanced Quantum Studies, Department of Physics, Beijing Normal University, Beijing, China

Haiwen Liu

3. International Center for Quantum Materials, School of Physics, Peking University, Beijing, China

Yi Liu, Yue Tang, Jian Wang & X. C. Xie

4. Beijing Academy of Quantum Information Sciences, Beijing, China

Jian Wang

5. Department of Physics, Brown University, Providence, RI, USA

James M. Valles Jr

6. Sichuan University, Chengdu, China

Yanrong Li

Contributions

J.X. and J.M.V. Jr conceived the study and supervised the project together with Y. Li and Jian Wang. C.Y. and Jiandong Wang fabricated the samples. C.Y., Y. Liu, S.W., D.Q., Y.W., Q.H., X.L., Y.T. and P.L. performed the experimental measurements. C.Y., H.L., J.M.V. Jr and J.X. analysed the data with contributions from Y. Liu, J.W. and Y. Li. X.C.X. participated in discussions. C.Y., J.M.V. Jr, H.L., J.X. and Y. Li wrote the manuscript with comments from J.W. and X.C.X.

Corresponding authors

Correspondence to [James M. Valles Jr](#) or [Jie Xiong](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Nicholas Breznay, Aavishkar Patel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

[Extended Data Fig. 1 Scanning electron microscopy image of a nanopatterned \$\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}\$ \(YBCO\) thin film.](#)

The 12-nm-thick nanopatterned YBCO thin film was fabricated by reactive ion etching through an anodic aluminium oxide (AAO) membrane directly placed atop the YBCO. By RIE, the anodized aluminium oxide pattern of a triangular array of holes with ~70-nm diameter and ~103-nm period was duplicated onto the YBCO film.

[Extended Data Fig. 2 First derivatives of the \$R\$ - \$T\$ curves of nanopatterned YBCO films.](#)

a–f, The first derivatives of resistance as a function of temperature for f2 (**a**), f3 (**b**), f4 (**c**), f7(**d**) and f8 (**e**). The table shows the yielded parameters of the statistical analysis (**f**).

[Source data](#)

[Extended Data Fig. 3 Data residuals after subtracting the linear fit for the \$R\$ - \$T\$ curves of nanopatterned YBCO films.](#)

Residuals for the R - T curves of f2 (**a**), f3 (**b**), f4 (**c**), f7(**d**) and f8 (**e**). The residual is defined by the resistance subtracting the linear fitting of the R - T curves with the slopes and interceptions shown in **f**. To delineate the

temperature regime for T -linear resistivity, the residual cut-off is set by $50\ \Omega$ which is around 0.5% of the normal-state sheet resistance R_N . The table shows the temperature regime for T -linear resistivity where the residual is within $50\ \Omega$ (f).

[Source data](#)

Extended Data Fig. 4 Nonlinear fitting for the R - T curves of nanopatterned YBCO films.

a–d, Least-squares nonlinear fitting of the R - T curves for f8 (a), f7 (b), f4 (c) and f2 (d). n is the yielded power from the fitting.

[Source data](#)

Extended Data Fig. 5 Scale-invariant B -linear resistance in nanopatterned YBCO thin films under perpendicular magnetic field.

a–f, The magnetoresistance for films: f0 (a), f2 (b), f3 (c), f5 (d), f6 (e) and superconducting (SC; f).

[Source data](#)

Extended Data Fig. 6 First derivatives and nonlinear fitting of the R - B curves of nanopatterned YBCO films.

a, b, The first derivative of resistance as a function of magnetic field for f4 (a) and f2 (b) at various temperatures. **c, d**, Least-squares nonlinear curve fitting of the R - B curves for f4 (c) and f2 (d).

[Source data](#)

Extended Data Fig. 7 B - T scaling in nanopatterned YBCO films.

a–c, $B-T$ scaling in nanopatterned YBCO thin films of f2 (**a**), f3 (**b**) and f5 (**c**).

[Source data](#)

Extended Data Fig. 8 Magnetotransport as a function of $(k_B T + \gamma \mu_B B)/k_B$ of nanopatterned YBCO films.

a–d, The resistance and magnetoresistance of f1 (**a**), f2 (**b**), f3 (**c**) and f5 (**d**) as a function of $(k_B T + \gamma \mu_B B)/k_B$, where the γ parameter can be estimated by adjusting it when the curves collapse best.

[Source data](#)

Extended Data Fig. 9 Electrode pattern for the measurement.

a, Illustration of the electrode pattern for standard four-probe measurements. **b**, Illustration of the electrode pattern for Hall measurements. The current was applied at electrode #1 and #5. The Hall resistance was measured from electrode #3 and #7, the longitudinal resistance is measured from electrode #2 and #3. STO, SrTiO₃.

Extended Data Fig. 10 Current voltage ($I-V$) curves and $R-T$ curves at different current excitations in nanopatterned YBCO film.

a–c, $R-T$ curves for representative films f4 (**a**), f3 (**b**) and f2 (**c**) with different currents. **d**, Current voltage ($I-V$) curves for f4.

[Source data](#)

Supplementary information

Supplementary Information

This file contains Supplementary Figures S1–S6; Supplementary Tables 1 and 2 and Supplementary References.

Supplementary Data 1

Source Data for Supplementary Fig. 1.

Supplementary Data 2

Source Data for Supplementary Fig. 2.

Supplementary Data 3

Source Data for Supplementary Fig. 3.

Supplementary Data 4

Source Data for Supplementary Fig. 4.

Supplementary Data 5

Source Data for Supplementary Fig. 5.

Supplementary Data 6

Source Data for Supplementary Fig. 6.

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 3

[**Source Data Fig. 4**](#)

[**Source Data Extended Data Fig. 2**](#)

[**Source Data Extended Data Fig. 3**](#)

[**Source Data Extended Data Fig. 4**](#)

[**Source Data Extended Data Fig. 5**](#)

[**Source Data Extended Data Fig. 6**](#)

[**Source Data Extended Data Fig. 7**](#)

[**Source Data Extended Data Fig. 8**](#)

[**Source Data Extended Data Fig. 10**](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Yang, C., Liu, H., Liu, Y. *et al.* Signatures of a strange metal in a bosonic system. *Nature* **601**, 205–210 (2022). <https://doi.org/10.1038/s41586-021-04239-y>

- Received: 10 May 2021
- Accepted: 09 November 2021
- Published: 12 January 2022

- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04239-y>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

[**A strange metal emerges from a failed superconductor**](#)

- Nicholas P. Breznay

News & Views 12 Jan 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04239-y>.

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 12 January 2022](#)

A crossbar array of magnetoresistive memory devices for in-memory computing

- [Seungchul Jung](#) ORCID: orcid.org/0000-0003-2727-0791¹,
- [Hyungwoo Lee](#)¹,
- [Sungmeen Myung](#) ORCID: orcid.org/0000-0002-2899-5092¹,
- [Hyunsoo Kim](#) ORCID: orcid.org/0000-0002-5194-0372¹,
- [Seung Keun Yoon](#)¹,
- [Soon-Wan Kwon](#)¹,
- [Yongmin Ju](#)¹,
- [Minje Kim](#)¹,
- [Wooseok Yi](#)¹,
- [Shinhee Han](#)²,
- [Baeseong Kwon](#)²,
- [Boyoung Seo](#)²,
- [Kilho Lee](#)³,
- [Gwan-Hyeob Koh](#)³,
- [Kangho Lee](#)²,
- [Yoonjong Song](#)³,
- [Changkyu Choi](#) ORCID: orcid.org/0000-0002-1470-8662¹,
- [Donhee Ham](#) ORCID: orcid.org/0000-0001-6925-2466^{1,4} &
- [Sang Joon Kim](#) ORCID: orcid.org/0000-0003-2286-3790¹

Nature volume **601**, pages 211–216 (2022)

- 4322 Accesses

- 198 Altmetric
- [Metrics details](#)

Subjects

- [Electrical and electronic engineering](#)
- [Neural circuits](#)

This article has been [updated](#)

Abstract

Implementations of artificial neural networks that borrow analogue techniques could potentially offer low-power alternatives to fully digital approaches^{1,2,3}. One notable example is in-memory computing based on crossbar arrays of non-volatile memories^{4,5,6,7} that execute, in an analogue manner, multiply–accumulate operations prevalent in artificial neural networks. Various non-volatile memories—including resistive memory^{8,9,10,11,12,13}, phase-change memory^{14,15} and flash memory^{16,17,18,19}—have been used for such approaches. However, it remains challenging to develop a crossbar array of spin-transfer-torque magnetoresistive random-access memory (MRAM)^{20,21,22}, despite the technology’s practical advantages such as endurance and large-scale commercialization⁵. The difficulty stems from the low resistance of MRAM, which would result in large power consumption in a conventional crossbar array that uses current summation for analogue multiply–accumulate operations. Here we report a 64×64 crossbar array based on MRAM cells that overcomes the low-resistance issue with an architecture that uses resistance summation for analogue multiply–accumulate operations. The array is integrated with readout electronics in 28-nanometre complementary metal–oxide–semiconductor technology. Using this array, a two-layer perceptron is implemented to classify 10,000 Modified National Institute of Standards and Technology digits with an accuracy of 93.23 per cent (software baseline: 95.24 per cent). In an emulation of a deeper, eight-layer Visual

Geometry Group-8 neural network with measured errors, the classification accuracy improves to 98.86 per cent (software baseline: 99.28 per cent). We also use the array to implement a single layer in a ten-layer neural network to realize face detection with an accuracy of 93.4 per cent.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Fig. 1: MRAM crossbar array.

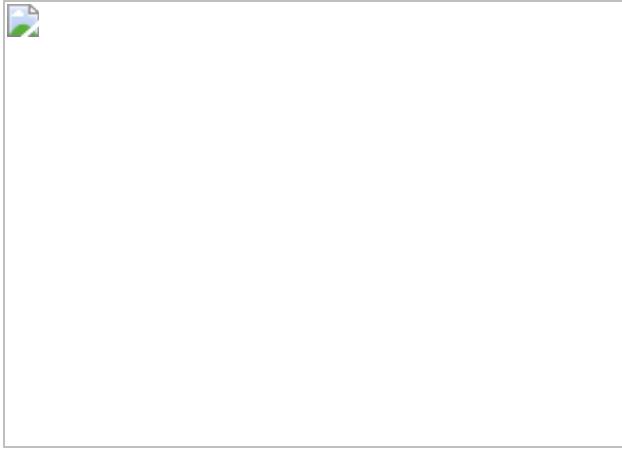


Fig. 2: Characterization of the MRAM crossbar array.



Fig. 3: Classification of 10,000 MNIST handwritten digits with a two-layer perceptron neural network.

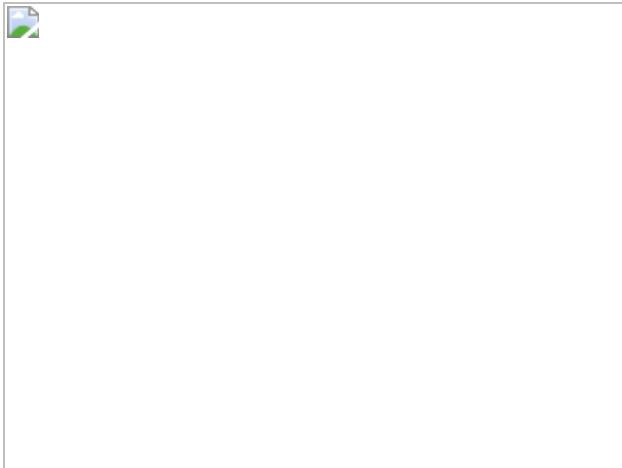
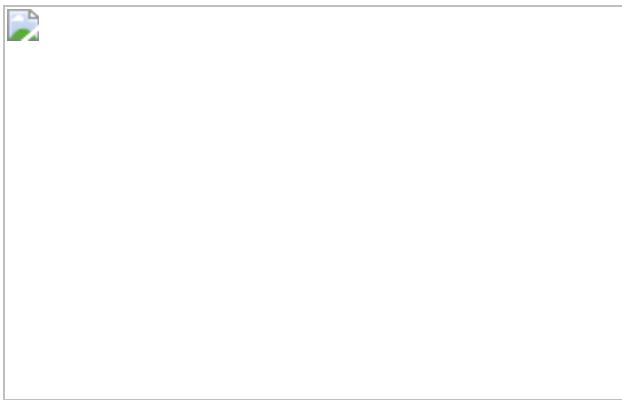


Fig. 4: Face detection for always-on sensing applications in edge devices.



Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Code availability

Computer codes are available from the corresponding authors on reasonable request.

Change history

- **14 January 2022**

In the version of this article initially published online, the posted article PDF was an earlier, incorrect version of the final article. The correct PDF has been posted as of 14 January 2022

References

1. 1. Horowitz, M. Computing's energy problem (and what we can do about it). In *Proc. International Solid-State Circuits Conference (ISSCC)* 10–14 (IEEE, 2014).
2. 2. Keckler, S. W., Dally, W. J., Khailany, B., Garland, M. & Glasco, D. GPUs and the future of parallel computing. *IEEE Micro* **31**, 7–17 (2011).
3. 3. Song, J. et al. An 11.5TOPS/W 1024-MAC butterfly structure dual-core sparsity-aware neural processing unit in 8nm flagship mobile SoC. In *2019 IEEE Int. Solid-State Circuits Conference Digest of Technical Papers (ISSCC)* 130–131 (IEEE, 2019).
4. 4. Sebastian, A. et al. Memory devices and applications for in-memory computing. *Nat. Nanotechnol.* **15**, 529–544 (2020).
5. 5. Wang, Z. et al. Resistive switching materials for information processing. *Nat. Rev. Mater.* **5**, 173–195 (2020).
6. 6.

Ielmini, D. & Wong, H. P. In-memory computing with resistive switching devices. *Nat. Electron.* **1**, 333–343 (2018).

7. 7.

Verma, N. et al. In-memory computing: advances and prospects. *IEEE Solid-State Circuits Mag.* **11**, 43–55 (2019).

8. 8.

Woo, J. et al. Improved synaptic behavior under identical pulses using AlO_x/HfO₂ bilayer RRAM array for neuromorphic systems. *IEEE Electron Device Lett.* **37**, 994–997 (2016).

9. 9.

Yao, P. et al. Face classification using electronic synapses. *Nat. Commun.* **8**, 15199 (2017).

10. 10.

Wu, H. et al. Device and circuit optimization of RRAM for neuromorphic computing. In *2017 IEEE International Electron Devices Meeting* 11.5.1–11.5.4 (IEEE, 2017).

11. 11.

Li, C. et al. Efficient and self-adaptive in-situ learning in multilayer memristor neural networks. *Nat. Commun.* **9**, 2385 (2018).

12. 12.

Chen, W. et al. CMOS-integrated memristive non-volatile computing-in-memory for AI edge processors. *Nat. Electron.* **2**, 420–428 (2019).

13. 13.

Yao, P. et al. Fully hardware-implemented memristor convolutional neural network. *Nature* **577**, 641–646 (2020).

14. 14.

Le Gallo, M. et al. Mixed-precision in-memory computing. *Nat. Electron.* **1**, 246–253 (2018).

15. 15.

Ambrogio, S. et al. Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* **558**, 60–67 (2018).

16. 16.

Merrikh-Bayat, F. et al. High-performance mixed-signal neurocomputing with nanoscale floating-gate memory cell arrays. *IEEE Trans Neural Netw. Learn. Syst.* **29**, 4782–4790 (2018).

17. 17.

Wang, P. et al. Three-dimensional NAND flash for vector-matrix multiplication. *IEEE Trans. VLSI Syst.* **27**, 988–991 (2019).

18. 18.

Xiang, Y. et al. Efficient and robust spike-driven deep convolutional neural networks based on NOR flash computing array. *IEEE Trans. Electron Dev.* **67**, 2329–2335 (2020).

19. 19.

Lin, Y.-Y. et al. A novel voltage-accumulation vector-matrix multiplication architecture using resistor-shunted floating gate flash memory device for low-power and high-density neural network applications. In *2018 IEEE International Electron Devices Meeting* 2.4.1–2.4.4 (IEEE, 2018).

20. 20.

Song, Y. J. et al. Demonstration of highly manufacturable STT-MRAM embedded in 28nm logic. In *2018 IEEE International Electron Devices*

Meeting 18.2.1–18.2.4 (IEEE, 2018).

21. 21.

Lee, Y. K. et al. Embedded STT-MRAM in 28-nm FDSOI logic process for industrial MCU/IoT application. In *2018 IEEE Symposium on VLSI Technology* 181–182 (IEEE, 2018).

22. 22.

Wei, L. et al. A 7Mb STT-MRAM in 22FFL FinFET technology with 4ns read sensing time at 0.9V using write-verify-write scheme and offset-cancellation sensing technique. In *2019 IEEE Int. Solid-State Circuits Conference Digest of Technical Papers* 214–216 (IEEE, 2019).

23. 23.

LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

24. 24.

Yu, S. Neuro-inspired computing with emerging nonvolatile memory. *Proc. IEEE* **106**, 260–285 (2018).

25. 25.

Patil, A. D. et al. An MRAM-based deep in-memory architecture for deep neural networks. In *2019 IEEE International Symposium on Circuits and Systems* (IEEE, 2019).

26. 26.

Zabihi, M. et al. In-memory processing on the spintronic CRAM: from hardware design to application mapping. *IEEE Trans. Comput.* **68**, 1159–1173 (2019).

27. 27.

Kang, S. H. Embedded STT-MRAM for energy-efficient and cost-effective mobile systems. In *2014 IEEE Symposium on VLSI Technology* (IEEE, 2014).

28. 28.

Zeng, Z. M. et al. Effect of resistance-area product on spin-transfer switching in MgO-based magnetic tunnel junction memory cells. *Appl. Phys. Lett.* **98**, 072512 (2011).

29. 29.

Kim, H. & Kwon, S.-W. Full-precision neural networks approximation based on temporal domain binary MAC operations. US patent 17/085,300.

30. 30.

Hung, J.-M. et al. Challenges and trends in developing nonvolatile memory-enabled computing chips for intelligent edge devices. *IEEE Trans. Electron Dev.* **67**, 1444–1453 (2020).

31. 31.

Jiang, Z., Yin, S., Seo, J. & Seok, M. C3SRAM: an in-memory-computing SRAM macro based on robust capacitive coupling computing mechanism. *IEEE J. Solid-State Circuits* **55**, 1888–1897 (2020).

32. 32.

Hubara, I. et al. Binarized neural networks. In *Advances in Neural Information Processing Systems* 4107–4115 (NeurIPS, 2016).

33. 33.

Rastegari, M., Ordonez, V., Redmon, J. & Farhadi, A. XNOR-Net: ImageNet classification using binary convolutional neural networks. In *2016 European Conference on Computer Vision* 525–542 (2016).

34. 34.

Lin, X., Zhao, C. & Pan, W. Towards accurate binary convolutional neural network. In *Advances in Neural Information Processing Systems* 345–353 (NeurIPS, 2017).

35. 35.

Zhuang, B. et al. Structured binary neural networks for accurate image classification and semantic segmentation. In *2019 IEEE Conference on Computer Vision and Pattern Recognition* 413–422 (IEEE, 2019).

36. 36.

Shafiee, A. et al. ISAAC: a convolutional neural network accelerator with in-situ analog arithmetic in crossbars. In *2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture* 14–26 (IEEE, 2016).

37. 37.

Liu, B. et al. Digital-assisted noise-eliminating training for memristor crossbar-based analog neuromorphic computing engine. In *2013 50th ACM/EDAC/IEEE Design Automation Conference* 1–6 (IEEE, 2013).

38. 38.

Wu, B., Iandola, F., Jin, P. H. & Keutzer, K. SqueezeDet: unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving. In *2017 IEEE Conference on Computer Vision and Pattern Recognition* 129–137 (IEEE, 2017).

39. 39.

Ham, D., Park, H., Hwang, S. & Kim, K. Neuromorphic electronics based on copying and pasting the brain. *Nat. Electron.* **4**, 635–644 (2021).

40. 40.

Wang, P. et al. Two-step quantization for low-bit neural networks. In *2018 IEEE Conference on Computer Vision and Pattern Recognition* 4376–4384 (IEEE, 2018).

Acknowledgements

We thank G. Jin (Corporate President of Samsung Electronics), S. Hwang (CEO of Samsung SDS), E. Shim (Corporate EVP of Samsung Electronics) and G. Jeong (Corporate EVP of Samsung Electronics) for technical discussions and support.

Author information

Affiliations

1. Samsung Advanced Institute of Technology, Samsung Electronics, Suwon-si, South Korea

Seungchul Jung, Hyungwoo Lee, Sungmeen Myung, Hyunsoo Kim, Seung Keun Yoon, Soon-Wan Kwon, Yongmin Ju, Minje Kim, Wooseok Yi, Changkyu Choi, Donhee Ham & Sang Joon Kim

2. Foundry Business, Samsung Electronics, Yongin-si, South Korea

Shinhee Han, Baeseong Kwon, Boyoung Seo & Kangho Lee

3. Semiconductor R&D Center, Samsung Electronics, Hwaseong-si, South Korea

Kilho Lee, Gwan-Hyeob Koh & Yoonjong Song

4. John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

Donhee Ham

Contributions

S.J., G.-H.K., Y.S., D.H. and S.J.K. devised this work. S.J., H.L., S.M. and Y.J. designed the analogue circuits, and S.-W.K. and M.K. designed the digital circuits for the CMOS chips. S.H., B.K., B.S., Kilho Lee, Kangho Lee, G.-H.K. and Y.S. characterized the MTJs and optimized the fabrication steps for the MRAM crossbar array. S.J. and Y.J. developed evaluation boards. S.J., H.L., S.M. and W.Y. evaluated the MRAM crossbar array. H.K. and C.C. designed and trained the two-layer, eight-layer and ten-layer neural networks. S.J., S.K.Y. and W.Y. performed the MNIST classification experiments. S.J., S.K.Y., S.J.K. and D.H. analysed dot product errors. S.K.Y. and H.K. developed the emulator. S.-W.K. and M.K. developed the face detection system and performed the face detection experiments. S.J., H.L., S.K.Y., S.J.K. and D.H. wrote the paper. G.-H.K., Y.S., C.C., D.H. and S.J.K. supervised this work. All authors read and discussed the manuscript.

Corresponding authors

Correspondence to [Donhee Ham](#) or [Sang Joon Kim](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Conventional memory crossbar array for ANN computing.

a, Conventional memory crossbar array to perform analogue vector–matrix multiplication. **b**, Vector–matrix multiplication is prevalent in ANN computing: it is used to transfer data from a layer to the next.

Extended Data Fig. 2 MTJ write/read operation.

a, For each column, two write/read data lines were added with access switches. **b**, Example of write operation. **c**, Example of read operation.

Extended Data Fig. 3 Estimated R from data-dependent delay and associated error.

Estimated $R = \tau/C$ calculated from equations (3) and (5) versus true R calculated from equation (4) for a broad variety of IN vectors, with $W = R_H$ for the entire column.

Extended Data Fig. 4 Error distribution after digital offsets.

Error distribution modified from Fig. 2d after applying digital offsets to select columns.

Extended Data Fig. 5 Performance with varying conditions.

a, Measured power efficiency and mean absolute error of the dot products as a function of the supply voltage of the TDC readout electronics for an 11.1 MHz operating frequency. **b**, Measured power efficiency and mean absolute error of the dot products as a function of the operating frequency for a 1.0 V supply voltage for the TDC readout electronics. For both **a** and **b**, each mean absolute error is obtained from 1,600 dot products as in Option 1 in Extended Data Table 1, except for the mean absolute error in

the case of the 1.0 V TDC readout supply and the 11.1 MHz operating frequency, for which the error is calculated from ~4 million dot products (this Option 2 in Extended Data Table 1).

Extended Data Fig. 6 Measurement set-up.

a, Evaluation board containing voltage regulators, clock generators, an MCU and the MRAM crossbar array chip. **b**, The MCU communicates with the PC via USB.

Extended Data Fig. 7 Distribution of dot product errors.

a, $N_C = 1$ and $N_\Delta = -16$ data group. **b**, $N_C = 1$ and $N_\Delta = 16$ data group.

Extended Data Table 1 Measured performance summary

Supplementary information

Supplementary Video 1

This video demonstrates a real-time face detection using the system of Fig. 4c, including a webcam. The stream of images captured by the webcam are shown on the bottom left. When a face appears, the neural network detects it and a smile icon shows on the top left. As the face disappears, the neural network recognizes it and the smile icon disappears. The graph on the bottom right shows the dissipated power in the four MRAM crossbar arrays combined. The 10-layer VGG-like neural network model used here is shown on the top right, where the 7th convolutional layer is realized by using 4 MRAM crossbar arrays.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Jung, S., Lee, H., Myung, S. *et al.* A crossbar array of magnetoresistive memory devices for in-memory computing. *Nature* **601**, 211–216 (2022). <https://doi.org/10.1038/s41586-021-04196-6>

- Received: 30 December 2020
- Accepted: 28 October 2021
- Published: 12 January 2022
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04196-6>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04196-6>

- Article
- [Published: 12 January 2022](#)

Elastomeric electrolytes for high-energy solid-state lithium batteries

- [Michael J. Lee](#) [ORCID: orcid.org/0000-0001-5087-3991¹](#), [na1](#),
- [Junghun Han](#) [ORCID: orcid.org/0000-0003-3064-0511²](#), [na1](#),
- [Kyungbin Lee](#) [ORCID: orcid.org/0000-0003-1046-7811¹](#),
- [Young Jun Lee](#) [ORCID: orcid.org/0000-0001-6406-890X²](#),
- [Byoung Gak Kim³](#),
- [Kyu-Nam Jung⁴](#),
- [Bumjoon J. Kim](#) [ORCID: orcid.org/0000-0001-7783-9689²](#) &
- [Seung Woo Lee](#) [ORCID: orcid.org/0000-0002-2695-7105¹](#)

[Nature](#) volume **601**, pages 217–222 (2022)

- 5368 Accesses
- 142 Altmetric
- [Metrics details](#)

Subjects

- [Batteries](#)
- [Polymers](#)

Abstract

The use of lithium metal anodes in solid-state batteries has emerged as one of the most promising technologies for replacing conventional lithium-ion

batteries^{1,2}. Solid-state electrolytes are a key enabling technology for the safe operation of lithium metal batteries as they suppress the uncontrolled growth of lithium dendrites. However, the mechanical properties and electrochemical performance of current solid-state electrolytes do not meet the requirements for practical applications of lithium metal batteries. Here we report a class of elastomeric solid-state electrolytes with a three-dimensional interconnected plastic crystal phase. The elastomeric electrolytes show a combination of mechanical robustness, high ionic conductivity, low interfacial resistance and high lithium-ion transference number. The *in situ*-formed elastomer electrolyte on copper foils accommodates volume changes for prolonged lithium plating and stripping processes with a Coulombic efficiency of 100.0 per cent. Moreover, the elastomer electrolytes enable stable operation of the full cells under constrained conditions of a limited lithium source, a thin electrolyte and a high-loading $\text{LiNi}_{0.83}\text{Mn}_{0.06}\text{Co}_{0.11}\text{O}_2$ cathode at a high voltage of 4.5 volts at ambient temperature, delivering a high specific energy exceeding 410 watt-hours per kilogram of electrode plus electrolyte. The elastomeric electrolyte system presents a powerful strategy for enabling stable operation of high-energy, solid-state lithium batteries.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Fig. 1: Design of plastic-crystal-embedded elastomer electrolyte.

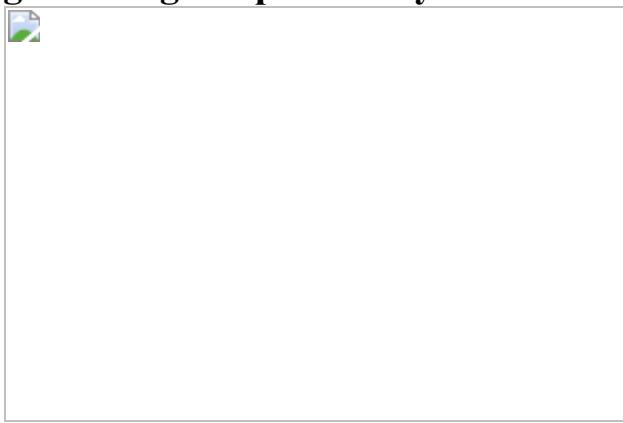


Fig. 2: Properties of the built-in PCEE.

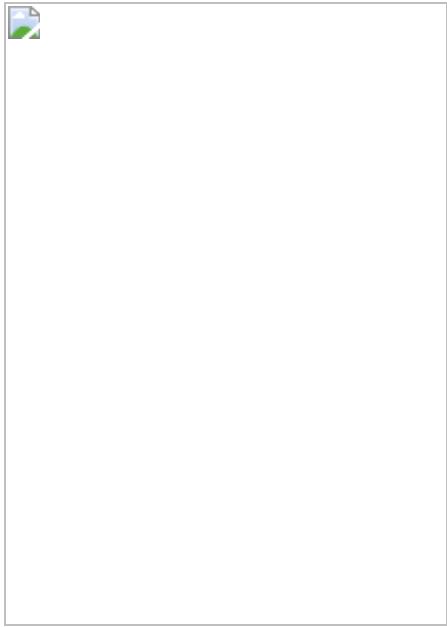


Fig. 3: Built-in PCEE in symmetric Li and asymmetric Li||Cu cells.

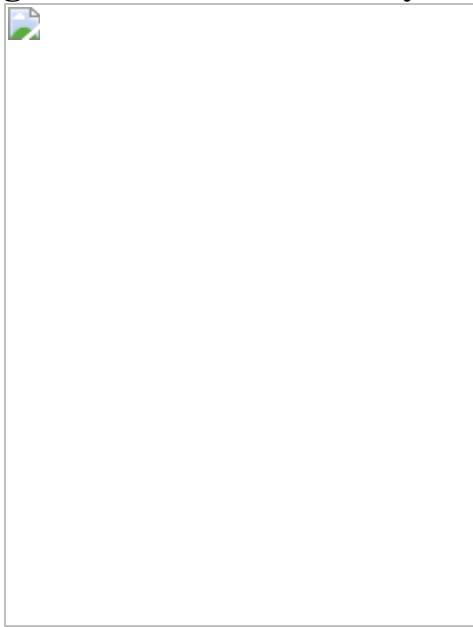
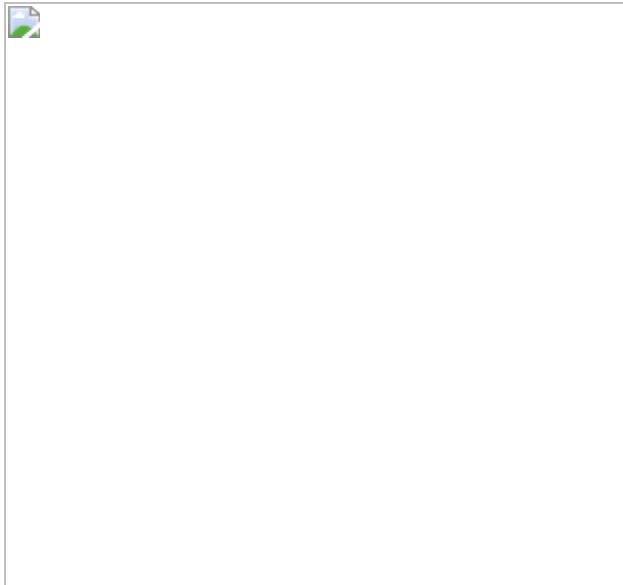


Fig. 4: Towards a high-energy all-solid-state LMB with elastomeric electrolyte.



Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request. [Source data](#) are provided with this paper.

References

1. 1.

Armand, M. & Tarascon, J.-M. Building better batteries. *Nature* **451**, 652–657 (2008).

2. 2.

Choi, J. W. & Aurbach, D. Promise and reality of post-lithium-ion batteries with high energy densities. *Nat. Rev. Mater.* **1**, 16013 (2016).

3. 3.

Bai, P., Li, J., Brushett, F. R. & Bazant, M. Z. Transition of lithium growth mechanisms in liquid electrolytes. *Energy Environ. Sci.* **9**, 3221–3229 (2016).

4. 4.

Lin, D. C., Liu, Y. & Cui, Y. Reviving the lithium metal anode for high-energy batteries. *Nat. Nanotechnol.* **12**, 194–206 (2017).

5. 5.

Wan, J. Y. et al. Ultrathin, flexible, solid polymer composite electrolyte enabled with aligned nanoporous host for lithium batteries. *Nat. Nanotechnol.* **14**, 705–711 (2019).

6. 6.

Lee, Y.-G. et al. High-energy long-cycling all-solid-state lithium metal batteries enabled by silver-carbon composite anodes. *Nat. Energy* **5**, 299–308 (2020).

7. 7.

Liu, W. et al. Enhancing ionic conductivity in composite polymer electrolytes with well-aligned ceramic nanowires. *Nat. Energy* **2**, 17035 (2017).

8. 8.

Zhao, Q., Liu, X. T., Stalin, S., Khan, K. & Archer, L. A. Solid-state polymer electrolytes with in-built fast interfacial transport for secondary lithium batteries. *Nat. Energy* **4**, 365–373 (2019).

9. 9.

Zhang, X. Y. et al. Long cycling life solid-state Li metal batteries with stress self-adapted Li/garnet interface. *Nano Lett.* **20**, 2871–2878 (2020).

10. 10.

Zhu, Y., Cao, J., Chen, H., Yu, Q. & Li, B. High electrochemical stability of a 3D cross-linked network PEO@nano-SiO₂ composite

polymer electrolyte for lithium metal batteries. *J. Mater. Chem. A* **7**, 6832–6839 (2019).

11. 11.

Lee, W. et al. Ceramic-salt composite electrolytes from cold sintering. *Adv. Funct. Mater.* **29**, 1807872 (2019).

12. 12.

Yang, X. et al. Determining the limiting factor of the electrochemical stability window for PEO-based solid polymer electrolytes: main chain or terminal –OH group? *Energy Environ. Sci.* **13**, 1318–1325 (2020).

13. 13.

Chen, R.-J. et al. Addressing the interface issues in all-solid-state bulk-type lithium ion battery via an all-composite approach. *ACS Appl. Mater. Interfaces* **9**, 9654–9661 (2017).

14. 14.

Bouchet, R. et al. Single-ion BAB triblock copolymers as highly efficient electrolytes for lithium-metal batteries. *Nat. Mater.* **12**, 452–457 (2013).

15. 15.

Zhou, D. et al. In situ synthesis of a hierarchical all-solid-state electrolyte based on nitrile materials for high-performance lithium-ion batteries. *Adv. Energy Mater.* **5**, 1500353 (2015).

16. 16.

Jiang, T. et al. Solvent-free synthesis of thin, flexible, nonflammable garnet-based composite solid electrolyte for all-solid-state lithium batteries. *Adv. Energy Mater.* **10**, 1903376 (2020).

17. 17.

Markvicka, E. J., Bartlett, M. D., Huang, X. & Majidi, C. An autonomously electrically self-healing liquid metal–elastomer composite for robust soft-matter robotics and electronics. *Nat. Mater.* **17**, 618–624 (2018).

18. 18.

Pan, C. et al. A liquid-metal–elastomer nanocomposite for stretchable dielectric materials. *Adv. Mater.* **31**, 1900663 (2019).

19. 19.

Kim, H. J., Chen, B., Suo, Z. & Hayward, R. C. Ionoelastomer junctions between polymer networks of fixed anions and cations. *Science* **367**, 773–776 (2020).

20. 20.

Park, M. et al. Highly stretchable electric circuits from a composite material of silver nanoparticles and elastomeric fibres. *Nat. Nanotechnol.* **7**, 803–809 (2012).

21. 21.

Chen, L. et al. PEO/garnet composite electrolytes for solid-state lithium batteries: From “ceramic-in-polymer” to “polymer-in-ceramic”. *Nano Energy* **46**, 176–184 (2018).

22. 22.

Wang, F. et al. Progress report on phase separation in polymer solutions. *Adv. Mater.* **31**, 1806733 (2019).

23. 23.

Seo, M. & Hillmyer, M. A. Reticulated nanoporous polymers by controlled polymerization-induced microphase separation. *Science* **336**, 1422–1425 (2012).

24. 24.

Schulze, M. W., McIntosh, L. D., Hilmyer, M. A. & Lodge, T. P. High-modulus, high-conductivity nanostructured polymer electrolyte membrane via polymerization-induced phase separation. *Nano Lett.* **14**, 122–126 (2014).

25. 25.

Alarco, P.-J., Abu-Lebdeh, Y., Abouimrane, A. & Armand, M. The plastic-crystalline phase of succinonitrile as a universal matrix for solid-state ionic conductors. *Nat. Mater.* **3**, 476–481 (2004).

26. 26.

Choi, K.-H. et al. Thin, deformable, and safety-reinforced plastic crystal polymer electrolytes for high-performance flexible lithium-ion batteries. *Adv. Funct. Mater.* **24**, 44–52 (2014).

27. 27.

White, T. J., Natarajan, L. V., Tondiglia, V. P., Bunning, T. J. & Guymon, C. A. Polymerization kinetics and monomer functionality effects in thiol-ene polymer dispersed liquid crystals. *Macromolecules* **40**, 1112–1120 (2007).

28. 28.

Serbutoviez, C., Kloosterboer, J. G., Boots, H. M. J. & Touwslager, F. J. Polymerization-induced phase separation. 2. Morphology of polymer-dispersed liquid crystal thin films. *Macromolecules* **29**, 7690–7698 (1996).

29. 29.

Phillip, W. A. et al. Diffusion and flow across nanoporous polydicyclopentadiene-based membranes. *ACS Appl. Mater. Interfaces* **1**, 472–480 (2009).

30. 30.

Watson, B. L., Rolston, N., Printz, A. D. & Dauskardt, R. H. Scaffold-reinforced perovskite compound solar cells. *Energy Environ. Sci.* **10**, 2500–2508 (2017).

31. 31.

Meng, J., Chu, F., Hu, J. & Li, C. Liquid polydimethylsiloxane grafting to enable dendrite-free Li plating for highly reversible Li-metal batteries. *Adv. Funct. Mater.* **29**, 1902220 (2019).

32. 32.

Albertus, P., Babinec, S., Litzelman, S. & Newman, A. Status and challenges in enabling the lithium metal electrode for high-energy and low-cost rechargeable batteries. *Nat. Energy* **3**, 16–21 (2018).

33. 33.

Bruce, P. G., Evans, J. & Vincent, C. A. Conductivity and transference number measurements on polymer electrolytes. *Solid State Ion.* **28**, 918–922 (1988).

34. 34.

Diederichsen, K. M., McShane, E. J. & McCloskey, B. D. Promising routes to a high Li^+ transference number electrolyte for lithium ion batteries. *ACS Energy Lett.* **2**, 2563–2575 (2017).

35. 35.

Timachova, K., Watanabe, H. & Balsara, N. P. Effect of molecular weight and salt concentration on ion transport and the transference number in polymer electrolytes. *Macromolecules* **48**, 7882–7888 (2015).

36. 36.

He, M. et al. Fluorinated electrolytes for 5-V Li-ion chemistry: probing voltage stability of electrolytes with electrochemical floating test. *J. Electrochem. Soc.* **162**, A1725–A1729 (2015).

37. 37.

Randau, S. et al. Benchmarking the performance of all-solid-state lithium batteries. *Nat. Energy* **5**, 259–270 (2020).

38. 38.

Duan, H. et al. Extended electrochemical window of solid electrolytes via heterogeneous multilayered structure for high-voltage lithium metal batteries. *Adv. Mater.* **31**, 1807789 (2019).

39. 39.

Yu, X. et al. Selectively wetted rigid-flexible coupling polymer electrolyte enabling superior stability and compatibility of high-voltage lithium metal batteries. *Adv. Energy Mater.* **10**, 1903939 (2020).

40. 40.

Lopez, J. et al. A dual-crosslinking design for resilient lithium-ion conductors. *Adv. Mater.* **30**, 1804142 (2018).

41. 41.

Zhang, W., Nie, J., Li, F., Wang, Z. L. & Sun, C. A durable and safe solid-state lithium battery with a hybrid electrolyte membrane. *Nano Energy* **45**, 413–419 (2018).

42. 42.

Wang, C. et al. Solid-state plastic crystal electrolytes: effective protection interlayers for sulfide-based all-solid-state lithium metal batteries. *Adv. Funct. Mater.* **29**, 1900392 (2019).

43. 43.

Sun, J. et al. Hierarchical composite-solid-electrolyte with high electrochemical stability and interfacial regulation for boosting ultra-stable lithium batteries. *Adv. Funct. Mater.* **31**, 2006381 (2021).

44. 44.

Yao, P. et al. PVDF/palygorskite nanowire composite electrolyte for 4 V rechargeable lithium batteries with high energy density. *Nano Lett.* **18**, 6113–6120 (2018).

45. 45.

Fu, C. et al. A dual-salt coupled fluoroethylene carbonate succinonitrile-based electrolyte enables Li-metal batteries. *J. Mater. Chem. A* **8**, 2066–2073 (2020).

46. 46.

Mackanic, D. G. et al. Decoupling of mechanical properties and ionic conductivity in supramolecular lithium ion conductors. *Nat. Commun.* **10**, 5384 (2019).

47. 47.

Fu, C. et al. Universal chemomechanical design rules for solid-ion conductors to prevent dendrite formation in lithium metal batteries. *Nat. Mater.* **19**, 758–766 (2020).

48. 48.

Xia, S. et al. High-rate and large-capacity lithium metal anode enabled by volume conformal and self-healable composite polymer electrolyte. *Adv. Sci.* **6**, 1802353 (2019).

49. 49.

Liu, Y. et al. Lithium-coated polymeric matrix as a minimum volume-change and dendrite-free lithium metal anode. *Nat. Commun.* **7**, 10992 (2016).

50. 50.

Chen, T. et al. Ionic liquid-immobilized polymer gel electrolyte with self-healing capability, high ionic conductivity and heat resistance for dendrite-free lithium metal batteries. *Nano Energy* **54**, 17–25 (2018).

51. 51.

Lu, Q. et al. Dendrite-free, high-rate, long-life lithium metal batteries with a 3D cross-linked network polymer electrolyte. *Adv. Mater.* **29**, 1604460 (2017).

52. 52.

Dong, T. et al. A multifunctional polymer electrolyte enables ultra-long cycle-life in a high-voltage lithium metal battery. *Energy Environ. Sci.* **11**, 1197–1203 (2018).

Acknowledgements

This work was performed in part at the Georgia Tech Institute for Electronics and Nanotechnology, a member of the National Nanotechnology Coordinated Infrastructure (NNCI), which is supported by the National Science Foundation (ECCS-2025462). This work was partially supported by KRICT core project (BSF20-242). J.H., Y.J.L. and B.J.K. acknowledge the support from the National Research Foundation of Korea (NRF-2019R1A2B5B03101123, 2017M3D1A1039553 and 2020R1A4A1018516).

Author information

Author notes

1. These authors contributed equally: Michael J. Lee, Junghun Han

Affiliations

1. George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Michael J. Lee, Kyungbin Lee & Seung Woo Lee

2. Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

Junghun Han, Young Jun Lee & Bumjoon J. Kim

3. Division of Advanced Materials, Korea Research Institute of Chemical Technology, Daejeon, Republic of Korea

Byoung Gak Kim

4. Renewable Energy Institute, Korea Institute of Energy Research, Daejeon, Republic of Korea

Kyu-Nam Jung

Contributions

S.W.L., B.J.K., M.J.L. and J.H. conceived the ideas and designed the experiments. K.L., Y.J.L., B.G.K. and K.-N.J. were involved with the methods and characterizations of materials. S.W.L., B.J.K., M.J.L. and J.H. co-wrote the manuscript. All authors reviewed the manuscript.

Corresponding authors

Correspondence to [Bumjoon J. Kim](#) or [Seung Woo Lee](#).

Ethics declarations

Competing interests

S.W.L., B.J.K., M.J.L. and J.H. have filed a US provisional patent application (63/209,140) covering the materials and lithium metal battery application described in this paper.

Peer review information

Nature thanks Enrique Gomez and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

[Extended Data Fig. 1 Fabrication process of in situ-polymerized PCEE within the electrochemical cell.](#)

a, Digital photo images of a homogeneous solution consisting of BA, SN, LiTFSI, PEGDA and AIBN for built-in polymerization (left) and haze-coloured PCEE on the bottom of a glass vial after polymerization at 70 °C for 2 h (right). **b**, Photo image of PCEE showing mechanical elasticity. **c**, Schematic illustration of the built-in polymerization process. The solution was injected into the electrochemical cells and then heated in an oven for built-in polymerization.

[Extended Data Fig. 2 Comparison of morphology, ion conductivity and mechanical property between polymerization induced-PCEE and blend systems.](#)

a, Morphology of PCEE. SEM image of PCEE shows the continuously-connected SN phases within the elastomeric matrix that was uniformly developed over a large area through PIPS. **b**, Morphology of blend consisting of elastomeric polymer (cross-linked poly(butylacrylate) and PEGDA) and plastic crystal (SN) with LiTFSI. A blend is prepared from a mixture of elastomeric polymers and SN–LiTFSI in chloroform, followed by a drying process. The same weight ratios of BA, SN, PEGDA and LiTFSI are used for constructing the PCEE and blend systems. SEM image of blend shows a macrophase separation with a length-scale of over μm . **c–d**, Comparison of ionic conductivity (**c**) and toughness (**d**) between the PCEE and blend systems.

[Source data](#)

Extended Data Fig. 3 Electrochemical characterization for the symmetric Li cells with built-in PCEE.

a, Time-dependent Nyquist plots of the symmetric Li cells configured with built-in PCEE. **b**, Nyquist plots of the symmetric Li cells configured with built-in PCEE after 25, 75, and 100 cycles. **c**, Cycling performance of the symmetric Li cells configured with built-in PCEE at different current densities. **d**, Voltage hysteresis of Li plating/stripping for built-in PCEE compared with previously reported literature data^{8,12,16,38,39,47,48,49,50,51,52}. **e**, Nyquist plots of the symmetric Li cells before and after polarization of 10 mV. **f**, Steady-state current measurement of the symmetric Li cells under 10 mV polarization for 10 h. EIS was measured at open-circuit voltage in the range of 10^5 to 10^0 Hz with an amplitude of 10 mV.

[Source data](#)

Extended Data Fig. 4 Characterization of the SEI components on the cycled Li metal anodes with built-in PCEE and SN100 by XPS.

The high-resolution Li 1s, C 1s, O 1s, N 1s, and F 1s XPS spectra of the Li metal anodes were measured after 100 cycles of the symmetric Li cells with

built-in PCEE and SN100 at a current density of 1 mA cm^{-2} with a capacity of 1 mAh cm^{-2} .

Source data

Extended Data Fig. 5 Li plating and stripping behaviour of built-in PCEE on bare Cu.

a, Cycling performance of the asymmetric Li||Cu cells at current densities of 0.5 and 1 mA cm^{-2} , respectively. **b–c**, Li stripping and plating profiles for built-in PCEE at a current density of 0.5 mA cm^{-2} with a capacity of 1 mAh cm^{-2} (**b**), and a current density of 1 mA cm^{-2} with a capacity of 2 mAh cm^{-2} (**c**).

Extended Data Fig. 6 Electrochemical stability of built-in PCEE paired with high-voltage NMC-622 cathode.

a, Electrochemical floating experiment was performed using Li||NMC-622 with built-in PCEE. The cell was charged to 4.2 V at 0.2C ($1 \text{ C} = 180 \text{ mA g}^{-1}$) and then held at gradually higher voltages for 10 h up to 4.7 V. **b**, Rate capability of the full cell (35- μm -thick Li anode; 25- μm -thick built-in PCEE; high-loading NMC-622 (9.7 mg cm^{-2}) in the voltage range of 2.7–4.5 V at equal current densities. (Inset: the capacity utilization at different areal current densities). **c**, Cycling performance of the full cell (excess Li; 25- μm -thick built-in PCEE; NMC-622 (2.1 mg cm^{-2})) as a function of cycle number in the voltage range of 2.7–4.5 V. The cell maintained a high capacity of $\sim 140 \text{ mAh g}^{-1}$ (82% capacity retention) with high CEs of 99.5% for 100 cycles, confirming the stable operation at high voltage. Cells were performed at 20°C .

Extended Data Fig. 7 Cycling performance of the Li||LiFePO₄ cell at 1 C without voltage holding.

a, Capacity and Coulombic efficiency as a function of cycle number. **b**, Corresponding voltage profiles. $1 \text{ C} = 170 \text{ mA g}^{-1}$.

Extended Data Fig. 8 Electrochemical performances of the full cells with high-voltage NMC-83 cathode.

a, The charge and discharge profiles of the full cell in the voltage range of 2.7–4.3V at 0.1 mA cm⁻². **b**, Temperature-dependent voltage profiles of the full cell charged/discharged at equal temperatures (60 to 0 °C) in the voltage range of 2.7–4.5 V. (Inset: the capacity utilization at different temperatures). All full cells were configured with 35-μm-thick Li anode; 25-μm-thick built-in PCEE; high-loading NMC-83 (>10 mg cm⁻²).

Extended Data Table 1 Comparison of battery performance with previously reported solid-state LMBs

Supplementary information

Supplementary Figures

This file contains Supplementary Figures 1–14.

Peer Review File

Source data

Source Data Extended Data Fig. 2

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 4

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Lee, M.J., Han, J., Lee, K. *et al.* Elastomeric electrolytes for high-energy solid-state lithium batteries. *Nature* **601**, 217–222 (2022).
<https://doi.org/10.1038/s41586-021-04209-4>

- Received: 15 March 2021
- Accepted: 05 November 2021
- Published: 12 January 2022
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04209-4>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04209-4>

- Article
- [Published: 12 January 2022](#)

The effect of rainfall changes on economic production

- [Maximilian Kotz](#) ORCID: orcid.org/0000-0003-2564-5043^{1,2},
- [Anders Levermann](#) ORCID: orcid.org/0000-0003-4432-4704^{1,2,3} &
- [Leonie Wenz](#) ORCID: orcid.org/0000-0002-8500-1568^{1,4}

[Nature](#) volume 601, pages 223–227 (2022)

- 2768 Accesses
- 406 Altmetric
- [Metrics details](#)

Subjects

- [Environmental economics](#)
- [Environmental health](#)
- [Interdisciplinary studies](#)

Abstract

Macro-economic assessments of climate impacts lack an analysis of the distribution of daily rainfall, which can resolve both complex societal impact channels and anthropogenically forced changes^{1,2,3,4,5,6}. Here, using a global panel of subnational economic output for 1,554 regions worldwide over the past 40 years, we show that economic growth rates are reduced by increases in the number of wet days and in extreme daily rainfall, in

addition to responding nonlinearly to the total annual and to the standardized monthly deviations of rainfall. Furthermore, high-income nations and the services and manufacturing sectors are most strongly hindered by both measures of daily rainfall, complementing previous work that emphasized the beneficial effects of additional total annual rainfall in low-income, agriculturally dependent economies^{4,7}. By assessing the distribution of rainfall at multiple timescales and the effects on different sectors, we uncover channels through which climatic conditions can affect the economy. These results suggest that anthropogenic intensification of daily rainfall extremes^{8,9,10} will have negative global economic consequences that require further assessment by those who wish to evaluate the costs of anthropogenic climate change.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.
VAT will be added later in the checkout.
Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Fig. 1: Assessing the distribution of daily rainfall via thresholds.



Fig. 2: The effect of four rainfall measures on economic growth rates.

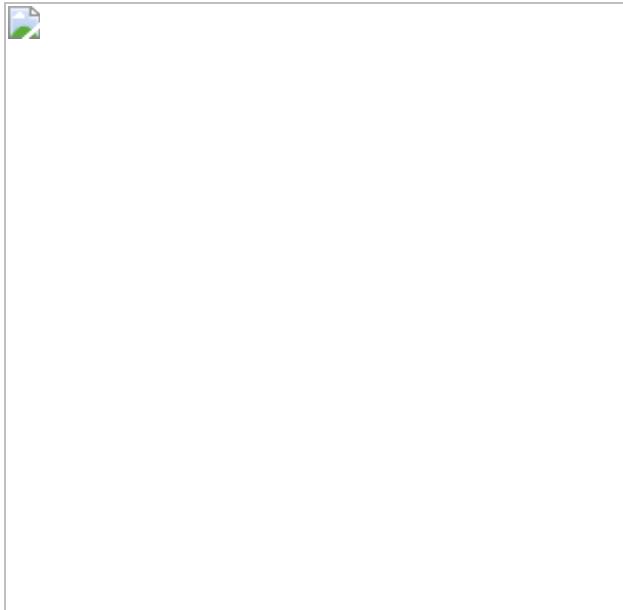


Fig. 3: Regional estimates of the historical effect on economic growth rates of a 1-s.d. shock in each of the four rainfall measures.

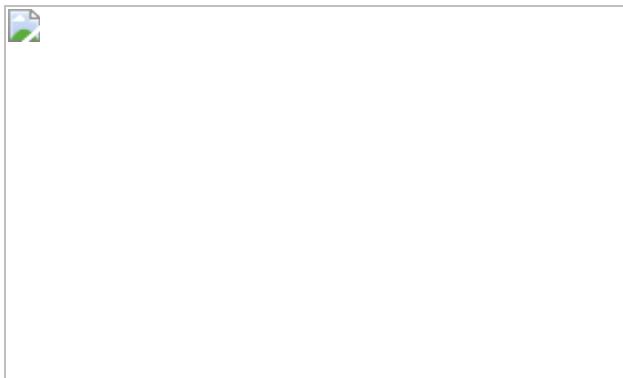
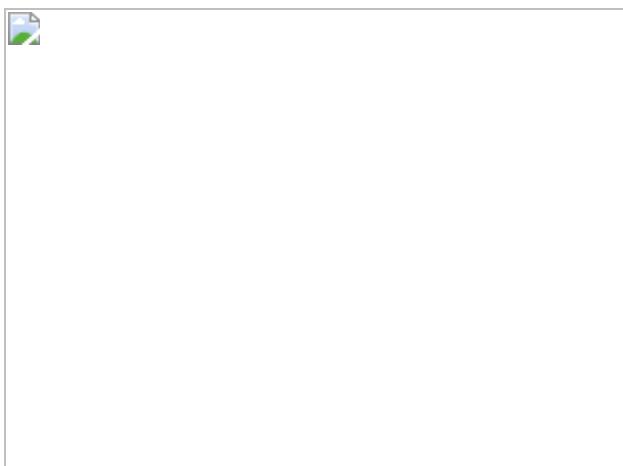


Fig. 4: Assessing the heterogeneity of the effect of rainfall by income and sector.



Data availability

The data on economic production and the ERA-5 climate data are both publicly available at <https://doi.org/10.5281/zenodo.4681306> and <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>, respectively. Secondary data are available at the public repository for this publication: <https://doi.org/10.5281/zenodo.5657457>. The maps were created using Matplotlib v. 3.4.2 (<https://matplotlib.org/>), Cartopy v.0.18.0 (Met Office UK, <https://pypi.python.org/pypi/Cartopy/0.18.0>), Geopandas v. 0.6.1 (<https://geopandas.org/>) and GADM administrative boundaries (<https://gadm.org/>). [Source data](#) are provided with this paper.

Code availability

The code to reproduce the analysis is available at the public repository for this publication: <https://doi.org/10.5281/zenodo.5657457>.

References

1. 1.
Dell, M., Jones, B. F. & Olken, B. A. Temperature shocks and economic growth: evidence from the last half century. *Am. Econ. J. Macroecon.* **4**, 66–95 (2012).
2. 2.
Burke, M., Hsiang, S. & Miguel, E. Global non-linear effect of temperature on economic production. *Nature* **527**, 235–239 (2015).
3. 3.
Kalkuhl, M. & Wenz, L. The impact of climate conditions on economic production. Evidence from a global panel of regions. *J. Environ. Econ. Manage.* **103**, 102360 (2020).

4. 4.

Damania, R., Desbureaux, S. & Zaveri, E. Does rainfall matter for economic growth? Evidence from global sub-national data (1990–2014). *J. Environ. Econ. Manage.* **102**, 102335 (2020).

5. 5.

Holtermann, L. Precipitation anomalies, economic production, and the role of “first-nature” and “second-nature” geographies: a disaggregated analysis in high-income countries. *Glob. Environ. Change* **65**, 102167 (2020).

6. 6.

Kotz, M. et al. Day-to-day temperature variability reduces economic growth. *Nat. Clim. Change* **11**, 319–325 (2021).

7. 7.

Barrios, S., Bertinelli, L. & Strobl, E. Trends in rainfall and economic growth in Africa: a neglected cause of the African growth tragedy. *Rev. Econ. Stat.* **92**, 350–366 (2010).

8. 8.

Min, S. K. et al. Human contribution to more-intense precipitation extremes. *Nature* **470**, 378–381 (2011).

9. 9.

Madakumbura, G. D. et al. Anthropogenic influence on extreme precipitation over global land areas seen in multiple observational datasets. *Nat. Commun.* **12**, 3944 (2021).

10. 10.

Fischer, E. & Knutti, R. Observed heavy precipitation increase confirms theory and early models. *Nat. Clim. Change* **6**, 986–991

(2016).

11. 11.

Marvel, K. & Bonfils, C. External influences on global precipitation. *Proc. Natl Acad. Sci. USA* **110**, 19301–19306 (2013).

12. 12.

Chadwick, R. et al. Large rainfall changes consistently projected over substantial areas of tropical land. *Nat. Clim. Change* **6**, 177–181 (2016).

13. 13.

Lehmann, J., Mempel, F. & Coumou, D. Increased occurrence of record-wet and record-dry months reflect changes in mean rainfall. *Geophys. Res. Lett.* **45**, 13468–13476 (2018).

14. 14.

Zhang, W. et al. Increasing precipitation variability on daily-to-multiyear time scales in a warming world. *Sci. Adv.* **7**, eabf8021 (2021).

15. 15.

Feng, X., Porporato, A. & Rodriguez-Iturbe, I. Changes in rainfall seasonality in the tropics. *Nat. Clim. Change* **3**, 811–815 (2013).

16. 16.

liang, X. Z. et al. Determining climate effects on US total agricultural productivity. *Proc. Natl Acad. Sci. USA* **114**, E2285–E2292 (2017).

17. 17.

Damania, R. The economics of water scarcity and variability. *Oxf. Rev. Econ. Policy* **36**, 24–44 (2020).

18. 18.

Desbureaux, S. & Rodella, A. S. Drought in the city: the economic impact of water scarcity in Latin American metropolitan areas. *World Dev.* **114**, 13–27 (2019).

19. 19.

Hsiang, S. M., Burke, M. & Miguel, E. Quantifying the influence of climate on human conflict. *Science* <https://doi.org/10.1126/science.1235367> (2013).

20. 20.

Davenport, F. V., Burke, M. & Diffenbaugh, N. S. Contribution of historical precipitation change to US flood damages. *Proc. Natl Acad. Sci. USA* **118**, e2017524118 (2021).

21. 21.

Willner, S. N., Otto, S. N. C. & Levermann, A. Global economic response to river floods. *Nat. Clim. Change* **8**, 594–598 (2018).

22. 22.

Beck, H. E. et al. Daily evaluation of 26 precipitation datasets using stage-IV gauge-radar data for the CONUS. *Hydrol. Earth Syst. Sci.* **23**, 207–224 (2019).

23. 23.

Garrick, D. E. et al. Valuing water for sustainable development. *Science* **358**, 1003–1005 (2017).

24. 24.

Ali, S. Climate change and economic growth in a rain-fed economy: How much does rainfall variability cost Ethiopia? *SSRN* <https://doi.org/10.2139/ssrn.2018233> (2012).

25. 25.

Sangkhaphan, S. & Shu, Y. The effect of rainfall on economic growth in Thailand: a blessing for poor provinces. *Economies* **8**, 1–17 (2020).

26. 26.

Loayza, N. V., Olaberría, E., Rigolini, J. & Christiaensen, L. Natural disasters and growth: going beyond the averages. *World Dev.* **40**, 1317–1336 (2012).

27. 27.

Kirchene, H. et al. Long-term impacts of tropical cyclones and fluvial floods on economic growth—empirical evidence on transmission channels at different levels of development. *World Dev.* **144**, 105475 (2021).

28. 28.

Moore, F. C. & Diaz, D. B. Temperature impacts on economic growth warrant stringent mitigation policy. *Nat. Clim. Change* **5**, 127–131 (2015).

29. 29.

Kikstra, J. S. et al. The social cost of carbon dioxide under climate–economy feedbacks and temperature variability. *Environ. Res. Lett.* **16**, 094037 (2021).

30. 30.

Hidalgo, C. A., Klinger, B., Barabási, A.-L. & Hausmann, R. The product space conditions the development of nations. *Science* **317**, 482–487 (2007).

31. 31.

Beck, H. E. et al. MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.* **21**, 589–615 (2017).

32. 32.

Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **19**, 3088–3111 (2006).

33. 33.

Wenz, L., Kalkuhl, M. & Kotz, M. DOSE v.1. The MCC and PIK database of subnational economic output: documentation. *Zenodo* <https://doi.org/10.5281/zenodo.4681306> (2021).

34. 34.

Bador, M., Alexander, L. V., Contractor, S. & Roca, R. Diverse estimates of annual maxima daily precipitation in 22 state-of-the-art quasi-global land observation datasets. *Environ. Res. Lett.* **51**, 035005 (2020).

35. 35.

Myhre, G. et al. Frequency of extreme precipitation increases extensively with event rareness under global warming. *Sci. Rep.* **9**, 16063 (2019).

36. 36.

Klein Goldewijk, K., Beusen, A., van Drecht, G. & de Vos, G. M. The HYDE 3.1 spatially explicit database of human induced land use change over the past 12,000 years. *Glob. Ecol. Biogeogr.* **20**, 73–86 (2011).

37. 37.

Swain, D. L. et al. Increasing precipitation volatility in twenty-first-century California. *Nat. Clim. Change* **8**, 427–433 (2018).

38. 38.

Auffhammer, M., Hsiang, S. M., Schlenker, W. & Sobel, A. Using weather data and climate model output in economic analyses of climate change. *Rev. Environ. Econ. Policy* **7**, 181–198 (2013).

Acknowledgements

We acknowledge funding from the Volkswagen Foundation and from the Horizon 2020 Framework Programme of the European Union (grant agreement number 820712). We thank M. Kalkuhl and S. Lange for discussions regarding economic and climate data, respectively.

Author information

Affiliations

1. Research Department of Complexity Science, Potsdam Institute for Climate Impact Research, Potsdam, Germany

Maximilian Kotz, Anders Levermann & Leonie Wenz

2. Institute of Physics, Potsdam University, Potsdam, Germany

Maximilian Kotz & Anders Levermann

3. Lamont-Doherty Earth Observatory, Columbia University, New York, NY, USA

Anders Levermann

4. Mercator Research Institute on Global Commons and Climate Change, Berlin, Germany

Leonie Wenz

Contributions

M.K. designed and conducted the analysis and contributed to the interpretation and presentation of the results. L.W. proposed the study, contributed to the design of the analysis and to the interpretation and presentation of the results. A.L. contributed to the interpretation and presentation of the results.

Corresponding author

Correspondence to [Leonie Wenz](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Xin-Zhong Liang, Chad W. Thackeray and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Historical means of the four principal rainfall measures.

Maps of the historical (1979–2019) means of each annual rainfall measure. **a**, The annual total rainfall. **b**, The monthly rainfall deviations (a weighted annual sum of anomalies of monthly rainfall from their climatological means which are, by definition, zero mean). **c**, The number of wet days. **d**, The extreme daily rainfall measure (the annual sum of rainfall on days exceeding the 99.9th percentile of the historical distribution).

Extended Data Fig. 2 Historical variability of the four principal rainfall measures.

Historical variability (the standard deviation of annual values over the years 1979–2019) for each measure of rainfall.

Extended Data Fig. 3 Rich and poor differentiated response of economic growth to changes in rainfall.

As Fig. 2 but having estimated economic responses to rainfall for rich and poor countries separately.

Extended Data Fig. 4 Response of sectoral growth to changes in rainfall.

As Fig. 2 but having estimated economic responses to rainfall for the agricultural (“ag”), manufacturing (“man”) and services (“serv”) sectors separately.

Extended Data Table 1 Results of the main econometric specification for the effect of temperature and rainfall changes on economic growth rates

Supplementary information

Supplementary Information

This Supplementary Information contains: Supplementary Sections 1–3, Figs. 1–6, Tables 1–17 and additional references.

Peer Review File

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Kotz, M., Levermann, A. & Wenz, L. The effect of rainfall changes on economic production. *Nature* **601**, 223–227 (2022).
<https://doi.org/10.1038/s41586-021-04283-8>

- Received: 18 June 2021
- Accepted: 25 November 2021
- Published: 12 January 2022

- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04283-8>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

[Extreme rainfall slows the global economy](#)

- Xin-Zhong Liang

News & Views 12 Jan 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04283-8>

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 12 January 2022](#)

Air pollution exposure disparities across US population and income groups

- [Abdulrahman Jbaily](#) ORCID: [orcid.org/0000-0001-9378-1819¹](https://orcid.org/0000-0001-9378-1819),
- [Xiaodan Zhou](#) ORCID: [orcid.org/0000-0002-9250-768X²](https://orcid.org/0000-0002-9250-768X),
- [Jie Liu²](#),
- [Ting-Hwan Lee²](#),
- [Leila Kamareddine³](#),
- [Stéphane Verguet¹](#) &
- [Francesca Dominici](#) ORCID: [orcid.org/0000-0002-9382-0141^{3,4}](https://orcid.org/0000-0002-9382-0141)

[Nature](#) volume 601, pages 228–233 (2022)

- 1622 Accesses
- 206 Altmetric
- [Metrics details](#)

Subjects

- [Environmental impact](#)

Abstract

Air pollution contributes to the global burden of disease, with ambient exposure to fine particulate matter of diameters smaller than 2.5 µm (PM_{2.5})

being identified as the fifth-ranking risk factor for mortality globally¹. Racial/ethnic minorities and lower-income groups in the USA are at a higher risk of death from exposure to PM_{2.5} than are other population/income groups^{2,3,4,5}. Moreover, disparities in exposure to air pollution among population and income groups are known to exist^{6,7,8,9,10,11,12,13,14,15,16,17}. Here we develop a data platform that links demographic data (from the US Census Bureau and American Community Survey) and PM_{2.5} data¹⁸ across the USA. We analyse the data at the tabulation area level of US zip codes (N is approximately 32,000) between 2000 and 2016. We show that areas with higher-than-average white and Native American populations have been consistently exposed to average PM_{2.5} levels that are lower than areas with higher-than-average Black, Asian and Hispanic or Latino populations. Moreover, areas with low-income populations have been consistently exposed to higher average PM_{2.5} levels than areas with high-income groups for the years 2004–2016. Furthermore, disparities in exposure relative to safety standards set by the US Environmental Protection Agency¹⁹ and the World Health Organization²⁰ have been increasing over time. Our findings suggest that more-targeted PM_{2.5} reductions are necessary to provide all people with a similar degree of protection from environmental hazards. Our study is observational and cannot provide insight into the drivers of the identified disparities.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

[Subscribe to Journal](#)

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

[Buy article](#)

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Fig. 1: Average PM_{2.5} concentration in 2000 and 2016 across ZCTAs in which Black or white populations are overrepresented.

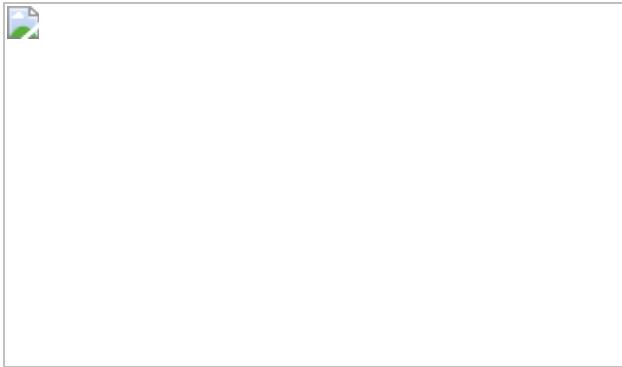


Fig. 2: Average PM_{2.5} concentration in 2000 and 2016 across low- and high-income ZCTAs.

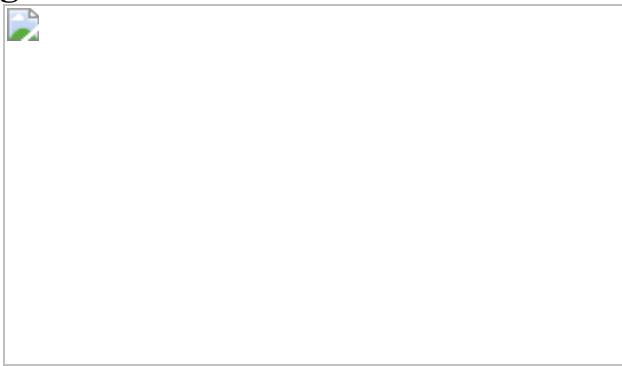


Fig. 3: US ZCTAs with average PM_{2.5} concentrations of more than 8 µg m⁻³ for Black and white populations in 2000 and 2016.

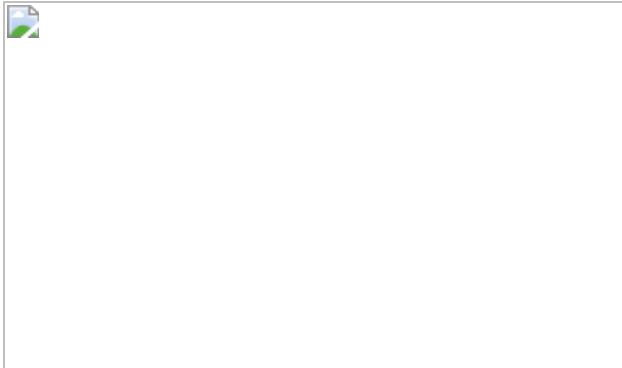
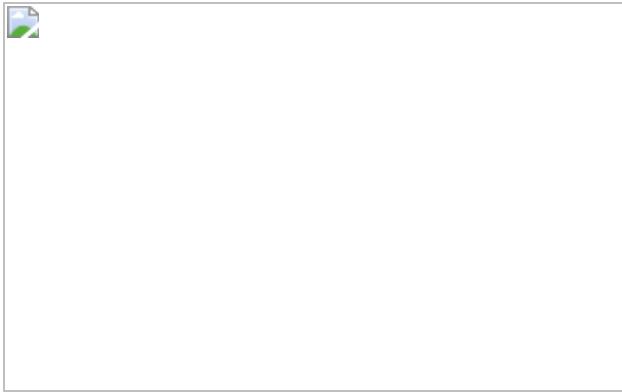


Fig. 4: Relative disparities in exposure to PM_{2.5} among racial/ethnic groups for 2000–2016.



Data availability

Data are available in the following GitHub repositories:

<https://github.com/NSAPH/National-Causal-Analysis/tree/master/Confounders/census> and https://github.com/xiaodan-zhou/pm25_and_disparity.

Code availability

Code is available in the following GitHub repository:

https://github.com/xiaodan-zhou/pm25_and_disparity.

References

1. 1.

Cohen, A. J. et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. *Lancet* **389**, 1907–1918 (2017).

2. 2.

Di, Q. et al. Air pollution and mortality in the Medicare population. *N. Engl. J. Med.* **376**, 2513–2522 (2017).

3. 3.

Bell, M. L., Zanobetti, A. & Dominici, F. Evidence on vulnerability and susceptibility to health risks associated with short-term exposure to particulate matter: a systematic review and meta-analysis. *Am. J. Epidemiol.* **178**, 865–876 (2013).

4. 4.

Wang, Y. et al. Long-term exposure to PM_{2.5} and mortality among older adults in the southeastern US. *Epidemiology* **28**, 207 (2017).

5. 5.

Kioumourtzoglou, M.-A., Schwartz, J., James, P., Dominici, F. & Zanobetti, A. PM_{2.5} and mortality in 207 US cities: modification by temperature and city characteristics. *Epidemiology* **27**, 221 (2016).

6. 6.

Bell, M. L. & Ebisu, K. Environmental inequality in exposures to airborne particulate matter components in the United States. *Environ. Health Perspect.* **120**, 1699–1704 (2012).

7. 7.

Rosofsky, A., Levy, J. I., Zanobetti, A., Janulewicz, P. & Fabian, M. P. Temporal trends in air pollution exposure inequality in Massachusetts. *Environ. Res.* **161**, 76–86 (2018).

8. 8.

Mikati, I., Benson, A. F., Luben, T. J., Sacks, J. D. & Richmond-Bryant, J. Disparities in distribution of particulate matter emission sources by race and poverty status. *Am. J. Public Health* **108**, 480–485 (2018).

9. 9.

Miranda, M. L., Edwards, S. E., Keating, M. H. & Paul, C. J. Making the environmental justice grade: the relative burden of air pollution

exposure in the United States. *Int. J. Environ. Res. Public Health* **8**, 1755–1771 (2011).

10. 10.

Mohai, P., Pellow, D. & Roberts, J. T. Environmental justice. *Annu. Rev. Environ. Resour.* **34**, 405–430 (2009).

11. 11.

Agyeman, J., Schlosberg, D., Craven, L. & Matthews, C. Trends and directions in environmental justice: from inequity to everyday life, community, and just sustainabilities. *Annu. Rev. Environ. Resour.* **41**, 321–340 (2016).

12. 12.

Banzhaf, S., Ma, L. & Timmins, C. Environmental justice: the economics of race, place, and pollution. *J. Econ. Perspect.* **33**, 185–208 (2019).

13. 13.

Kelly, J. T. et al. Examining PM_{2.5} concentrations and exposure using multiple models. *Environ. Res.* **196**, 110432 (2020).

14. 14.

Fann, N., Coffman, E., Timin, B. & Kelly, J. T. The estimated change in the level and distribution of PM_{2.5}-attributable health impacts in the United States: 2005–2014. *Environ. Res.* **167**, 506–514 (2018).

15. 15.

Tessum, C. W. et al. PM_{2.5} polluters disproportionately and systemically affect people of color in the United States. *Sci. Adv.* **7**, (2021).

16. 16.

Harper, S. et al. Using inequality measures to incorporate environmental justice into regulatory analyses. *Int. J. Environ. Res. Public Health* **10**, 4039–4059 (2013).

17. 17.

Colmer, J., Hardman, I., Shimshack, J. & Voorheis, J. Disparities in PM_{2.5} air pollution in the United States. *Science* **369**, 575–578 (2020).

18. 18.

Meng, J. et al. Estimated long-term (1981–2016) concentrations of ambient fine particulate matter across North America from chemical transport modeling, satellite remote sensing, and ground-based measurements. *Environ. Sci. Technol.* **53**, 5071–5079 (2019).

19. 19.

US Environmental Protection Agency. Process of reviewing the National Ambient Air Quality Standards. <https://www.epa.gov/criteria-air-pollutants/process-reviewing-national-ambient-air-quality-standards>

20. 20.

World Health Organization. *Air Quality Guidelines. Global Update 2005. Particulate Matter, Ozone, Nitrogen Dioxide, and Sulfur Dioxide* (World Health Organization, 2005).

21. 21.

Dominici, F. et al. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *J. Am. Med. Assoc.* **295**, 1127–1134 (2006).

22. 22.

Bell, M. L. et al. Seasonal and regional short-term effects of fine particles on hospital admissions in 202 US counties, 1999–2005. *Am. J. Epidemiol.* **168**, 1301–1310 (2008).

23. 23.

Kloog, I. et al. Short term effects of particle exposure on hospital admissions in the mid-atlantic states: a population estimate. *PLoS One* **9**, (2014).

24. 24.

Bravo, M. A. et al. Airborne fine particles and risk of hospital admissions for understudied populations: effects by urbanicity and short-term cumulative exposures in 708 US counties. *Environ. Health Perspect.* **125**, 594–601 (2017).

25. 25.

Dominici, F., McDermott, A., Zeger, S. L. & Samet, J. M. National maps of the effects of particulate matter on mortality: exploring geographical variation. *Environ. Health Perspect.* **111**, 39–44 (2003).

26. 26.

Beelen, R. et al. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre escape project. *Lancet* **383**, 785–795 (2014).

27. 27.

Crouse, D. L. et al. Ambient PM_{2.5}, O₃, and NO₂ exposures and associations with mortality over 16 years of follow-up in the Canadian census health and environment cohort (CanCHEC). *Environ. Health Perspect.* **123**, 1180–1186 (2015).

28. 28.

Makar, M. et al. Estimating the causal effect of fine particulate matter levels on death and hospitalization: are levels below the safety standards harmful? *Epidemiology* **28**, 627 (2017).

29. 29.

Di, Q. et al. Association of short-term exposure to air pollution with mortality in older adults. *J. Am. Med. Assoc.* **318**, 2446–2456 (2017).

30. 30.

Wu, X., Braun, D., Schwartz, J., Kioumourtzoglou, M. & Dominici, F. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. *Sci. Adv.* **6**, eaba5692 (2020).

31. 31.

Liu, C. et al. Ambient particulate air pollution and daily mortality in 652 cities. *N. Engl. J. Med.* **381**, 705–715 (2019).

32. 32.

Samoli, E. et al. Estimating the exposure–response relationships between particulate matter and mortality within the aphea multicity project. *Environ. Health Perspect.* **113**, 88–95 (2005).

33. 33.

Samet, J. M., Dominici, F., Curriero, F. C., Coursac, I. & Zeger, S. L. Fine particulate air pollution and mortality in 20 us cities, 1987–1994. *N. Engl. J. Med.* **343**, 1742–1749 (2000).

34. 34.

Shah, A. S. et al. Global association of air pollution and heart failure: a systematic review and meta-analysis. *Lancet* **382**, 1039–1048 (2013).

35. 35.

American Psychological Association. Racial and ethnic identity.
<https://apastyle.apa.org/style-grammar-guidelines/bias-free-language/racial-ethnic-minorities>

36. 36.

EPA proposes to retain NAAQS for particulate matter.
<https://www.epa.gov/newsreleases/epa-proposes-retain-naaqs-particulate-matter>

37. 37.

Clark, L. P., Millet, D. B. & Marshall, J. D. National patterns in environmental injustice and inequality: outdoor NO₂ air pollution in the United States. *PLoS One* **9**, e94431 (2014).

38. 38.

Levy, J. I., Chemerynski, S. M. & Tuchmann, J. L. Incorporating concepts of inequality and inequity into health benefits analysis. *Int. J. Equity Health* **5**, 2 (2006).

39. 39.

Lambert, P. J., Millimet, D. L. & Slottje, D. Inequality aversion and the natural rate of subjective inequality. *J. Public Econ.* **87**, 1061–1090 (2003).

40. 40.

Currie, J., Voorheis, J. & Walker, R. *What Caused Racial Disparities in Particulate Exposure to Fall? New Evidence From the Clean Air Act and Satellite-Based Measures of Air Quality*. Technical report (National Bureau of Economic Research, 2020).

41. 41.

Auffhammer, M., Bento, A. M. & Lowe, S. E. Measuring the effects of the clean air act amendments on ambient PM₁₀ concentrations: the critical importance of a spatially disaggregated analysis. *J. Environ. Econ. Manage.* **58**, 15–26 (2009).

42. 42.

Grainger, C. A. The distributional effects of pollution regulations: do renters fully pay for cleaner air? *J. Public Econ.* **96**, 840–852 (2012).

43. 43.

Di, Q. et al. Assessing PM_{2.5} exposures with high spatiotemporal resolution across the continental United States. *Environ. Sci. Technol.* **50**, 4712–4721 (2016).

44. 44.

Di, Q. et al. An ensemble-based model of PM_{2.5} concentration across the contiguous United States with high spatiotemporal resolution. *Environ. Int.* **130**, 104909 (2019).

Acknowledgements

We thank R. Martin and J. D. Schwartz for providing the air-pollution data; B. Sabbath for cleaning and preparing the data sets; and L. Bennett for comments and discussions. We also thank J. Kodros for his comments on an earlier draft. This work was supported financially by grants from the Health Effects Institute (4953- RFA14-3/16-4), the National Institutes of Health (DP2MD012722, P50MD010428), the National Institutes of Health and Yale University (R01MD012769), the National Institutes of Health and National Institute of Environmental Health Sciences (R01 ES028033, R01ES026217, R01AG066793-01, R01ES029950, R01ES028033-S1), the National Institutes of Health and Columbia University (1R01ES030616), the Environmental Protection Agency (83587201-0), The Climate Change Solutions Fund, and a Harvard Star Friedman Award.

Author information

Affiliations

1. Department of Global Health and Population, Harvard T. H. Chan School of Public Health, Boston, MA, USA

Abdulrahman Jbaily & Stéphane Verguet

2. Environmental Systems Research Institute, Redlands, CA, USA

Xiaodan Zhou, Jie Liu & Ting-Hwan Lee

3. Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA

Leila Kamareddine & Francesca Dominici

4. Harvard Data Science Initiative, Cambridge, MA, USA

Francesca Dominici

Contributions

A.J., S.V. and F.D. contributed to the study design. A.J. led the research, with support from X.Z. and supervision from F.D. Maps and videos were prepared by X.Z., J.L. and T.-H.L. A.J. drafted the manuscript, with support from L.K., S.V. and F.D. All authors read and approved the final manuscript for submission.

Corresponding authors

Correspondence to [Abdulrahman Jbaily](#) or [Francesca Dominici](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Corbett Grainger, Jonathan Levy, Arden Pope III and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

[Extended Data Fig. 1 Summary PM_{2.5} metrics across racial/ethnic and income groups.](#)

a, The population-weighted average of PM_{2.5} decreased by 40.4% from the year 2000 to 2016. **b**, Population-weighted average PM_{2.5} concentration across the different racial/ethnic communities for 2000 to 2016, showing that Black and Native American populations live in the most- and least-polluted areas, respectively. **c**, Population-weighted average PM_{2.5} concentration across racial/ethnic communities as a function of ZCTA racial/ethnic population (%) for 2016. For example, when the racial/ethnic population percentage is equal to 0.2, the red curve includes every ZCTA where the Black population is 20% or more, and the blue curve includes every ZCTA where the white population is 20% or more. As a ZCTA's Black and Hispanic or Latino populations increase, the PM_{2.5} concentration levels increase. The opposite effect is seen for the white and Native American communities. **d**, The population-weighted average PM_{2.5} concentration across the income groups reveals that the low-income group

has been exposed to only slightly higher PM_{2.5} levels than the high-income groups since 2004. **e**, Population-weighted average PM_{2.5} concentrations across the different racial/ethnic communities that are in the low-income group, for 2000–2016. **f**, Population-weighted average PM_{2.5} concentrations across the different racial/ethnic communities that are in the high-income group, for 2000–2016. Panels **e**, **f** show similar differences in average PM_{2.5} concentrations across the racial/ethnic groups as seen in **b**.

Extended Data Fig. 2 Average PM_{2.5} concentrations across the US.

a, Distribution of PM_{2.5} in 2000. **b**, Distribution of PM_{2.5} in 2016. Supplementary Video [1](#) shows the change in the distribution of PM_{2.5} concentration levels in the US from 2000 to 2016. Note that Hawaii and Alaska are not shown. Imagery provided courtesy of Esri, HERE, Garmin, FAO, NOAA, USGS, ©OpenStreetMap contributors, and the GIS User Community.

Extended Data Fig. 3 Average PM_{2.5} concentrations across ZCTAs in which different racial/ethnic groups are overrepresented.

a, Distribution of PM_{2.5} across five different maps for 2000, each showing the ZCTAs in which one race/ethnicity group is overrepresented. **b**, Distribution of PM_{2.5} across five different maps for 2016, each showing the ZCTAs in which one race/ethnicity group is overrepresented. Supplementary Videos [2](#), [3](#) show the change in the distribution of PM_{2.5} concentrations across the five maps from 2000 to 2016. Note that Hawaii and Alaska are not shown. Imagery provided courtesy of Esri, HERE, Garmin, FAO, NOAA, USGS, ©OpenStreetMap contributors, and the GIS User Community.

Extended Data Fig. 4 Distribution of racial/ethnic populations above a PM_{2.5} threshold of 8 µg m⁻³ for 2000 and 2016.

a, US ZCTAs for each race/ethnicity are ranked on the basis of the ratio of the race/ethnicity population to the total ZCTA population. Dark blue indicates fractions close to 1 (ZCTAs in which the corresponding race/ethnicity most lives), and light yellow indicates fractions close to 0 (ZCTAs in which the corresponding race/ethnicity least lives). **b**, US ZCTAs with PM_{2.5} concentrations higher than 8 µg m⁻³ in 2000. **c**, US ZCTAs with PM_{2.5} concentrations higher than 8 µg m⁻³ in 2016. Supplementary Videos 5–8 show the distribution of the different racial/ethnic groups across multiple ranges of PM_{2.5} concentrations for 2000 and 2016. Note that Hawaii and Alaska are not shown. Imagery provided courtesy of Esri, HERE, Garmin, FAO, NOAA, USGS, ©OpenStreetMap contributors, and the GIS User Community.

Extended Data Fig. 5 Supplementary measures of relative disparities in exposure to PM_{2.5} among racial/ethnic groups for 2000–2016.

a, The Atkinson index is computed to measure relative disparities among the racial/ethnic groups (Black, white, Asian, Native American and Hispanic or Latino). **b**, The Gini index is computed to measure relative disparities among the racial/ethnic groups (Black, white, Asian, Native American and Hispanic or Latino). The trends in both indices are similar to that measured by CoV (Fig. 4): racial/ethnic disparities in exposure to air pollution relative to pollution levels at or below the EPA standard are increasing. The Atkinson and Gini indices were computed using the inequality package ‘ineq’ in R software. The input is the proportion of the racial/ethnic (or income) groups living above the set PM_{2.5} threshold. We set the Atkinson aversion parameter, ε , to 0.75 (ref. 7); the sensitivity of the index to different values of ε is shown in Extended Data Fig. 6.

Extended Data Fig. 6 Sensitivity of the Atkinson index to the inequality aversion parameter ε .

a, Sensitivity of the Atkinson index relative to a PM_{2.5} threshold of 8 $\mu\text{g m}^{-3}$. **b**, Sensitivity of the Atkinson index relative to a PM_{2.5} threshold of 10 $\mu\text{g m}^{-3}$. **c**, Sensitivity of the Atkinson index relative to a PM_{2.5} threshold of 12 $\mu\text{g m}^{-3}$. A consistent trend is shown across the parameter values.

Extended Data Fig. 7 Replication of the main findings across urban and rural areas.

A ZCTA's population density is used as a metric to control for urbanicity in our study. We classify urban and rural areas on the basis of the percentage of the urban population in each ZCTA; such percentages are available from the US Census Bureau for 2010. ZCTAs with an urban population of more than 50% are classified as urban, whereas those with an urban population of less than 50% are classified as rural. For nationwide, urban and rural US, we reproduce our main results: namely, the average PM_{2.5} concentrations for the total population (**a–c**), for racial/ethnic groups (**d–f**) and for income groups (**g–i**), as well as disparities among racial/ethnic groups (**j–l**). Similarities in the results across the national, urban and rural US are apparent and findings are consistent regardless of the urbanicity of ZCTAs. Note that in the case of the rural US, we only compute disparities (**l**) for the years in which the proportion of the population exposed to PM_{2.5} concentrations above the thresholds of interest is non-zero. For example, the proportion of the population in the rural US that is exposed to PM_{2.5} concentrations above $T = 12 \mu\text{g m}^{-3}$ reaches near-zero levels in 2009, and hence disparities after this year are not computed.

Extended Data Fig. 8 Sensitivity of our main findings to estimates of PM_{2.5}.

We replicated our analysis using an independent pollution data set^{43,44}, and we show here the sensitivity of our findings to the new PM_{2.5} estimates. **a**, Replication of Extended Data Fig. 1b using the alternative data set. **b**, Replication of Extended Data Fig. 1d using the alternative data set. **c**, Replication of Fig. 4 using the alternative data set. Our main findings are robust and consistent across the two data sets. (Minor differences resulting from the different pollution estimates can be spotted, as expected.).

Supplementary information

Peer Review File

Supplementary Video 1

Average PM_{2.5} concentration levels across the US by ZCTA and by year from 2000 to 2016. The colour ramps from green to red represent PM_{2.5} levels of 0–7, 7–8, 8–9, 9–10, 10–11, 11–12, 12–30 µg m⁻³. As the animation moves forward, we sequentially see the PM_{2.5} levels from 2000, 2001, 2002, up to 2016. Note that Hawaii and Alaska are not shown. Imagery provided courtesy of Esri, HERE, Garmin, FAO, NOAA, USGS, ©OpenStreetMap contributors, and the GIS User Community.

Supplementary Video 2

Average PM_{2.5} concentration levels for the ZCTAs where the racial/ethnic communities are overrepresented for the years 2000 to 2016. The colour ramps from green to red represent PM_{2.5} levels of 0–7, 7–8, 8–9, 9–10, 10–11, 11–12, 12–30 µg m⁻³. At the top-left, we highlight ZCTAs where the Black population fraction is higher than 7%. At the top-right we highlight ZCTAs where the white population fraction is higher than 84%. At the bottom-left we highlight ZCTAs where the Hispanic/Latino population fraction is higher than 9%. At the bottom-right we highlight ZCTAs where the Asian population fraction is higher than 2%. As the animation moves forward, we sequentially see the PM_{2.5} levels from 2000, 2001, 2002, up to 2016. Note that Hawaii and Alaska are not shown. Imagery provided

courtesy of Esri, HERE, Garmin, FAO, NOAA, USGS, ©OpenStreetMap contributors, and the GIS User Community.

Supplementary Video 3

An extension of Supplementary Video 2 to the Native American population. We show ZCTAs where the Native American population fraction is higher than 1%. Note that Hawaii and Alaska are not shown. Imagery provided courtesy of Esri, HERE, Garmin, FAO, NOAA, USGS, ©OpenStreetMap contributors, and the GIS User Community.

Supplementary Video 4

Average PM_{2.5} concentration levels across low- and high- income ZCTAs for the years 2000 to 2016. The colour ramps from green to red represent PM_{2.5} levels of 0–7, 7–8, 8–9, 9–10, 10–11, 11–12, 12–30 µg m⁻³. On the left, we highlight low-income ZCTAs where the median household income is at the bottom 30%. On the right, we highlight high-income ZCTAs where the median household income is at the top 30%. As the animation moves forward, we sequentially see the PM_{2.5} levels from 2000, 2001, 2002, up to 2016. Note that Hawaii and Alaska are not shown. Imagery provided courtesy of Esri, HERE, Garmin, FAO, NOAA, USGS, ©OpenStreetMap contributors, and the GIS User Community.

Supplementary Video 5

Distribution of the racial/ethnic communities across levels of PM_{2.5} concentrations in 2000. The continuous colour ramps from light yellow to dark blue and represents the quantile of the percentage of racial/ethnic communities across ZCTAs from low to high. As the animation moves forward, we sequentially see which racial/ethnic communities are exposed to PM_{2.5} levels above 0, 7, 8, 9, 10, 11, 12 µg m⁻³. Note that Hawaii and Alaska are not shown. Imagery provided courtesy of Esri, HERE, Garmin, FAO, NOAA, USGS, ©OpenStreetMap contributors, and the GIS User Community.

Supplementary Video 6

An extension of Supplementary Video 5 to the Native American population. Note that Hawaii and Alaska are not shown. Imagery provided courtesy of Esri, HERE, Garmin, FAO, NOAA, USGS, ©OpenStreetMap contributors, and the GIS User Community.

Supplementary Video 7

Distribution of the racial/ethnic communities across levels of PM_{2.5} concentrations in 2016. The continuous colour ramps from light yellow to dark blue and represents the quantile of the percentage of racial/ethnic communities across ZCTAs from low to high. As the animation moves forward, we sequentially see which racial/ethnic communities are exposed to PM_{2.5} levels above 0, 7, 8, 9, 10, 11, 12 µg m⁻³. Note that Hawaii and Alaska are not shown. Imagery provided courtesy of Esri, HERE, Garmin, FAO, NOAA, USGS, ©OpenStreetMap contributors, and the GIS User Community.

Supplementary Video 8

An extension of Supplementary Video 7 to the Native American population. Note that Hawaii and Alaska are not shown. Imagery provided courtesy of Esri, HERE, Garmin, FAO, NOAA, USGS, ©OpenStreetMap contributors, and the GIS User Community.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Jbaily, A., Zhou, X., Liu, J. *et al.* Air pollution exposure disparities across US population and income groups. *Nature* **601**, 228–233 (2022).
<https://doi.org/10.1038/s41586-021-04190-y>

- Received: 13 July 2020
- Accepted: 26 October 2021
- Published: 12 January 2022
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04190-y>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04190-y>

- Article
- Open Access
- [Published: 20 December 2021](#)

Ostrich eggshell beads reveal 50,000-year-old social network in Africa

- [Jennifer M. Miller](#) [ORCID: orcid.org/0000-0002-3444-8650^{1,2}](#) &
- [Yiming V. Wang](#) [ORCID: orcid.org/0000-0003-3228-5592¹](#)

[Nature](#) volume 601, pages 234–239 (2022)

- 12k Accesses
- 1 Citations
- 669 Altmetric
- [Metrics details](#)

Subjects

- [Archaeology](#)
- [Palaeoclimate](#)

Abstract

Humans evolved in a patchwork of semi-connected populations across Africa^{1,2}; understanding when and how these groups connected is critical to interpreting our present-day biological and cultural diversity. Genetic analyses reveal that eastern and southern African lineages diverged sometime in the Pleistocene epoch, approximately 350–70 thousand years ago (ka)^{3,4}; however, little is known about the exact timing of these interactions, the cultural context of these exchanges or the mechanisms that drove their separation. Here we compare ostrich eggshell bead variations between eastern and southern Africa to explore population dynamics over the past 50,000 years. We found that ostrich eggshell bead technology probably originated in eastern Africa and spread southward approximately 50–33 ka via a regional network. This connection breaks down approximately 33 ka, with populations remaining

isolated until herders entered southern Africa after 2 ka. The timing of this disconnection broadly corresponds with the southward shift of the Intertropical Convergence Zone, which caused periodic flooding of the Zambezi River catchment (an area that connects eastern and southern Africa). This suggests that climate exerted some influence in shaping human social contact. Our study implies a later regional divergence than predicted by genetic analyses, identifies an approximately 3,000-kilometre stylistic connection and offers important new insights into the social dimension of ancient interactions.

[Download PDF](#)

Main

Unresolved questions in human evolution concern the ancient distribution and diversification of our species (*Homo sapiens*) across Africa^{2,5}. The metapopulation model suggests that anatomical modernity and behavioural complexity arose within a pan-African patchwork of populations who experienced pulses of connection and isolation⁶, possibly in response to environmental circumstances^{1,7}. Research into these shifting connections is increasingly derived from DNA and ancient DNA analyses, which reveal that present-day African hunter-gatherer populations diverged into regional lineages sometime in the Pleistocene, including a deep division between southern and eastern groups approximately 350–70 ka^{3,4,8}. Although ancient DNA is a powerful tool for acquiring information about biological exchange, it is unable to address the cultural context of ancient interactions. Many questions about these ancient interactions remain, such as where and when did ancient populations connect, what social exchanges took place and what mechanisms provoked their eventual isolation.

Beginning in Marine Isotope Stage 3 (approximately 57 ka), African populations underwent substantial social reorganization^{9,10,11}. Numerous advancements appear around this time, but an important new feature is the manufacture of beads¹² (Supplementary Discussion 1). The systematic production of beads is a considerable labour investment, and signals the increasing scale and importance of social interactions in Marine Isotope Stage 3 (ref. ¹³), perhaps relating to the growing population size and social systems evident around this time¹¹. These societal reforms signal that the African Late Pleistocene is a crucial period for understanding the development of complex social networks.

Ostrich eggshell (OES) beads are the oldest fully manufactured beads and could be key to revealing Late Pleistocene social dynamics in Africa. They emerged in eastern Africa by 52 ka¹², in southern Africa by 42 ka¹⁴ and are still produced in some areas

today. Modern ethnographic research in Africa indicates that a finished piece of OES beadwork (for example, a beaded skirt) carries symbolic meaning¹⁵. However, individual beads can also preserve social information, as every step in their production is a deliberate choice that intensifies morphological differences¹⁶ (Supplementary Discussion 2). These manufacturing decisions are cultural norms that are commonly shared between neighbouring groups, while long distances reduce transmission opportunities leading to cultural variation or drift^{17,18,19,20}. Therefore, the characteristics of OES beads can be used as a means to reconstruct population interaction. Previous studies linked the introduction of herding into southern Africa (approximately 2 ka) with the appearance of larger-diameter OES beads^{21,22}, indicating possible connections with eastern African populations, as supported by archaeological and genetic evidence^{4,21,23}. Some recent studies have reported stylistic variation within Late Pleistocene sites^{24,25,26,27}; however, to our knowledge, there has been no attempt to use similar variation to explore population contact in the Pleistocene.

Episodes of population connection and isolation have been linked with environmental shifts^{1,2}, and over the past 50,000 years (kyr), climatic events have triggered temperature fluctuations and hydroclimatic reorganization in Africa^{28,29,30}. These shifts could have fragmented habitable areas, in turn affecting where and when regional populations could interact. Therefore, it is critical to explore how intergroup connectivity may correspond with climatic and environmental changes in the Late Pleistocene.

In this study, we analysed OES bead characteristics from the past 50 kyr in search of patterns that reveal population connections, and their association with hydroclimate shifts in Africa. We compiled data from 31 sites in eastern (22.5–40° E, 9° N to 9° S) and southern Africa (8–35° E, 20–35° S), totalling 1,516 individual beads (Fig. 1, Supplementary Table 1), with 1,238 of these being fully reported for the first time. We recorded three metric variables wherever possible (bead diameter, aperture diameter and shell thickness). Our database comprises securely dated Pleistocene sites with available data, and well-dated sequences in each region, with age estimates drawn from direct radiocarbon dates, dated archaeological layers or bracketing layers. To understand the potential effects of climate on these patterns, we divided the past 50–2 kyr into four periods based on major glacial and interglacial shifts (Supplementary Discussion 3): phase I: 50–33 ka (Marine Isotope Stage 3 to the reinvigoration of ice-sheet growth); phase II: 33–19 ka (ice-sheet growth to the end of Last Glacial Maximum); phase III: 19–11.6 ka (last deglaciation); and phase IV: 11.6–2 ka (Early Holocene epoch to before the spread of herding into southern Africa). Phase V (2 ka to present) marks the previously identified shift in bead sizes that emerges as herding spreads into southern Africa. We expect to see population connections indicated by similar bead characteristics, and that periods of isolation may parallel climatic shifts.

Fig. 1: Locations of sites included in this study and palaeoclimate records.

 figure1

Base map modified from Natural Earth. **a**, Kakapel Rockshelter (1); Enkapune ya Muto (2); Mumba Rockshelter (3); Panga ya Saidi (4); Daumboy 3 Rockshelter (5); Kisese II Rockshelter (6); Mlambalasi Rockshelter (7); Magubike Rockshelter (8); White Paintings Shelter (9); Geduld (10); Lower Numas Cave (11); Lower Orabes Shelter (12); Leopard Cave (13); Eros (14); Wortel (15); Bushman Rockshelter (16); Border Cave (17); Apollo 11 Cave (18); Wonderwerk Cave (19); Dikbosch 1 Shelter (20); Sehonghong (21); SK2001.026 (22); Rooiwal Hollow/Midden (23); Varsche Rivier 003 (24); Paternoster (25); Grassridge Shelter (26); Witklip (27); Kasteelberg A + B (28); Geelbek Dunes (29); Voelvlei (30); Nelson Bay Cave (31). **b**, Representative OES beads from sites in eastern Africa. **c**, Representative OES beads from sites in southern Africa.

Regional and chronological bead metrics

Our results reveal that eastern and southern African OES beads take unique stylistic trajectories through time (Fig. 2a). Phases and regions are both important factors driving the variation in OES bead characteristics (Pillai's trace = 0.60, $F_{3,1319} = 664.8$, $P < 0.001$ for region and Pillai's trace = 0.18, $F_{12,3963} = 21.34$, $P < 0.001$ for phase), although interaction between phases and regions do not appear to significantly influence OES bead characteristics (Pillai's trace = 0.02, $F_{9,3963} = 2.22$, $P = 0.02$; Supplementary Table 2).

Fig. 2: OES bead diameter, thickness and aperture diameter distribution through the past 50 kyr in eastern and southern Africa.

 **figure2**

a, Generalized additive model plots to show bead characteristic evolutionary trajectories. The mean (curved lines) and 95% confidence interval are shown for each parameter. **b**, Split violin plots of bead parameters. The violins represent the kernel density of the frequency distribution, and the points are presented as mean values \pm one standard error. Statistical results are shown in Supplementary Tables 5–7, 11. The asterisks denote the significance of region differences (* $P < 0.01$ and *** $P < 0.001$). When no significance between two regions was found, P values are presented.

In eastern Africa, the range of bead and aperture diameters remain consistent over 50 kyr, with only minor fluctuations. Eastern beads average 6.9 ± 1.2 mm in diameter and 2.6 ± 0.6 mm in aperture diameter (Fig. 2a), with a wide range of variation. By contrast, southern bead characteristics have changed through time, with larger bead and aperture diameters in phase I (50–33 ka) and significantly smaller characteristics in the younger phases (Pillai's trace = 0.113, $F_{9,3147} = 13.4$, $P < 0.001$; Fig. 2a, Supplementary Table 3). While southern beads virtually disappear from the archaeological record in phase II (33–19 ka), they re-emerged around the onset of deglaciation (approximately 19 ka) with consistently smaller sizes. From phases III–V (19 ka to present), southern bead diameters and aperture diameters are smaller with narrower ranges (4.5 ± 0.9 mm and 1.8 ± 0.4 mm, respectively) than their eastern counterparts. They remained in this consistently smaller style until after 2 ka when larger bead characteristics, associated with the movement of pastoral communities, appeared in southern Africa (multivariate analysis of variance (MANOVA) Pillai's trace = 0.004, $F_{3,700} = 1.05$, $P = 0.371$; Supplementary Table 4) (Figs. 2b, 3).

Fig. 3: Comparison of bead characteristics between eastern and southern Africa during the past 50 kyr.

 figure3



a, Principal component analyses (PCA) of OES bead metric parameters for phases III–V. Diameter, aperture diameter and thickness account for more than 89% of the variation, separating eastern and southern Africa into distinct groups. PCA clustering for two regions for phase II and phase I is not possible due to insufficient data. **b**, Paired diameter and aperture diameter for each phase. Newly reported data: collected by authors, reported as individual beads; published data: drawn from published metrics, reported as individual beads; average data: drawn from published metrics, reported as averaged values.

We found distinct regional clusters with varying degrees of overlap throughout phases III–V (19 ka to present) using principal component analysis for specimens with all three metric parameters ($n = 1,333$) (Fig. 3a). PC1 and PC2 explain 92%, 91% and 93% of variations between southern Africa and eastern Africa for phase III (19–11.6 ka), phase IV (11.6–2 ka) and phase V (2 ka to present), respectively (Fig. 3a).

The univariate analysis of variance (ANOVA) performed on the MANOVA outputs showed that all three parameters have a role in driving the regional differences in phases III–V (ANOVA $P < 0.001$ for all tests; Supplementary Tables 5–7). We further explored these regional differences using the two most commonly reported variables (bead diameter and aperture diameter), which slightly increased sample size to 1,445 beads (Fig. 3b). Our MANOVA results using only these two variables confirmed that bead characteristics are significantly different between the two regions during phases III–V (19 ka to present) (Fig. 3b, Supplementary Tables 8–10). Compared with the more distinct regional bead clusters in phase III (19–11.6 ka) and phase IV (11.6–2 ka), the beads in phase V (2 ka to present) show increased overlap between eastern and southern Africa. Despite this overlap, most southern beads in phase V (2 ka to present) remain smaller, consistent with phase III (19–11.6 ka) and phase IV (11.6–2 ka) (Figs. 2b, 3b).

Bead characteristics in phase I are nearly identical for eastern and southern Africa (Pillai's trace = 0.15, $F_{2,36} = 3.2$, $P = 0.052$; Figs. 2b, 3b, Supplementary Table 11), with similarities driven by bead diameter and aperture diameter (ANOVA $P = 0.08$ and 0.02, respectively; Supplementary Table 11). The average OES bead diameters in southern Africa are larger in phase I (6.7 mm) than those in other time periods by more than 2 mm, making them more similar to sizes in eastern Africa (average diameters of more than 6.9 mm) (Fig. 3b). The majority of southern beads (12 out of 14) derive from a single site—Border Cave—which has a wide range of diameters (4.3–8.1 mm). The remaining beads are one each from VR003 and White Paintings Shelter. Both sites are located significantly further west, but each bead is 5.7 mm in diameter, which falls within the range of diameters from Border Cave.

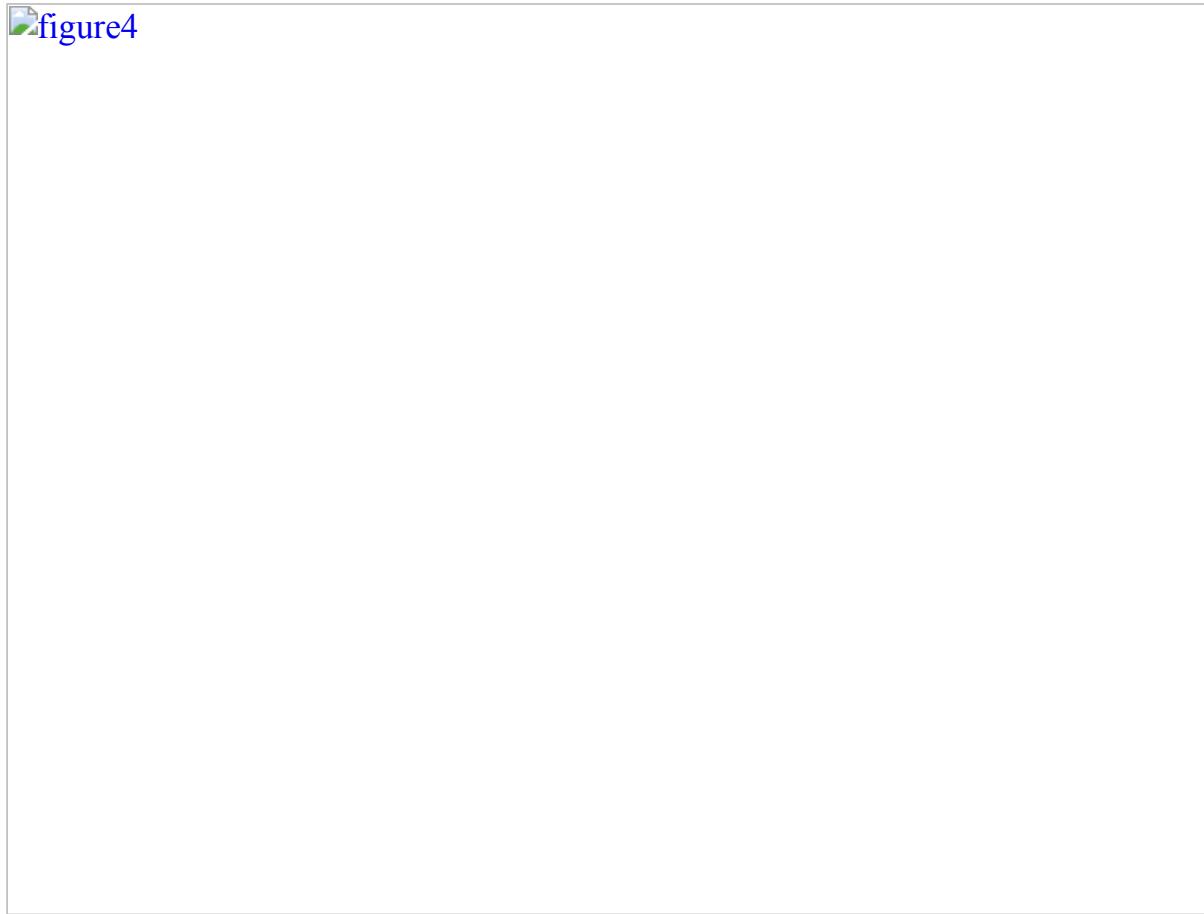
Shell thickness is not a stylistic trait, but instead may reflect a complex relationship between environment and ostrich. Both regions maintain consistent shell thickness over the entire 50 kyr period, with eastern African shells averaging 1.7 ± 0.2 mm, and southern shells averaging 1.5 ± 0.2 mm (Fig. 2b). This appears to contradict previous suggestions that shell thickness varies in response to temperature and aridity³¹. While thickness does not vary within each region through time, it is significantly different between the two regions ($P < 0.004$ for phases III–V; Supplementary Tables 5–7), and may represent different ostrich sub-species^{32,33}. The thinner southern African shell may have encouraged the production of smaller beads, and future studies should test this hypothesis, although this would not account for larger beads in southern African phase I (50–33 ka) and phase V (2 ka to present).

Discussion

Stylistic connection at 50–33 ka

Despite the substantial distance (more than 3,000 km) between eastern and southern Africa, the available OES beads from phase I (50–33 ka) share stylistic similarities. This is the oldest (and only) time period that the two regions have the same bead diameter range, strongly suggesting some form of socially mediated exchange during this time, marking the furthest Pleistocene stylistic connection ever documented. On the basis of age, site locations and bead characteristics, OES bead technology appears to have originated in eastern Africa. The oldest directly dated eastern beads are approximately 10 kyr older than those from southern Africa^{[12,14,34](#)}. Most southern beads in this phase come from Border Cave, which is located towards eastern Africa (Fig. [1a](#)); however, none of the three phase I sites from southern Africa have signs of in situ bead production. This apparent spread of beadmaking technology, evident mainly from the traits at Border Cave, corresponds with the relatively wet climatic conditions in eastern Africa during phase I (50–33 ka) (Fig. [4](#)).

Fig. 4: Bead-derived social connection and comparison climate proxies from eastern and southern Africa.



a, Annual mean rainfall map of Africa (map modified from Wikimedia Commons under the CC BY-SA 4.0 licence), showing the position of palaeoclimate data (yellow squares) used in **c**: Lake Tanganyika^{[29](#)}; Lake Malawi^{[46](#)}; Core GIK16160-3 (ref. [30](#));

Core 64PE304-80 (ref. ⁴³); Zambezi River catchment area (red) and major tributaries (white). The two black boxes indicate the modelled climate data area for eastern Africa ($22.5\text{--}40^\circ \text{E}$, $-9\text{ to }9^\circ \text{N}$) and southern Africa ($8\text{--}35^\circ \text{E}$, $20\text{--}35^\circ \text{S}$) from the LOVECLIM transient climate model³⁵. The black dashed lines indicate the location of present-day Intertropical Convergence Zone (ITCZ) in January and July. **b**, Latitudinal north to south sequence of climate data series (from top to bottom): total precipitation for eastern Africa ($22.5\text{--}40^\circ \text{E}$, $-9\text{ to }9^\circ \text{N}$) derived from the LOVECLIM transient climate model³⁵; moisture availability derived from the δD of leaf wax record from Lake Tanganyika²⁹ is overlaid with the June–July–August (JJA) mean insolation at $0\text{--}20^\circ \text{N}$ (red dashed line); lake level is derived from PCA analyses (PC1) in Lake Malawi; moisture availability derived from the δD of leaf wax record (ice volume corrected) from the Zambezi River catchment area⁴⁷ overlaid with the December–January–February (DJF) mean insolation at $0\text{--}20^\circ \text{S}$; the neodymium (Nd) isotope signature of lithogenic fractions in marine sediments from the Mozambique Shelf margin⁴³; the total precipitation for southern Africa ($8\text{--}35^\circ \text{E}$, $20\text{--}35^\circ \text{S}$) derived from the LOVECLIM transient climate model³⁵. HE3, Heinrich Event III; LGM, Last Glacial Maximum; VSMOW, Vienna Standard Mean Ocean Water. **c**, Social connections between eastern and southern Africa, derived through bead styles, over the past 50 kyr.

Disconnection and climatic links

The regional network seems to break down sometime in phase II (33–19 ka), raising questions about the influence of climate on social connections (Fig. ^{4a}). By 33 ka, precipitation in eastern Africa decreased (Fig. ^{4b}), modulated by the Indian Winter Monsoon and decreasing sea surface temperature of the Indian Ocean²⁹. These drier conditions persisted until approximately 16 ka^{29,35}, resulting in the lowest net primary production values over the past 50 kyr, according to climate model simulations (Supplementary Figs. ^{2, 3}). This reduction of net primary production would have altered the distribution of vegetation and fauna on the landscape⁷ (Supplementary Fig. ⁴), requiring humans to adjust mobility and foraging strategies^{36,37,38}. This in turn could have reorganized the distribution of people on the landscape, depopulating areas, and rendering some previous social networks unsustainable^{1,39,40}. The breakdown between phase I and phase II also coincides with the lowest effective population size in Africa predicted by ref. ⁴¹, and may suggest that shrinking population sizes contributed to regional disconnection. By contrast, the Zambezi River catchment (the large region connecting eastern and southern Africa) became wetter from 30 to 16 ka, according to climate proxy data (Fig. ^{4b}). This enhanced rainfall was due to the southward migration of the Intertropical Convergence Zone to $10\text{--}20^\circ \text{S}$, largely controlled by Heinrich Event III in the North Atlantic where massive iceberg melting occurred approximately 30 ka⁴² (Supplementary Discussion ⁴). The increased

precipitation resulted in periodic flooding of Zambezi River and its tributaries⁴³, which could have formed a geographical barrier to connections between eastern and southern Africa (Supplementary Discussion 4, Fig. 4b). The drying trend in eastern Africa and the flooded Zambezi River catchment may have instigated the regional disconnection that appeared by phase II (33–19 ka), suggesting that climate induced behavioural responses could be an important mechanism for driving cultural isolation in the Late Pleistocene (Fig. 4c).

Southern OES beads became rare, even seeming to disappear by 33 ka, and did not re-emerge until after 19 ka (Supplementary Discussion 5). Their absence coincides with the lowest net primary production and the coolest glacial temperatures in southern Africa, which may have limited the population size in the Late Pleistocene. If social group sizes are small, the mass production of standardized beads can be more costly than beneficial (Supplementary Discussion 1). This could explain why OES beadmaking did not become part of the cultural repertoire, even after the technology was introduced in phase I (50–33 ka). When southern beads do re-emerge (approximately 19 ka), they are in an exclusively smaller style. This regionalization of styles reflects a prolonged period of social isolation, and corresponds with a gradual increase in precipitation and temperature in southern Africa (Fig. 4b, Supplementary Fig. 1). Finally, bead styles document another episode of connection after 2 ka when mobile pastoralists enter southern Africa²¹.

Human resilience and regional adaptions

The distinct trajectories of bead characteristics suggest that populations in each region responded to environmental changes with different social strategies. The eastern bead tradition is continuous, and its characteristics remain steady, regardless of any climatic shifts. This consistency hints at the presence of resilient intraregional social networks that remain intact even throughout 50 kyr of environmental uncertainty. Owing to the overall higher net primary production and carrying capacity, populations in eastern Africa may have sustained larger sizes or more robust social networks as a strategy to mitigate climate change. By contrast, southern African OES bead characteristics vary widely, and bead use was rare from 30 to 19 ka. This may reflect a strategy where populations lived in smaller, disconnected groups, with less need for symbolic behaviour (Supplementary Discussion 1). Other archaeological evidence from this time seems to support this, showing a staggered technological transition in Marine Isotope Stage 2–3 with possible coexisting but culturally unique sub-populations in southern Africa^{44,45}. The proliferation of consistently sized beads after 19 ka suggests an increasing reliance on symbolic behaviour after climate conditions improved. These regional differences highlight the flexibility of human social behaviour and illustrate variable strategies for coping with environmental challenges in the Late Pleistocene.

Perspective

Our research presents a new line of evidence to help to disentangle complex interactions between ancient populations that are difficult to understand through genetic data alone. The stylistic variation of OES beads reveals intermittent connections between eastern and southern African populations over the past 50 kyr, including the oldest regional stylistic connection ever identified. Furthermore, our findings suggest that cultural contact persisted long after the genetic divergence estimate of 70 ka. This raises interesting questions about whether these social connections existed independently from population admixture or coexisted with biological introgression. Future research is warranted to explore these scenarios. In addition, we find it plausible that climatic variability and human behavioural responses affected interregional social networks by conditioning where and when people could meet. Researchers can build on this foundation by incorporating OES bead data from site-based studies to refine the broader regional comparisons (Supplementary Discussion 6).

Methods

Data collection

This study reports data from 1,516 OES beads from 31 sites across sub-Saharan Africa (Fig. 1). Of these, 1,238 beads from 11 sites are described, while previously published data includes 290 beads from 21 sites (Supplementary Table 1). We used all relevant data with no statistical methods used to predetermine sample size. We selected completed beads, based on the criteria by Orton⁴⁸: ground to a circular shape, presence of use-wear, or were completed and broken with more than 50% remaining. The majority of the specimens were analysed in-person ($n = 1,148$), under low-power magnification, and photographed with a digital microscope. We recorded three metric variables (bead diameter, aperture diameter and thickness) wherever possible. These are the most frequently reported, standardized characteristics in published literature. Bead colour and shape are less commonly reported and may be more subject to interobserver error, so these and other qualitative variables have been excluded from this study. Bead diameter and aperture diameter both result from cultural behaviour, whereas shell thickness instead may reflect a complex relationship between pore density, environmental aridity³¹ and ostrich sub-species³². To obtain diameter values in-person, multiple measurements were taken around the perimeter of the bead using digital calipers. Beads measured from photos with a visible scale were processed in ImageJ to obtain diameter and aperture diameter measurements. As not all beads or apertures are perfectly round, minimum and maximum measurements were used to generate an average, and this value was used in the analysis. Published measurements

could not be assessed in the same manner, and in these cases, the reported average diameter of completed beads was used. We only included beads that have estimated age (either by direct dates, dating of an excavation layer or by averaging bracketing dates from surrounding layers). Wherever possible, we calibrated the original radiocarbon age with either Intcal13⁴⁹ or Intcal20⁵⁰. All phase V data were calibrated with Intcal20, and randomly selected ages from other phases were also calibrated with Intcal20⁵⁰. Differences between the two were minor enough that recalibration of the entire dataset was unwarranted. No blinding or randomization were required for this study.

Statistical analyses

All statistical analyses were performed in R version 4.0.1⁵¹ with RStudio interface version 1.3.959. We grouped the last 50 kyr into five periods: phase I: 50–33 ka (Marine Isotope Stage 3 to time of renewed ice-sheet growth); phase II: 33–19 ka (the onset of global ice sheet growth to the Last Glacial Maximum); phase III: 19–11.6 ka (last deglaciation); phase IV: 11.6–2 ka (Early Holocene to the spread of herding into southern Africa); and phase V: 2 ka to present (spread of herding into southern Africa to present). Phase I and phase II contain the majority of data points, whereas phase V has the least, and southern Africa only has one point for phase IV.

We grouped the 31 sites into two geographical regions (southern Africa and eastern Africa) instead of examining bead characteristics between sites, for two reasons. First, the number of data points between sites was extremely uneven. The largest dataset (Nelson Bay Cave, South Africa) has $n = 529$, whereas 15 sites have less than six data entries, and seven of these only have one data entry. OES bead data distributed unevenly through time with more OES beads in the later phases than phase I and phase II, probably due to myriad factors. Notably, phase I has 39, phase II has 97 and phases III–V contain 1,380 samples. The sampling difference between regions and time periods is an unavoidable outcome of archaeological data. Second, the differences between sites were negligible compared with the differences between regions. A two-sample *t*-test between southern and eastern Africa shows that for regional difference, the OES diameter was significantly different (two-sample *t*-test, mean diameter = 6.9 mm and 4.5 mm, respectively, $t = 34.1$, d.f. = 510.9, $P < 0.0001$). For example, the mean diameters from the southern African sites of Nelson Bay Cave and Wonderwerk Cave (both of which have more than 350 data points) are 4.4 mm and 4.6 mm, respectively. Whereas the average diameters for eastern African sites Enkapune ya Muto and Mumba Rockshelter (both of which have approximately 80 data points) are 6.7 mm and 6.2 mm. Therefore, we suggest that classifying OES data by region is appropriate. Furthermore, despite eastern Africa having far fewer numbers of beads than southern Africa, they are consistently present throughout all five phases.

By contrast, OES beads are largely absent during phase II (33–19 ka), and our dataset only includes one bead from this period.

We applied principal component analysis (PCA; R-package vegan⁵²) using multi-dimensional information (bead diameter, aperture diameter and shell thickness) to examine variation in eastern and southern Africa over the past 50 kyr. Out of 1,516 beads, 1,333 had all three parameters available, so these are included in the PCA. A covariance matrix PCA was used to preserve variance because the range and scale of variables are in the same units of measure. We also applied MANOVA to examine the OES bead variation in these two regions through time using the following steps: (1) we conducted a two-way MANOVA to investigate whether the three bead parameters were influenced by region, phases and/or the interactions between region and phases. (2) We applied Pillai's trace MANOVA to test the null hypothesis that there is no significant difference in OES characteristics between southern Africa and eastern Africa for phases I–III. (3) We applied Pillai's trace MANOVA to test the null hypothesis that there is no significant difference in OES characteristics through time for each region. We did not conduct any statistical tests for phase II because there are insufficient data available from southern Africa. (4) For phase V, we compared OES beads of two regions based on the diameter and aperture diameter using MANOVA, as thickness records from eastern Africa are incomplete (only one sample has recorded thickness). (5) Finally, we used univariate ANOVA performed on the output from all MANOVA to assess which bead parameters are important for driving the differences in OES sizes in regions and through time.

Although not every specimen has all three parameters recorded, every entry at least has a bead diameter, so analysis by diameter provides the larger dataset. Shell thickness is the least likely to be absent as it cannot be accurately measured from photos and is inconsistently reported. Therefore, to increase sample size and visually demonstrate bead variability, we also created plots that include fewer variables. Specifically, we plotted all reported bead parameters against time, running a generalized additive model for each of these variables in both regions to show understanding of how sizes of these parameters evolve through time. Unless otherwise stated, statistical significance is assessed at $P < 0.01$. All PCA figures were made using ‘ggplot2’ packages⁵³.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

All data generated or analysed for this study are included in this published Article (and Supplementary Table 1). All statistical analyses were performed in R version 4.0.151 with RStudio interface version 1.3.959. All PCA figures were made using ‘ggplot2’ version 3.3.5. The data and R code are available from GitHub (<https://github.com/alsjmonsoon/Ostrich-egg-shell-bead-data>).

References

1. 1. Scerri, E. M. L. et al. Did our species evolve in subdivided populations across Africa, and why does it matter? *Trends Ecol. Evol.* **33**, 582–594 (2018).
2. 2. Scerri, E. M. L., Chikhi, L. & Thomas, M. G. Beyond multiregional and simple out-of-Africa models of human evolution. *Nat. Ecol. Evol.* **3**, 1370–1372 (2019).
3. 3. Schlebusch, C. M. et al. Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* **358**, 652–655 (2017).
4. 4. Skoglund, P. et al. Reconstructing prehistoric African population structure. *Cell* **171**, 59–71.e21 (2017).
5. 5. Will, M., Conard, N. J. & Tryon, C. in *Modern Human Origins and Dispersal* (eds Sahle, Y., Reyes-Centeno, H. & Bentz, C.) 25–72 (Kerns, 2019).
6. 6. Bergström, A., Stringer, C., Hajdinjak, M., Scerri, E. M. L. & Skoglund, P. Origins of modern human ancestry. *Nature* **590**, 229–237 (2021).
7. 7. Faith, J. T., Tryon, C. A. & Peppe, D. J. in *Africa from MIS 6-2* (eds Jones, S. C. & Stewart, B. A.) 233–245 (Springer Netherlands, 2016).

8. 8.

Lipson, M. et al. Ancient West African foragers in the context of African population history. *Nature* **577**, 665–670 (2020).

9. 9.

Tryon, C. A. The Middle/Later Stone Age transition and cultural dynamics of Late Pleistocene East Africa. *Evol. Anthropol.* **28**, 267–282 (2019).

10. 10.

Mirazón Lahr, M. The shaping of human diversity: filters, boundaries and transitions. *Phil. Trans. R. Soc. B* **371**, 20150241 (2016).

11. 11.

Powell, A., Shennan, S. & Thomas, M. G. Late Pleistocene demography and the appearance of modern human behavior. *Science* **324**, 1298–1301 (2009).

12. 12.

McBrearty, S. & Brooks, A. S. The revolution that wasn't: a new interpretation of the origin of modern human behavior. *J. Hum. Evol.* **39**, 453–563 (2000).

13. 13.

Kuhn, S. L. & Stiner, M. C. in *Rethinking the Human Revolution: New Behavioural and Biological Perspectives on the Origin and Dispersal of Modern Humans* (eds Mellars, P. et al.) 45–54 (Oxbow Books, 2007).

14. 14.

d'Errico, F. et al. Early evidence of San material culture represented by organic artifacts from Border Cave, South Africa. *Proc. Natl Acad. Sci. USA* **109**, 13214–13219 (2012).

15. 15.

d'Errico, F. & Vanhaeren, M. in *Death Rituals, Social Order and the Archaeology of Immortality in the Ancient World* (eds Renfrew, C., Boyd, M. J. & Morley, I.) 45–62 (Cambridge Univ. Press, 2016).

16. 16.

Wiessner, P. Style and social information in Kalahari San projectile points. *Am. Antiq.* **48**, 253–276 (1983).

17. 17.

Sackett, J. R. Approaches to style in lithic archaeology. *J. Anthropol. Archaeol.* **1**, 59–112 (1982).

18. 18.

Shennan, S. J., Crema, E. R. & Kerig, T. Isolation-by-distance, homophily, and “core” vs. “package” cultural evolution models in Neolithic Europe. *Evol. Hum. Behav.* **36**, 103–109 (2015).

19. 19.

Crema, E. R., Kerig, T. & Shennan, S. Culture, space, and metapopulation: a simulation-based study for evaluating signals of blending and branching. *J. Archaeol. Sci.* **43**, 289–298 (2014).

20. 20.

Wenger, E. *Communities of Practice: Learning, Meaning, and Identity* (Cambridge Univ. Press, 1998).

21. 21.

Miller, J. M. & Sawchuk, E. A. Ostrich eggshell bead diameter in the Holocene: regional variation with the spread of herding in eastern and southern Africa. *PLoS ONE* **14**, e0225143 (2019).

22. 22.

Jacobson, L. The size variability of ostrich eggshell beads from central Namibia and its relevance as a stylistic and temporal marker. *S. Afr. Archaeol. Bull.* **42**, 55–58 (1987).

23. 23.

Pickrell, J. K. et al. Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl Acad. Sci. USA* **111**, 2632–2637 (2014).

24. 24.

Dayet, L., Erasmus, R., Val, A., Feyfant, L. & Porraz, G. Beads, pigments and Early Holocene ornamental traditions at Bushman Rock Shelter, South Africa. *J. Arch. Sci. Rep.* **13**, 635–651 (2017).

25. 25.

d'Errico, F. et al. Trajectories of cultural innovation from the Middle to Later Stone Age in eastern Africa: personal ornaments, bone artifacts, and ochre from Panga Ya Saidi, Kenya. *J. Hum. Evol.* **141**, 102737 (2020).

26. 26.

Tryon, C. A. et al. Middle and Later Stone Age chronology of Kisese II Rockshelter (UNESCO World Heritage Konda Rock-Art Sites), Tanzania. *PLoS ONE* **13**, e0192029 (2018).

27. 27.

Collins, B., Wojcieszak, M., Nowell, A., Hodgsiss, T. & Ames, C. J. H. Beads and bead residues as windows to past behaviours and taphonomy: a case study from Grassridge Rockshelter, eastern Cape, South Africa. *Archaeol. Anthropol. Sci.* **12**, 192 (2020).

28. 28.

Ziegler, M. et al. Development of Middle Stone Age innovation linked to rapid climate change. *Nat. Commun.* **4**, 1905 (2013).

29. 29.

Tierney, J. E. et al. Northern hemisphere controls on tropical southeast African climate during the past 60,000 years. *Science* **322**, 252–255 (2008).

30. 30.

Wang, Y. V. et al. Northern and southern hemisphere controls on seasonal sea surface temperatures in the Indian Ocean during the last deglaciation. *Paleoceanography* **28**, 619–632 (2013).

31. 31.

Ecker, M., Botha-Brink, J., Lee-Thorp, J. A., Piuz, A. & Horwitz, L. K. in *Changing Climates, Ecosystems and Environments Within Arid Southern Africa and Adjoining Region* vol. 33 (ed. Runge, J.) 95–115 (CRC Press, 2015).

32. 32.
- Harrison, T. & Msuya, C. P. Fossil Struthionid eggshells from Laetoli, Tanzania: taxonomic and biostratigraphic significance. *J. Afr. Earth. Sci.* **41**, 303–315 (2005).
33. 33.
- Cooper, R. G. et al. The wild ostrich (*Struthio camelus*): a review. *Trop. Anim. Health Prod.* **41**, 1669–1678 (2009).
34. 34.
- Miller, J. M. & Willoughby, P. R. Radiometrically dated ostrich eggshell beads from the Middle and Later Stone Age of Magubike Rockshelter, southern Tanzania. *J. Hum. Evol.* **74**, 118–122 (2014).
35. 35.
- Timmermann, A. & Friedrich, T. Late Pleistocene climate drivers of early human migration. *Nature* **538**, 92–95 (2016).
36. 36.
- Morgan, C. Climate change, uncertainty and prehistoric hunter–gatherer mobility. *J. Anthropol. Archaeol.* **28**, 382–396 (2009).
37. 37.
- Dyson-Hudson, R. & Smith, E. A. Human territoriality: an ecological reassessment. *Am. Anthropol.* **80**, 21–41 (1978).
38. 38.
- Kusimba, S. B. Hunter–gatherer land use patterns in Later Stone Age East Africa. *J. Anthropol. Archaeol.* **18**, 165–200 (1999).
39. 39.
- d'Errico, F. et al. Additional evidence on the use of personal ornaments in the Middle Paleolithic of North Africa. *Proc. Natl Acad. Sci. USA* **106**, 16051–16056 (2009).
40. 40.

Hollfelder, N., Breton, G., Sjödin, P. & Jakobsson, M. The deep population history in Africa. *Hum. Mol. Genet.* **30**, R2–R10 (2021).

41. 41.

Schlebusch, C. M. et al. Khoe-San genomes reveal unique variation and confirm the deepest population divergence in *Homo sapiens*. *Mol. Biol. Evol.* **37**, 2944–2954 (2020).

42. 42.

Ziemer, F. A., Kapsch, M.-L., Klockmann, M. & Mikolajewicz, U. Heinrich events show two-stage climate response in transient glacial simulations. *Clim. Past* **15**, 153–168 (2019).

43. 43.

van der Lubbe, H. J. L., Frank, M., Tjallingii, R. & Schneider, R. R. Neodymium isotope constraints on provenance, dispersal, and climate-driven supply of Zambezi sediments along the Mozambique Margin during the past ~45,000 years. *Geochem. Geophys. Geosyst.* **17**, 181–198 (2016).

44. 44.

Mackay, A., Stewart, B. A. & Chase, B. M. Coalescence and fragmentation in the Late Pleistocene archaeology of southernmost Africa. *J. Hum. Evol.* **72**, 26–51 (2014).

45. 45.

Bousman, C. B. & Brink, J. S. The emergence, spread, and termination of the Early Later Stone Age event in South Africa and southern Namibia. *Quat. Int.* **495**, 116–135 (2018)

46. 46.

Lyons, R. P. et al. Continuous 1.3-million-year record of East African hydroclimate, and implications for patterns of evolution and biodiversity. *Proc. Natl Acad. Sci. USA* **112**, 15568–15573 (2015).

47. 47.

Wang, Y. V. et al. What does leaf wax δD from a mixed C3/C4 vegetation region tell us? *Geochim. Cosmochim. Acta* **111**, 128–139 (2013).

48. 48.

Orton, J. Later Stone Age ostrich eggshell bead manufacture in the Northern Cape, South Africa. *J. Archaeol. Sci.* **35**, 1765–1775 (2008).

49. 49.

Reimer, P. J. et al. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).

50. 50.

Reimer, P. J. et al. The IntCal20 Northern hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).

51. 51.

R Core Team. *R: a Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).

52. 52.

Oksanen, J. et al. vegan: community ecology package (2019).

53. 53.

Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).

Acknowledgements

We thank the University of Alberta, Department of Anthropology, for hosting PhD research by J.M.M. For permissions and assistance with accessing archaeological collections for this research, we recognize the following: Tanzanian Commission on Science and Technology (COSTECH), Tanzanian Division of Antiquities; Apollo 11 Cave—National Museum of Namibia, Windhoek; Border Cave—McGregor Museum, Kimberley, F. d’Errico; Daumboy 3 Rockshelter—M. Prendergast (also A. Mabulla); Dikbosch Rockshelter—McGregor Museum, Kimberley, D. Morris; Enkapune ya Muto—S. Ambrose and P. Slater; Kakapel Rockshelter—S. Goldstein (also E. Ndiema and N. Boivin); Magubike and Mlambalasi Rockshelters—P. Willoughby; Mumba Rockshelter—M. Prendergast (also A. Mabulla and M. Domingo-Rodriguez); Nelson Bay Cave—Iziko Museums of South Africa; White Paintings Shelter—Ministry of Natural Resources and Tourism, Government of the Republic of Botswana, Ministry of Environment, Wildlife and Tourism, National Museum of Gaborone, L. Robins; and

Wonderwerk Cave—McGregor Museum, Kimberley, M. Chazan. We thank U. Krebs-Kanzow for downloading the LOVECLIM data and insightful discussion on palaeoclimate model results; X. Xu for discussing palaeoresults; E. Zeller for writing a script for downloading LOVECLIM data on the Python interface; and K. Bishop, E. Finestone, S. Goldstein, A. Kay, V. van der Haas, J. Thompson, E. Sawchuk, M. Stewart, M. Storozum and J. Wolfhagen for providing feedback on this manuscript.

Funding

Open access funding provided by Max Planck Society.

Author information

Affiliations

1. Department of Archaeology, Max Planck Institute for the Science of Human History, Jena, Germany

Jennifer M. Miller & Yiming V. Wang

2. Department of Anthropology, University of Alberta, Edmonton, Alberta, Canada

Jennifer M. Miller

Contributions

Conception, data procurement, data analysis and writing of the manuscript were done by J.M.M. Conception, data analysis and manuscript writing were done by Y.V.W.

Corresponding authors

Correspondence to [Jennifer M. Miller](#) or [Yiming V. Wang](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Benjamin Collins, Alex Mackay and Axel Timmermann for their contribution to the peer review of this work.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information

Supplementary Information

This file contains Supplementary Discussion; Supplementary Figs. 1 – 5; legend for Supplementary Table 1; Supplementary Tables 2–11 and their accompanying legends and Supplementary References.

Reporting Summary

Supplementary Table 1

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Reprints and Permissions

About this article

Cite this article

Miller, J.M., Wang, Y.V. Ostrich eggshell beads reveal 50,000-year-old social network in Africa. *Nature* **601**, 234–239 (2022). <https://doi.org/10.1038/s41586-021-04227-2>

- Received: 04 June 2021

- Accepted: 10 November 2021
- Published: 20 December 2021
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04227-2>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [**Beads reveal long-distance connections in early Africa**](#)
 - Benjamin R. Collins
 - Amy Hatton

Nature (2022)

[**Beads reveal long-distance connections in early Africa**](#)

- Benjamin R. Collins
- Amy Hatton

News & Views 20 Dec 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04227-2>

- Article
- [Published: 08 December 2021](#)

Adaptive stimulus selection for consolidation in the hippocampus

- [Satoshi Terada](#)^{1,2},
- [Tristan Geiller](#) [ORCID: orcid.org/0000-0001-8957-7786](#)^{1,2} na1,
- [Zhenrui Liao](#)^{1,2} na1,
- [Justin O'Hare](#)^{1,2} na1,
- [Bert Vancura](#)^{1,2} na1 &
- [Attila Losonczy](#) [ORCID: orcid.org/0000-0002-7064-0252](#)^{1,2,3}

[Nature](#) volume **601**, pages 240–244 (2022)

- 5381 Accesses
- 35 Altmetric
- [Metrics details](#)

Subjects

- [Consolidation](#)
- [Hippocampus](#)
- [Learning algorithms](#)

Abstract

Associative memories guide behavioural adaptation by binding together outcome-predictive sensory stimuli^{1,2}. However, in a feature-rich environment, only a subset of stimuli may predict a desired outcome^{3,4}.

How neural circuits enable behavioural adaptation by selectively and durably representing subsets of sensory stimuli that are pertinent to a specific outcome is not known. We investigated this feature selection process in the hippocampus during memory acquisition and subsequent consolidation. Two-photon calcium imaging of CA3 axonal projections to CA1 combined with simultaneous local field potential recordings revealed that CA3 projections that encode behaviourally informative sensory stimuli were selectively recruited during the memory replay events that underlie hippocampal memory consolidation⁵. These axonal projections formed sequential assemblies that conjunctively link sensory features to spatial location and thus reward proximity. By contrast, axons encoding uninformative, peripatetic sensory cues were notably suppressed during memory replay. Thus, while the hippocampus comprehensively encodes the real-time sensory environment, it implements a flexible filtering mechanism to maximize the utility of memories destined for long-term storage. We propose that utility-dependent recruitment of sensory experience during memory consolidation is a general coding principle for associative learning.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Fig. 1: Rapid reorganization of CA3SC activity during sensory experience.



Fig. 2: Suppressed reactivation of task-irrelevant sensory information during SWRs.

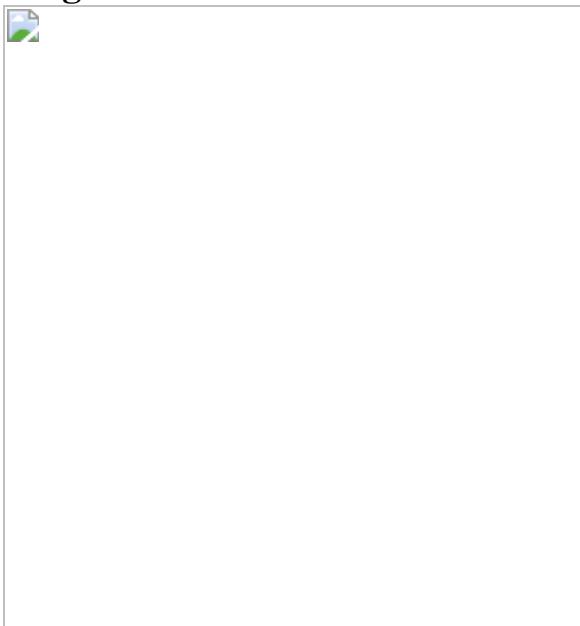


Fig. 3: Adaptive switch in reactivation of cue-CA3SCs during SWRs.

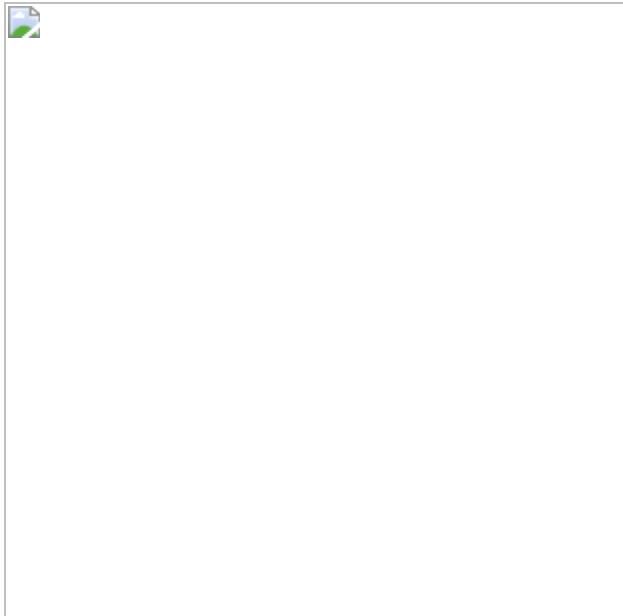


Fig. 4: Replay of sensory cue-driven cognitive maps in CA3SCs.



Data availability

Datasets included in this study are available from the corresponding author upon reasonable request. [Source data](#) are provided with this paper.

Code availability

Custom Matlab code supporting this study is available at https://github.com/losonczylab/Terada_Nature2021. Source data is provided with this paper.

References

1. 1.

Eichenbaum, H. A cortical–hippocampal system for declarative memory. *Nat. Rev. Neurosci.* **1**, 41-50, (2000).

2. 2.

Whittington, J. C. R. et al. The Tolman–Eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell* **183**, 1249-1263, (2020).

3. 3.

Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Ann. Rev. Neurosci.* **18**, 193-222, (1995).

4. 4.

Knudsen, E. I. Fundamental components of attention. *Ann. Rev. Neurosci.* **30**, 57-78, (2007).

5. 5.

Buzsaki, G. Hippocampal sharp wave-ripple: a cognitive biomarker for episodic memory and planning. *Hippocampus* **25**, 1073-1188, (2015).

6. 6.

Kentros, C. G., Agnihotri, N. T., Streater, S., Hawkins, R. D. & Kandel, E. R. Increased attention to spatial context increases both place field stability and spatial memory. *Neuron* **42**, 283-295 (2004).

7. 7.

Rowland, D. C. & Kentros, C. G. Potential anatomical basis for attentional modulation of hippocampal neurons. *Ann. N.Y. Acad. Sci.*

1129, 213-224, (2008).

8. 8.

Aronov, D., Nevers, R. & Tank, D. W. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature* **543**, 719-722, (2017).

9. 9.

Constantinescu, A. O., O'Reilly, J. X. & Behrens, T. E. J. Organizing conceptual knowledge in humans with a gridlike code. *Science* **352**, 1464-1468, (2016).

10. 10.

Kesner, R. P. & Rolls, E. T. A computational theory of hippocampal function, and tests of the theory: new developments. *Neurosci. Biobehav. Rev.* **48**, 92-147, (2015).

11. 11.

Nakazawa, K. et al. Hippocampal CA3 NMDA receptors are crucial for memory acquisition of one-time experience. *Neuron* **38**, 305-315 (2003).

12. 12.

Kudrimoti, H. S., Barnes, C. A. & McNaughton, B. L. Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and EEG dynamics. *J. Neurosci.* **19**, 4090-4101 (1999).

13. 13.

Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676-679 (1994).

14. 14.

Jadhav, S. P., Kemere, C., German, P. W. & Frank, L. M. Awake hippocampal sharp-wave ripples support spatial memory. *Science* **336**, 1454-1458, (2012).

15. 15.

Nakashiba, T., Buhl, D. L., McHugh, T. J. & Tonegawa, S. Hippocampal CA3 output is crucial for ripple-associated reactivation and consolidation of memory. *Neuron* **62**, 781-787, (2009).

16. 16.

Nakashiba, T., Young, J. Z., McHugh, T. J., Buhl, D. L. & Tonegawa, S. Transgenic inhibition of synaptic transmission reveals role of CA3 output in hippocampal learning. *Science* **319**, 1260-1264, (2008).

17. 17.

Csicsvari, J., Hirase, H., Mamiya, A. & Buzsaki, G. Ensemble patterns of hippocampal CA3–CA1 neurons during sharp wave-associated population events. *Neuron* **28**, 585-594, (2000).

18. 18.

Diba, K. & Buzsaki, G. Forward and reverse hippocampal place-cell sequences during ripples. *Nat. Neurosci.* **10**, 1241-1242, (2007).

19. 19.

Danielson, N. B. et al. Sublayer-specific coding dynamics during spatial navigation and learning in hippocampal area CA1. *Neuron* **91**, 652-665, (2016).

20. 20.

Zaremba, J. D. et al. Impaired hippocampal place cell dynamics in a mouse model of the 22q11.2 deletion. *Nat. Neurosci.* **20**, 1612-1623, (2017).

21. 21.

O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171-175 (1971).

22. 22.

Hunt, D. L., Linaro, D., Si, B., Romani, S. & Spruston, N. A novel pyramidal cell type promotes sharp-wave synchronization in the hippocampus. *Nat. Neurosci.* **21**, 985-995, (2018).

23. 23.

Soltesz, I. & Losonczy, A. CA1 pyramidal cell diversity enabling parallel information processing in the hippocampus. *Nat. Neurosci.* **21**, 484-493, (2018).

24. 24.

Mysore, S. P. & Knudsen, E. I. Reciprocal inhibition of inhibition: a circuit motif for flexible categorization in stimulus selection. *Neuron* **73**, 193-205, (2012).

25. 25.

Geiller, T. et al. Large-scale 3D two-photon imaging of molecularly identified CA1 interneuron dynamics in behaving mice. *Neuron* **108**, 968-983, (2020).

26. 26.

Gupta, A. S., van der Meer, M. A., Touretzky, D. S. & Redish, A. D. Hippocampal replay is not a simple function of experience. *Neuron* **65**, 695-705, (2010).

27. 27.

Wu, C. T., Haggerty, D., Kemere, C. & Ji, D. Hippocampal awake replay in fear memory retrieval. *Nat. Neurosci.* **20**, 571-580, (2017).

28. 28.

Ahmed, M. S. et al. Hippocampal network reorganization underlies the formation of a temporal association memory. *Neuron*, <https://doi.org/10.1016/j.neuron.2020.04.013> (2020).

29. 29.

Liu, Y., Mattar, M. G., Behrens, T. E. J., Daw, N. D. & Dolan, R. J. Experience replay is associated with efficient nonlocal learning. *Science* **372**, <https://doi.org/10.1126/science.abf1357> (2021).

30. 30.

Grosmark, A. D. & Buzsaki, G. Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. *Science* **351**, 1440-1443, (2016).

31. 31.

Kaufman, A. M., Geiller, T. & Losonczy, A. A role for the locus coeruleus in hippocampal CA1 place cell reorganization during spatial reward learning. *Neuron* **105**, 1018-1026, (2020).

32. 32.

Kaifosh, P., Lovett-Barron, M., Turi, G. F., Reardon, T. R. & Losonczy, A. Septo-hippocampal GABAergic signaling across multiple modalities in awake mice. *Nat. Neurosci.* **16**, 1182-1184, (2013).

33. 33.

Lovett-Barron, M. et al. Dendritic inhibition in the hippocampus supports fear learning. *Science* **343**, 857-863, (2014).

34. 34.

Pachitariu, M. et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. Preprint at *bioRxiv*, 061507, <https://doi.org/10.1101/061507> (2016).

35. 35.

Hainmueller, T. & Bartos, M. Parallel emergence of stable and dynamic memory engrams in the hippocampus. *Nature* **558**, 292-296, (2018).

36. 36.

Dong, C., Madar, A. D. & Sheffield, M. E. J. Distinct place cell dynamics in CA1 and CA3 encode experience in new environments. *Nat. Commun.* **12**, 2977, (2021).

37. 37.

Rajasethupathy, P. et al. Projections from neocortex mediate top-down control of memory retrieval. *Nature*, <https://doi.org/10.1038/nature15389> (2015).

38. 38.

Ali, F. & Kwan, A. C. Interpreting *in vivo* calcium signals from neuronal cell bodies, axons, and dendrites: a review. *Neurophotonics* **7**, 011402, (2020).

39. 39.

Cox, C. L., Denk, W., Tank, D. W. & Svoboda, K. Action potentials reliably invade axonal arbors of rat neocortical neurons. *Proc. Natl Acad. Sci. USA* **97**, 9724-9728, (2000).

40. 40.

Petreanu, L. et al. Activity in motor-sensory projections reveals distributed coding in somatosensation. *Nature* **489**, 299-303, (2012).

41. 41.

Dugladze, T., Schmitz, D., Whittington, M. A., Vida, I. & Gloveli, T. Segregation of axonal and somatic activity during fast network oscillations. *Science* **336**, 1458-1461, (2012).

42. 42.

Williams, A. H. et al. Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron* **98**, 1099-1115, (2018).

43. 43.

Skaggs, W. E., Mcnaughton, B. L., Markus, E. J. & Gothard, K. M. in *Advances in Neural Information Process Systems* (eds Hanson, S. et al.) Vol. 5, 1030–1037 (Morgan Kaufmann, 1993).

44. 44.

Mackevicius, E. L. et al. Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. *eLife* **8**, <https://doi.org/10.7554/eLife.38471> (2019).

45. 45.

Malvache, A., Reichinnek, S., Villette, V., Haimerl, C. & Cossart, R. Awake hippocampal reactivations project onto orthogonal neuronal assemblies. *Science* **353**, 1280-1283, (2016).

46. 46.

Ecker, A. et al. Hippocampal sharp wave-ripples and the associated sequence replay emerge from structured synaptic interactions in a network model of area CA3. Preprint at *bioRxiv*,

2021.2002.2018.431868, <https://doi.org/10.1101/2021.02.18.431868> (2021).

47. 47.

Friedrich, J., Zhou, P. & Paninski, L. Fast online deconvolution of calcium imaging data. *PLoS Comput. Biol.* **13**, e1005423, (2017).

Acknowledgements

S.T. is supported by a JSPS Overseas Fellowship. A.L. is supported by National Institute of Mental Health (NIMH) 1R01MH124047 and 1R01MH124867, National Institute of Neurological Disorders and Stroke (NINDS) 1U19NS104590 and 1U01NS115530, and the Kavli Foundation. B.V. is supported by grants (NIH) T32GM007367 and (NIMH) F30MH125628, J.O. is supported by F32MH118716. Z.L. is supported by NINDS F31NS120783-01 and NIH T32GM007367. We thank S. Fusi, M. Shadlen, J. Gottlieb, F. Polleux and I. Soltesz, as well as members of the Losonczy laboratory and D. Hadjiabadi (Soltesz laboratory), for their invaluable comments on previous versions of the manuscript.

Author information

Author notes

1. These authors contributed equally: Tristan Geiller, Bert Vancura, Justin O'Hare, Zhenrui Liao

Affiliations

1. Department of Neuroscience, Columbia University, New York, NY, USA

Satoshi Terada, Tristan Geiller, Zhenrui Liao, Justin O'Hare, Bert Vancura & Attila Losonczy

2. Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA

Satoshi Terada, Tristan Geiller, Zhenrui Liao, Justin O'Hare, Bert Vancura & Attila Losonczy

3. The Kavli Institute for Brain Science, Columbia University, New York, NY, USA

Attila Losonczy

Contributions

S.T. and A.L. designed the project and experiments. S.T. performed the experiments with help from T.G. and B.V. S.T. analysed the data with help from Z.L., J.O. and T.G. All authors wrote the manuscript.

Corresponding author

Correspondence to [Attila Losonczy](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Aman Saleem and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Labeling of CA3 Schaffer Collaterals for *in vivo* two-photon calcium imaging and identification of ROIs.

a, Representative coronal slice of the dorsal hippocampus of a *Grik4-Cre* mouse showing Cre-dependent rAAV-driven GCaMP6f expression restricted to CA3. The CA3 pyramidal cell layer is outlined with the dotted white contour, and the approximate locations of the CA3 subregions (*CA3a-c*) are shown. **b**, Quantification of relative fluorescence intensity along the pyramidal cell layer of CA3 subregions (mean \pm s.e.m.). Fluorescence intensity at 10 locations along the CA3 pyramidal cell layer was measured, starting from the hilus of the dentate gyrus ($x=0\text{ }\mu\text{m}$) to the approximate CA2/3 border ($x=900\text{ }\mu\text{m}$). **c**, Example *in vivo* two-photon field of view (FOV) in *CA1 SR* from one representative mouse ($n=6$ in this study). Imaging focal plane was located 50–150 μm below CA1 SP. *Right*, the same FOV after ROI segmentation with the Suite2p analysis package (*middle*, overlapping display). **d**, Distributions of correlation coefficients between pairs of CA3SC ROIs during spontaneous activity for each individual mouse (black lines) and mean for all mice (green). **e**, Estimation of number of unique CA3 units recorded using hierarchical clustering and silhouette score ([Methods](#)). *Top*, silhouette values as number of clusters K (to the total number of detected ROIs) in an example FOV (green: average, pink: median, light blue: 75th percentile, red: 25th percentile). *Bottom*, marginal change in silhouette values each time K is incremented by 1. The estimated number of unique CA3 units recorded on each FOV was defined as the number of clusters that maximized the rate of change (528 clusters in this case, and 644 ± 91 clusters, mean \pm s.e.m.). **f**, *Left*, correlation matrix of ROI signals from the example FOV shown in (c). *Right top*, zoom: example clusters of putative same-axon ROIs (red), and non-clustered ROIs, likely recorded from distinct axons (blue). **g**, Median silhouette values over percentage of putative distinct CA3 axons as a proportion of total number of ROIs detected in the FOVs (color scheme as in e). The pink shaded area indicates the ranges between the mice. Dashed lines indicate the cluster numbers that maximized rate of change in 25th (red shaded area on right) and 75th percentiles (blue shaded area on left). **h**, Boxplots of within-cluster

correlations of ROI signals all together (*left*, n=175 ROIs) and by mouse (*right*, 29.17 ± 9.5 ROIs, mean \pm s.e.m.). The whiskers indicate maxima and minima. **i**, Fraction of ROIs putatively originating from the same cell across mice (n=6810 ROIs from 6 mice). The whiskers indicate maxima and minima, outliers excluded.

[Source data](#)

Extended Data Fig. 2 Stability of two-photon imaging.

a, *Top*, cue onsets in a representative RC-NS session (orange: odor; green: visual; light blue: reward). *Bottom*, trial-averaged FOVs corresponding to each odor trial from one representative mouse (n=6 in this study). Trial averages were computed from the frame of odor stimulus onset through the last frame 5 s prior to the next stimulus onset. **b**, Mean pixel-wise correlations between all pairs of trial-averaged FOVs from the above session. **c**, Mean pixel-wise correlations of trial-averaged FOVs with whole-session time-averaged FOVs in RC-NS sessions. Colors correspond to individual mice. Correlations were recalculated with whole-session time-averaged FOVs on different mice (shuffled) and plotted in black (mean, shaded area: 90% confidence interval). **d**, *Left*, empirical cumulative distribution function (CDF) of mean squared error (MSE) between each frame and time-averaged FOVs from corresponding sessions (green: all frames, orange: intra-trial frames, black: comparison to other FOVs). *Middle*, violin plot of data shown in CDFs. *Right*, MSEs between each frame and trial-averaged FOVs. MSEs were averaged every 10 trials and shaded areas indicate percentiles from 5% to 95%. **e–i**, Same as (a) for each individual mouse.

[Source data](#)

Extended Data Fig. 3 Reward-predictive behavior during sensory-place associative paradigm.

a, Schematic of sensory stimulation paradigm and belt types ([Methods](#)). **b**, Examples of velocity in individual laps under different conditions. Each shaded color area corresponds to textural elements of the spatial belt where

the sensory stimulations were presented and dashed lines indicate cue onsets. **c**, Representative lap-by-lap (top) and mean (bottom) velocity during FC-S. **d**, Velocity as predicted by generalized linear model (GLM, Method). Lines indicate mean GLM-predicted velocity and shaded areas indicate 25th and 75th percentiles across mice. **e**, Comparison of velocities before (black) and after (green) the onset of visual stimulus. *Top*, cumulative histograms of mean velocities 50 cm before and after visual stimulus. *Bottom*, velocity modulation by visual stimulus calculated with real pairs of velocities (green line) vs null distribution of scores calculated on 1000 random permutations of the onset ID (black histogram). Dotted lines indicate significance thresholds corresponding to p-values of 0.05 and 0.01. **f**, Examples of cue-CA3SC ROIs showing the different trial-by-trial transitions (*left and middle*; odor-, right; multi-sensory CA3SC ROIs). Heatmap shows $\Delta F/F_0$ by trial, chunked by cue modality and centered on cue onset with mean values plotted on right. PSTHs calculated on subsets of trials are shown below. **g**, Heatmap of mean cue-CA3SC activity centered on running bout initiations during random cue/non-spatial paradigm. **h**, *Top*, velocity (mean±s.e.m.) from all running bout initiations. *Bottom*, mean cue-CA3SC activity shown in (a). **i**, Fraction of significantly running-modulated CA3SCs during RC-NS (Other: n=5,836; Cue: n=1,219 ROIs from 6 mice). 1000-time shuffling test with Bonferroni correction was used to obtain significance thresholds for $P < 0.05$ in each ROIs (two-tailed test, $\alpha=2.5\%$ on each side). **j**, Correlation between cue response and running speed during trial periods represented as boxplots with median and interquartile range (Up: n=276 odor, n=143 visual, n=203 reward; Down: n=136 odor, n=77 visual, 78 reward; Other: n=41 odor, n=196 visual, n=69 reward). The whiskers indicate maxima and minima, outliers excluded. The red plus symbols indicate outliers.

[Source data](#)

[Extended Data Fig. 4 Multi-axis dimensionality reduction of trial-by-trial changes in cue-CA3SC activity via Tensor Component Analysis.](#)

a, Schematic of tensor component analysis (TCA). Imaging data were arranged into a 4th order tensor with dimensions N×T×K×C where axes correspond to the number of CA3SC ROIs, within-trial temporal dynamics, evolution over multiple trials, and cue modality. **b**, Reconstructed (*top*) and original (*bottom*) $\Delta F/F_0$ heatmaps, centered on cue onset, for cue-CA3SCs in response to their respective 1st preferred (*left*), 2nd preferred (*middle*), and least preferred (*right*) cues. A 10-component TCA model, shown in (*e*), was used for reconstruction. **c**, Error plot showing normalized reconstruction error for TCA (inset) and reduction in error with each additional component until 50 for 1st preferred (cyan), 2nd preferred (pink), and least preferred (black) cue. **d**, Median coefficients (R^2) \pm s.d. of reconstruction for each cue-CA3SC with the number of components for TCA ranging from 1 to 20 (n=1219 ROIs from 6 mice). **e**, Extracted tensor components (TCs) from the 5, 10, 15-component models. Columns show temporal factor (*left*), trial factor (*middle left*), cue factor (*middle right*), and ROI factor (*right*). Cyan trial factor: TCs dominated by a single sensory modality (#1-5) Pink trial factor: multimodal TCs (#6-10). Unimodal TCs strongly contributed to reconstructed responses to their 1st preferred cues shown in (*b*) and represent modality-specific features within and across trials. **f**, Distributions of averaged $\Delta F/F_0$ response after cue onsets, by trial, in each cue-CA3SC cluster.

[Source data](#)

Extended Data Fig. 5 Latency and reliability of cue-CA3SC activity across trials.

a, *Left*, Heatmap activity showing responses triggered by all cue onsets across ROIs from each cluster (*Up* and *Down*). Row indicates individual trial. *Right*, response profiles of each cluster. **b**, Latency between cue onset and peak time of cue-CA3SC response in each trial. Each symbol indicates mean response latency of all cue-CA3SCs in each cluster (*Up* (purple): n=276 odor, n=143 visual, n=203 reward; *Down* (yellow): n=136 odor, n=77 visual, 78 reward. Two-way ANOVA with post-hoc Tukey's tests. For odor-CA3SCs: Up x Down: P =0.74.; main effect of trial: P=3.6x10⁻¹⁴; interaction: P=1.8x10⁻⁵; for visual-CA3SCs, Up x Down: P=0.62; main

effect of trial: $P=5.7 \times 10^{-5}$; interaction: $P=2 \times 10^{-4}$; for reward-CA3SCs, Up x Down: $P=0.74$; main effect of trial: $P=9.6 \times 10^{-6}$; interaction: $P=0.0193$). **c**, Reliability of cue response, defined as the number of trials with response divided by the total number of trials (10 trials per modality) represented as boxplots with median and interquartile range. Whiskers denote minimum and maximum values. **d**, Extracted 1st principal component (PC) of the cue response of individual cue-CA3SCs assigned to Up and Down clusters using the TCA/K-means approach. To validate TCA and K-means classification, PCA was applied to K x T (trial x trial time) arrays of individual cue-CA3SCs. The 1st PC of each CA3SC robustly captured the initially-identified cue response. Note that 1st PC coefficients across individual trials represent the trial-by-trial evolution in the cue response of each CA3SC. **e**, Using K-means clustering, coefficient trends across trials were assigned to Up or Down clusters. Mean coefficients are plotted across trials of classified cue-CA3SCs. **f**, Fraction of cue-CA3SCs in inferred clusters. **g**, Overlap rates between TCA and PCA classifications of trial-by-trial trends.

Source data

Extended Data Fig. 6 Additional data on SWR recruitment of cue-CA3SCs in random cue conditions.

a, Top, Example period of velocity and cue onsets. Blue shaded area corresponds to time window in the below heatmap showing examples of peri-SWR CA3SC activity. **Bottom**: SWR band-filtered LFP is plotted on top with detected CA1 SWRs indicated by black ticks and white dashed lines. **b, Top**, Average peri-SWR response in each group shown in Fig. 1 for cue-CA3SCs. **Bottom**, Fraction of SWR recruitments of cue-CA3SCs. Number of ROIs for single-modality: *odor*: 349, *visual*: 256, *reward*: 350. Multi-modality ROIs, $n=264$; with *odor-visual*: 74, *odor-reward*: 30, *visual-reward*: 107, *all-modalities*: 53 (two-sided Fisher's exact test with Bonferroni correction. $P=3.7 \times 10^{-6}$ for Odor-vs-Reward. $P=1.5 \times 10^{-5}$ for Visual-vs-Reward. $P=0.8967$ for Odor-vs-Visual. $P=3.7 \times 10^{-6}$ for down-Reward-vs-down-Odor, -vs-up-Odor ($P=1.6 \times 10^{-7}$), vs-down-Visual ($P=3.6 \times 10^{-5}$), vs-up-Visual ($P=3.4 \times 10^{-5}$). $P=0.0033$ for up-Reward-vs-

down-Odor, and -vs-up-Odor ($P=3.7 \times 10^{-5}$). $P=0.2418$ for up-Reward-vs-down-Visual. $P = 0.0985$ for up-Reward-vs-up-Visual. **c**, *Top*, Cumulative histograms of peri-SWR z-scored $\Delta F/F_0$ of all CA3SCs shown in Fig. 2g (RC-S, Day 4. Other: $n=3,371$; Place: $n=1,644$; Cue-Place: $n=198$, Cue: $n=501$ from 4 mice). *Bottom left*, fraction of cue-, cue-place, place-, and non-responsive CA3SC ROIs-. *Bottom right*, fraction of SWR recruitments of CA3SCs (One-way ANOVA with post-hoc Tukey's test. ROI-type: $P<1.3 \times 10^{-16}$). **d**, *Left*, mean SWR-triggered power spectrogram of contralateral CA1 LFP (black line: an example filtered LFP trace). *Right*, peri-cue onset time histogram of SWR rate for all mice and average rate stem plot juxtaposed to random, non-cue-centered onset times ('outside'). Dashed line indicates onset of cue presentation. While SWR events around cue onset were excluded in the following analysis. **e**, *Top*, a single-component TCA model fit to cue-CA3SC peri-SWR suppression over individual SWR events in each mouse ($n=17.76 \pm 1.39$ SWRs in each inter-trial-interval, mean \pm s.e.m. from 6 mice). *Left*, sorted ROI factor for each CA3SC. *Middle-left and middle-right*, temporal factor and trial factor (mean \pm s.e.m.). Trial factor TC scores were extracted in each SWR event and indexed to inter-trial intervals where these events were observed. *Right*, ROI-averaged $\Delta F/F_0$ indexed to inter-trial intervals (mean \pm s.e.m.). *Bottom*, a single-component TCA model of other-CA3SC peri-SWR activation over SWR events. The trial factor TC score is likely to correlate with synchronicity of all ROIs to SWR events. To more clearly isolate any trend in the trial factor other-CA3SCs, we included only other-CA3SCs with strong responses to SWRs for model fitting (exceeding average amplitude shown in Fig. 2c). The error bars indicate s.e.m. **f**, *Left*, extracted PCs representing CA3SC suppression and activation driven by SWRs during different task conditions in each class of CA3SCs (cue, place, and other). K-means clustering was applied to the PCA loadings to classify ROIs as "activated" or "suppressed". *Middle*, percentages of variance accounted for by each PC (up to 20). *Right*, mean peri-SWR response of classified CA3SCs (dark green: activated, magenta: suppressed).

[Source data](#)

Extended Data Fig. 7 Reproducibility of SWR suppression of cue-responsive CA3SCs with automated and manual sparse ROI segmentation and cell body ROI.

a, *Left*, representative time-averaged FOV from one representative mouse ($n=6$ in this study). *Right*, zoomed-in patches of FOV on left (yellow dashed square) showing original (default Suite2p performance), sparse automated, and sparse manual segmentation ([Methods](#)). ROIs overlaid onto FOV in random transparent colors at top with ROI outlines plotted below. **b**, Cumulative histogram of Euclidean distance between each ROI and its nearest neighboring ROI (black: $n=7055$ original ROIs; pink: $n=2945$ sparse automated ROIs). Median values \pm s.d. are shown in bar plot (inset). **c**, Density of surrounding ROIs calculated for each ROI as a function of distance. **d–g**, Analysis of automatically-detected CA3SC ROIs including mean cue responses \pm s.e.m. of cue-CA3SCs during RC-NS (d), mean peri-SWR response \pm s.e.m. of all cue-CA3SCs (Cue, red) and non-cue CA3SCs (Other, black) recorded (e), fraction of cue-responsive and non-cue CA3SCs (910 cue-CA3SC ROIs from 6 mice) (f), and fraction of SWR recruitment of CA3SCs during RC-NS and POST (Other: $n=1,945$; Cue: $n=910$ ROIs from 6 mice. two-sided Fisher's exact test with Bonferroni correction. $P=2.6\times 10^{-24}$ for Cue-vs-All and Cue-vs-Other) (g). **h–k**, Analysis of manually-curated sparse CA3SC ROIs plotted as in (d-g) (Other: $n=1,083$; Cue: $n=255$ ROIs from 6 mice. Fisher's exact test with Bonferroni correction. $P=1.2\times 10^{-7}$ for Cue-vs-All and Cue-vs-Other). Fraction of hand-drawn cue-responsive and non-cue CA3SC ROIs. **k**, Fraction of SWR recruitment. **l**, Schematic of window implantation above CA3 stratum pyramidale and CA3 pyramidal cells injected with rAAV2/1:Syn(GCaMP7f)^{Cre} in Grk4-Cre mice. A 4-channel linear silicon probe was implanted in contralateral CA1. **m**, Example motion-corrected, time-averaged two photon imaging FOV from one representative mouse ($n=6$ in this study). Scale bar=100 μm **n**, *Left*, two example Suite2p-detected ROIs from (b). *Right*, representative GCaMP calcium signals ($\Delta F/F_0$) extracted from CA3 pyramidal cell ROIs. **o**, Mean cue responses \pm s.e.m. of identified cue-CA3 cells during RC-NS in $n=6$ mice. **p**, Mean peri-SWR response \pm s.e.m. of cue- and non-cue-CA3 cells recorded during RC-NS. **q**, Fraction of cue-responsive and non-cue CA3 cells in RC-

NS. **r**, Fraction of SWR recruitment during RC-NS and POST (Other: n=238; Cue: n=76 ROIs from 6 mice. two-sided Fisher's exact test with Bonferroni correction. P=2.7x10⁻⁹ for Cue-vs-All and Cue-vs-Other).

[Source data](#)

Extended Data Fig. 8 Spatial coding profiles of CA3SCs in FC-S.

a, Design of session block on Days 5 and 6 ([Methods](#)). **b**, *Left*, diagram of CA3SC ROI classification scheme on Days 5 and 6. *Right*, Fraction of active CA3SCs on Days 5 and Day 6. Each color indicates classified CA3SCs (Cue: n = 661, Place: n = 1,024, Other: n = 459 from 3 mice). **c**, Mean responses during RC-NS. *Middle*, Heatmap of place fields of individual cue-CA3SCs during FC-S with ROIs sorted according to location of peak activity to illustrate place coding (*bottom*). **d**, Cumulative histograms of cue-related and spatial information (bits). Mean±s.d. information values are shown in the insets. Cue- (red) and place- (blue) CA3SCs showed higher place information than other- (gray) CA3SCs on Day 5 (Cue: n=470, Place: n=1,170, Other: n = 418 from 3 mice). One-way ANOVA with post-hoc Tukey's test. Main effect of ROI-type for both FC-S and RC-NS: P =4.6x10⁻⁹) and on Day 6 (Cue: n=661, Place: n=1,024, Other: n=459 from 3 mice. One-way ANOVA with post-hoc Tukey's test. Main effect of ROI-type for both FC-S and RC-NS: P=1.5x10⁻⁵). **e**, peri-SWR activity in PRE and POST (*left*, mean±s.e.m.) summary of peri-SWR peak activity (*right*, mean±s.d.) for cue-, place- and other-CA3SCs on Day 6 (Two-way ANOVA with post-hoc Tukey's test, main effect of RC-vs-FC: P=8.7x10⁻¹²; main effect of ROI-type: P=2.3x10⁻¹²; interaction: P=7.9x10⁻¹⁰). **f**, Cumulative distribution plots of peri-SWR, z-scored $\Delta F/F_0$ for all CA3SCs shown in (e).

[Source data](#)

Extended Data Fig. 9 SWR-associated high-synchrony events containing recurring CA3SC assemblies.

a, Example SWR-associated high-synchrony event (HSE). *Top*, heatmap showing $\Delta F/F_0$ of individual CA3SC ROIs with place fields. *Bottom*, mean $\Delta F/F_0$ and the first principal component (PC) of these ROIs (*bottom*). 0 point indicates onset of an SWR event. Red line indicates duration of the detected HSE with threshold ([Methods](#)). **b**, Distribution of HSE peak latencies and co-occurrence rate (inset) with SWRs ($n=500$ HSEs from 4 mice). Red center line in the box plot indicates the median, bottom and top edges of the box denote 25th and 75th percentiles. The whiskers indicate maxima and minima. **c**, Distribution of individual CA3SC latencies within SWR-associated HSEs. **d**, *Left*, normalized cross-correlation (CC) at zero lag between all pairs of CA3SC ROIs replotted from Fig. [4b](#). Purple dashed line denotes result including all detected ROIs for comparison. *Middle*, distribution of correlation coefficients between ROIs putatively from the same cells (black: $n=139$ ROIs, all together; each color: $n = 34.75 \pm 13.84$ clustered ROIs, Mean \pm s.e.m, individual mice). Red center line in the box plot indicates the median, bottom and top edges of the box denote 25th and 75th percentiles. The whiskers indicate maxima and minima. *Right*, fraction of clustered ROIs putatively originating from common cells ($n=3315$ ROIs from 4 mice, [Methods](#)). Red center line in the box plot indicates the median, bottom and top edges of the box denote 25th and 75th percentiles. The whiskers indicate maxima and minima. **e**, *Left*, similarity matrix of all HSEs containing recurring assemblies in a representative session. *Middle*, assembly recruitment for each identified HSE cluster. Colored lines and traces with ID numbers correspond to identified HSE clusters and recurring assemblies in the similarity map, respectively. *Right*, example plots showing participation of 2 recurring assemblies across identified HSE clusters in Fig. [4d](#) (activation quantified as normalized covariance; clusters represented by different colors). **f**, Boxplot of the number of CA3SC ROIs in each identified recurring assembly (median with 25th and 75th percentiles from $n=4$ mice). The whiskers indicate maxima and minima. **g**, Distribution of correlation coefficients between place fields within the recurring assemblies (magenta, mean with 25th and 75th percentiles), and all cue-place CA3SC ROIs in a field of view (black, mean of 4 mice). Pairs with coefficients >0.7 are considered as ‘Near’ fields, and coefficients <0 are considered as ‘Other’ field pairs. *Middle*, boxplots of correlation coefficients between the distribution within assemblies ($n=28$ HSE clusters

from 4 mice, median with 25th and 75th percentiles). *Right*, fractions of ‘Other’ fields and ‘Near’ fields pairs within assemblies. *Right*, scatter plot based on these fractions (magenta circles for each assembly, black cross for mean value of all ROIs from each mouse). Notably, the lack of bias for recurring assemblies to comprise CA3SCs representing similar spatial locations suggests that individual structured replay sequences of CA3SC activity represent past experience in a relatively unbiased manner. Red center lines in the box plots indicate the median, bottom and top edges of the boxes denote 25th and 75th percentiles. The whiskers indicate maxima and minima, outliers excluded.

Source data

Extended Data Fig. 10 Validation of convNMF on simulated CA3 replay sequences.

a, Randomly, recurrently-connected spiking network model (pyramidal cells, blue triangles; interneurons, red circles) of CA3 used to generate synthetic replay sequences ([Methods](#)). **b**, Replay occurs spontaneously at rest in the simulated CA3 network (n=8000 simulated pyramidal cells, 150 inhibitory interneurons): 10 s simulation epoch, sampled at 10 kHz. Cell x time raster. **c**, Simulated spike rasters from (b) are re-binned to the calcium sampling rate (30 Hz). **d**, Re-binned spikes are then convolved with a calcium kernel ($\tau=0.65$ s), with added white Gaussian noise to simulate experimental conditions. **e**, Schematic of convNMF technique used to detect sequential reactivation events, vs standard NMF ([Methods](#)). The convNMF approximates the cell x time raster as the sum of K matrices, each which can be decomposed into the convolution of a rank-1 temporal component with cell x lag matrix. Unlike standard NMF, convNMF extracts recurring, localized sequential activity patterns. **f**, High-fidelity reconstruction of the simulated replay/calcium dataset in (b) is achieved using the convNMF-identified replay ensembles (left panels) and temporal factors (top traces). Each cell-component (left panel) represents a recurring sequential pattern of cell activations detected in the network. **g**, Slopes and intercepts were calculated for inferred replay events which were then compared to ground truth. ConvNMF accurately identifies replay events and extracted sequences of replaying cells. *Left*, scatterplot of true (x-axis)

vs inferred (y-axis) replay slopes (a.u.). Note that this simulation also demonstrates that forward and reverse replay events are reliably differentiated by convNMF. *Right*, distribution of true (x-axis) vs inferred (y-axis) replay intercepts. **h**, ROC curve of sensitivity/specificity of replay event detection by convNMF. Specificity is prioritized over sensitivity. 20–60% of “true” replay events were detected at an 80% specificity threshold.

Supplementary information

[Reporting Summary](#)

Source data

[Source Data Fig. 1](#)

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

[Source Data Extended Data Fig. 1](#)

[Source Data Extended Data Fig. 2](#)

[Source Data Extended Data Fig. 3](#)

[Source Data Extended Data Fig. 4](#)

[Source Data Extended Data Fig. 5](#)

[Source Data Extended Data Fig. 6](#)

[Source Data Extended Data Fig. 7](#)

[Source Data Extended Data Fig. 8](#)

[Source Data Extended Data Fig. 9](#)

[Source Data Extended Data Fig. 10](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Terada, S., Geiller, T., Liao, Z. *et al.* Adaptive stimulus selection for consolidation in the hippocampus. *Nature* **601**, 240–244 (2022).
<https://doi.org/10.1038/s41586-021-04118-6>

- Received: 03 February 2021
- Accepted: 08 October 2021
- Published: 08 December 2021
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04118-6>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04118-6>

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 15 December 2021](#)

NLRs guard metabolism to coordinate pattern- and effector-triggered immunity

- [Keran Zhai](#)^{1,na1},
- [Di Liang](#)^{1,2,na1},
- [Helin Li](#)^{1,3},
- [Fangyuan Jiao](#)^{1,4},
- [Bingxiao Yan](#)^{1,2},
- [Jing Liu](#)^{1,2},
- [Ziyao Lei](#)^{1,2},
- [Li Huang](#)^{1,2},
- [Xiangyu Gong](#)^{1,2},
- [Xin Wang](#)¹,
- [Jiashun Miao](#) [ORCID: orcid.org/0000-0002-5227-5431](#)¹,
- [Yichuan Wang](#)⁵,
- [Ji-Yun Liu](#)¹,
- [Lin Zhang](#)⁶,
- [Ertao Wang](#) [ORCID: orcid.org/0000-0002-5178-3530](#)¹,
- [Yiwen Deng](#)¹,
- [Chi-Kuang Wen](#) [ORCID: orcid.org/0000-0002-5345-0756](#)¹,
- [Hongwei Guo](#)⁵,
- [Bin Han](#)¹ &
- [Zuhua He](#) [ORCID: orcid.org/0000-0002-6098-7893](#)^{1,4}

[Nature](#) volume 601, pages 245–251 (2022)

- 10k Accesses
- 1 Citations
- 72 Altmetric
- [Metrics details](#)

Subjects

- [Biotic](#)
- [Effectors in plant pathology](#)

Abstract

Pattern-triggered immunity (PTI) and effector-triggered immunity (ETI) in plants enable them to respond to pathogens by activating the production of defence metabolites that orchestrate immune responses^{1,2,3,4}. How the production of defence metabolites is promoted by immune receptors and coordinated with broad-spectrum resistance remains elusive. Here we identify the deubiquitinase PICI1 as an immunity hub for PTI and ETI in rice (*Oryza sativa*). PICI1 deubiquitinates and stabilizes methionine synthetases to activate methionine-mediated immunity principally through biosynthesis of the phytohormone ethylene. PICI1 is targeted for degradation by blast fungal effectors, including AvrPi9, to dampen PTI. Nucleotide-binding domain, leucine-rich-repeat-containing receptors (NLRs) in the plant immune system, such as PigmR, protect PICI1 from effector-mediated degradation to reboot the methionine–ethylene cascade. Natural variation in the *PICI1* gene contributes to divergence in basal blast resistance between the rice subspecies *indica* and *japonica*. Thus, NLRs govern an arms race with effectors, using a competitive mode that hinges on a critical defence metabolic pathway to synchronize PTI with ETI and ensure broad-spectrum resistance.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Fig. 1: PICI1 interacts with PigmR and regulates PigmR-mediated blast resistance and PTI responses.



Fig. 2: PICI1 stabilizes OsMETS through deubiquitination.

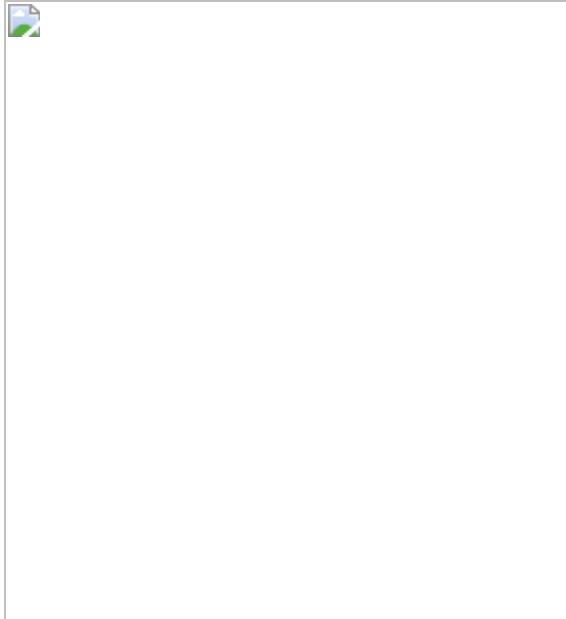


Fig. 3: Deubiquitination-mediated accumulation of OsMETS enhances blast resistance through promoting Met–ethylene biosynthesis.

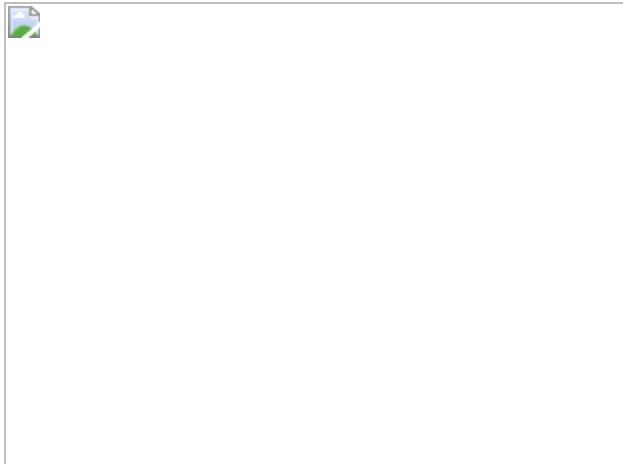


Fig. 4: PigmR protects PICI1 from degradation by AvrPi9, and natural variation in *PICI1* contributes to subspecies immune divergence.



Data availability

All data are available within this Article and its Supplementary Information. Original gel blots are shown in Supplementary Fig. 1. Original data points in graphs are shown in the Source Data files. Statistical analyses of this study are provided in Supplementary Table 4. The sequences of PICI1, OsMETS1 and OsMETS2 have been deposited and made publicly available in GenBank with accession codes [MT920667](#), [MT920668](#) and [MT920669](#), respectively. Protein structure models of Protein Data Bank (PDB) ID [2WP7](#) and PDB ID [1U1U](#) were obtained from the PDB. [Source data](#) are provided with this paper.

References

1. 1.

Jones, J. D. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).

2. 2.

Couto, D. & Zipfel, C. Regulation of pattern recognition receptor signalling in plants. *Nat. Rev. Immunol.* **16**, 537–552 (2016).

3. 3.

Xin, X.-F. Kvitko, B. & He, S. Y. *Pseudomonas syringae*: what it takes to be a pathogen. *Nat. Rev. Microbiol.* **16**, 316–328 (2018).

4. 4.

Kourelis, J. & van der Hoorn, R. A. L. Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *Plant Cell* **30**, 285–299 (2018).

5. 5.

Li, W., Deng, Y., Ning, Y., He, Z. & Wang, G. L. Exploiting broad-spectrum disease resistance in crops: from molecular dissection to breeding. *Annu. Rev. Plant Biol.* **71**, 575–603 (2020).

6. 6.

Ngou, B. P. M., Ahn, H.-K., Ding, P. & Jones, J. D. G. Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature* **592**, 110–115 (2021).

7. 7.

Yuan, M. et al. Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature* **592**, 105–109 (2021).

8. 8.

Bürger, M. & Chory, J. Stressed out about hormones: how plants orchestrate immunity. *Cell Host Microbe* **26**, 163–172 (2019).

9. 9.

Broekaert, W. F., Delauré Sl, M. F. C., Bolle, D., De Bolle, Mf, B. P. A. & Cammue, B. P. The role of ethylene in host-pathogen interactions. *Annu. Rev. Phytopathol.* **44**, 393–416 (2006).

10. 10.

Washington, E. J. et al. *Pseudomonas syringae* type III effector HopAF1 suppresses plant immunity by targeting methionine recycling to block ethylene induction. *Proc. Natl Acad. Sci. USA* **113**, E3577–E3586 (2016).

11. 11.

Zhao, S. et al. A viral protein promotes host SAMS1 activity and ethylene production for the benefit of virus infection. *eLife* **6**, e27529 (2017).

12. 12.

Dean, R. et al. The top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* **13**, 414–430 (2012).

13. 13.

Mosquera, G., Giraldo, M. C., Khang, C. H., Coughlan, S. & Valent, B. Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1–4 as biotrophy-associated secreted proteins in rice blast disease. *Plant Cell* **21**, 1273–1290 (2009).

14. 14.

Mentlak, T. A. et al. Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. *Plant Cell* **24**, 322–335 (2012).

15. 15.

Orbach, M. J., Farrall, L., Sweigard, J. A., Chumley, F. G. & Valent, B. A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta*. *Plant Cell* **12**, 2019–2032 (2000).

16. 16.

Kanzaki, H. et al. Arms race co-evolution of *Magnaporthe oryzae* *AVR-Pik* and rice *Pik* genes driven by their physical interactions. *Plant J.* **72**, 894–907 (2012).

17. 17.

Sweigard, J. A. et al. Identification, cloning, and characterization of PWL2, a gene for host species-specificity in the rice blast fungus. *Plant Cell* **7**, 1221–1233 (1995).

18. 18.

Wu, J. et al. Comparative genomics identifies the *Magnaporthe oryzae* avirulence effector AvrPi9 that triggers Pi9-mediated blast resistance in rice. *New Phytol.* **206**, 1463–1475 (2015).

19. 19.

Yoshida, K. et al. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* **21**, 1573–1591 (2009).

20. 20.

Li, W. et al. The *Magnaporthe oryzae* avirulence gene *AvrPiz-t* encodes a predicted secreted protein that triggers the immunity in rice mediated by the blast resistance gene *Piz-t*. *Mol. Plant Microbe Interact.* **22**, 411–420 (2009).

21. 21.

Deng, Y. et al. Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. *Science* **355**, 962–965 (2017).

22. 22.

Xie, Z. et al. A nucleotide-binding site-leucine-rich repeat receptor pair confers broad-spectrum disease resistance through physical association in rice. *Phil. Trans. R. Soc. B* **374**, 20180308 (2019).

23. 23.

Zhou, B. et al. The eight amino-acid differences within three leucine-rich repeats between *Pi2* and *Piz-t* resistance proteins determine the resistance specificity to *Magnaporthe grisea*. *Mol. Plant Microbe Interact.* **19**, 1216–1228 (2006).

24. 24.

Xie, X. et al. PPPDE1 promotes hepatocellular carcinoma development by negatively regulate p53 and apoptosis. *Apoptosis* **24**, 135–144 (2019).

25. 25.

Shimizu, T. et al. Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. *Plant J.* **64**, 204–214 (2010).

26. 26.

Suh, H. Y. et al. Crystal structure of DeSI-1, a novel deSUMOylase belonging to a putative isopeptidase superfamily. *Proteins* **80**, 2099–

2104 (2012).

27. 27.

Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229 (2012).

28. 28.

Chen, X. L. et al. Proteomic analysis of ubiquitinated proteins in rice (*Oryza sativa*) after treatment with pathogen-associated molecular pattern (PAMP) elicitors. *Front. Plant Sci.* **9**, 1064 (2018).

29. 29.

Iwai, T., Miyasaka, A., Seo, S. & Ohashi, Y. Contribution of ethylene biosynthesis for resistance to blast fungus infection in young rice plants. *Plant Physiol.* **142**, 1202–1215 (2006).

30. 30.

Singh, R. et al. Magnaporthe oryzae effector AVR-Pii helps to establish compatibility by inhibition of the rice NADP-malic enzyme resulting in disruption of oxidative burst and host innate immunity. *Mol. Cells* **39**, 426–438 (2016).

31. 31.

De la Concepcion, J. C. et al. Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen. *Nat. Plants* **4**, 576–585 (2018).

32. 32.

Maqbool, A. et al. Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife* **4**, 08709 (2015).

33. 33.

Wang, W. et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**, 43–49 (2018).

34. 34.

Huang, X. et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* **44**, 32–39 (2011).

35. 35.

Hu, C. et al. Metabolic variation between *japonica* and *indica* rice cultivars as revealed by non-targeted metabolomics. *Sci. Rep.* **4**, 5067 (2014).

36. 36.

Liao, J. et al. Pathogen effectors and plant immunity determine specialization of the blast fungus to rice subspecies. *eLife* **5**, e19377 (2016).

37. 37.

Axtell, M. J. & Staskawicz, B. J. Initiation of *RPS2*-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4. *Cell* **112**, 369–377 (2003).

38. 38.

Mackey, D., Holt, B. F. III, Wiig, A. & Dangl, J. L. RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* **108**, 743–754 (2002).

39. 39.

He, J. et al. A LysM receptor heteromer mediates perception of arbuscular mycorrhizal symbiotic signal in rice. *Mol. Plant* **12**, 1561–1576 (2019).

40. 40.

Ma, X. et al. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot Plants. *Mol. Plant* **8**, 1274–1284 (2015).

41. 41.

Zhai, K. et al. RRM transcription factors interact with NLRs and regulate broad-spectrum blast resistance in rice. *Mol. Cell* **74**, 996–1009 (2019).

42. 42.

Park, C.-H. et al. The *Magnaporthe oryzae* effector AvrPiz-t targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern-triggered immunity in rice. *Plant Cell* **24**, 4748–4762 (2012).

Acknowledgements

We thank J. Zhou and J. Li for critical reading and discussion; X. Wei and G. Zhang for providing the rice germplasm and CSSL; J. Zhang for help with ethylene measurement; S. Wang for help with Met measurement; B. Liu, C. Wang and Z. Gu for help with bioinformatic analysis; and X. Fu for sharing HA-Ub plasmids. This work was supported by grants from the National Natural Science Foundation of China (32088102, 31720103913 and U20A2021), the Chinese Academy of Sciences (XDB27040201 and XDA24010304), the National Key Research and Development Program of China (2016YFD0100600), the Science and Technology Commission of Shanghai Municipality (19391900300), the Project of Special Funding for Crop Science Discipline Development of Yangzhou University (yzuxk202006) and the National Key Laboratory of Plant Molecular Genetics. K.Z. was supported by the National Postdoctoral Program for Innovative Talents (BX20190346) and by the Shanghai Postdoctoral Excellence Program.

Author information

Author notes

1. These authors contributed equally: Keran Zhai, Di Liang

Affiliations

1. National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China

Keran Zhai, Di Liang, Helin Li, Fangyuan Jiao, Bingxiao Yan, Jing Liu, Ziyao Lei, Li Huang, Xiangyu Gong, Xin Wang, Jiashun Miao, Ji-Yun Liu, Ertao Wang, Yiwen Deng, Chi-Kuang Wen, Bin Han & Zuhua He

2. University of the Chinese Academy of Sciences, Beijing, China

Di Liang, Bingxiao Yan, Jing Liu, Ziyao Lei, Li Huang & Xiangyu Gong

3. School of Life Sciences, Henan University, Kaifeng, China

Helin Li

4. School of Life Science and Technology, Shanghai Tech University, Shanghai, China

Fangyuan Jiao & Zuhua He

5. Department of Biology, Institute of Plant and Food Science, Southern University of Science and Technology, Shenzhen, China

Yichuan Wang & Hongwei Guo

6. Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding, Yangzhou University, Yangzhou, China

Lin Zhang

Contributions

K.Z. and Z.H. conceived and designed the experiments. K.Z. and D.L. performed most experiments, including the Y2H, SLC, BiFC, co-IP, subcellular localization, cell-free degradation assay, deubiquitination assay, ROS detection and western blots. K.Z., H.L. and Y.W. performed Met and ethylene measurement. K.Z., H.L. and F.J. performed RNA analysis. K.Z., D.L., B.Y. and X.G. performed pathoassays. D.L. and H.L. performed agronomic traits analysis. K.Z., J.L., Z.L., L.H., X.W. and J.-Y.L. generated material used in this study. J.M. performed the domestication analysis. H.G. and B.H. supervised Y.W. and J.M., respectively. L.Z., E.W., Y.D., C.-K.W., H.G. and B.H. provided theoretical contributions to the project. K.Z., D.L. and Z.H. analysed the data and wrote the paper.

Corresponding author

Correspondence to [Zuhua He](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Nicholas Talbot and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Screening of PICIs and PigmR–PICI1 interaction analysis.

a, The workflow for identifying the proteins involved in both PTI and ETI. Tandem Mass Tag (TMT)-based proteomic analysis was performed to investigate the global changes of rice protein profiling with chitin treatment, the blue circle denotes the chitin-induced 413 candidate proteins. Total 11 PigmR-CC interacting proteins were identified through yeast-two-hybrid screen and shown with the pink circle. The overlap between the blue and pink circles indicates 3 PICI candidates. **b**, Protein alignment of PICI1 and its homologs in *A. thaliana*, *S. cerevisiae*, *C. elegans*, *M. musculus* and *H. sapiens* (NP_194296.1, NP_594707.2, NP_741592.1, NP_077244.1 and NP_057160.2, respectively) with MegAlign software. The conserved PPPDE region is underlined in red and the conserved site Cys115 residue is marked with red arrow. **c**, Immunodetection of protein expression in yeast. pDEST22 (EV), pDEST22-PICI1 (1) and pDEST22-Os07g16130 (2) detected using anti-GAL4-AD antibody. pDEST32-PigmR (BD-PigmR), pDEST32-PigmR-CC (BD-PigmR-CC), pDEST32-PigmR-NBS (BD-NBS) and pDEST32-PigmR-LRR (BD-LRR) detected using anti-GAL4-BD antibody. The asterisks indicate the target proteins. **d**, Protein alignment of PICI1 and its homologs in rice. **e**, PigmR and PigmR-CC specifically interact with PICI1 but not its homologs in Y2H assay. **f**, Immunodetection of protein levels in yeast. Upper panel (AD), pDEST22 (lane 1), pDEST22-PICI1 (lane 2), -Os02g43840 (lane 3), -Os03g01130 (lane 4), -Os03g10200 (lane 5), -Os04g46290 (lane 6), -Os06g01780 (lane 7), -Os06g08360 (lane 8), -Os06g36490 (lane 9), detected using anti-GAL4-AD antibody; middle panel (BD), pDEST32-PigmR-CC, detected using anti-GAL4-BD antibody. The asterisks indicate the target proteins. **g**, SLC assay of PigmR-PICI1 interaction in *N. benthamiana*. PIBP1 and Os07g16130 served as positive and negative control, respectively. Fluorescence signal intensity is indicated. **h**, Western blot confirming expression of proteins in *N. benthamiana*. Note that cLuc-PICI1 was detected using anti-PICI1 antibody. Short, long = short or long exposure. **i**, PICI1-YFP was co-expressed with various known organelle markers as indicated in rice

protoplasts. mCherry = cytoplasmic marker, NLS-RFP = nucleus marker, ST-mCherry = trans-Golgi cisternae marker, SYP61-mCherry = trans-Golgi network/early endosome (TGN/EE) marker, ER-mCherry = endoplasmic reticulum marker, OsRac1-mCherry = cell member marker, Chl Auto = Chloroplast autofluorescence. **j**, Western blot confirming expression of PICI1-YFP without free YFP in rice protoplast. **k**, Localization of PICI1-GFP in representative root cells of stable transgenic plants. DAPI staining indicates the nucleus. **l**, Bimolecular fluorescence complementation (BiFC) verifies the PigmR-CC/PICI1 interaction in rice protoplasts. Pish-CC served as a negative control. Ponceau S staining (**c**, **f**) or Actin (**h**, **j**) served as loading control. Scale bars, 5 μ m (**i**, **l**) or 10 μ m (**k**). Experiments were repeated twice (**c**, **f**, **h**) or three times with similar results (**e**, **g**, **i–l**).

Extended Data Fig. 2 PICI1 is involved in PigmR-, Pizt- and Pi9- but not Pish-mediated blast resistance.

a, i, l, o, Schematic of two independent *PICI1* knockout lines in NIL-*Pigm* (**a**), ZH11 (*Pizt*-containing) (**i**), Ky-*Pi9* (*Pi9*-containing) (**l**) and NIPB (*Pish*-containing) (**o**), respectively. **b, j, m, p**, Protein analysis of PICI1 in the corresponding transgenic plants using an anti-PICI1 antibody. The asterisk indicates the endogenous PICI1 and the band around 63 kDa indicates the PICI1-GFP fusion protein (**b**, **p**). **c**, Blast resistance of NIL-*Pigm*, *PICI1*-KO/NIL-*Pigm* and *PICI1*-OE/NIL-*Pigm* lines at 7dpi with punch inoculation inoculation (TH12). NIPB served as a susceptible control. **d**, Y2H assay of interactions between PICI1 and the CC domains of PigmR, Pi9, Pizt and Pish. EV, empty vector. **e**, Immunodetection of protein levels in yeast. Upper panel (AD), pDEST22-PICI1, detected using anti-GAL4-AD antibody; middle panel (BD), PigmR-CC, Pi9-CC, Pizt-CC and Pish-CC, detected using anti-GAL4-BD antibody. **f**, SLC confirmation of the PICI1-Pizt-CC and PICI1/Pi9-CC interactions in *N. benthamiana*. PigmR-CC and Pish-CC served as positive and negative control, respectively. Fluorescence signal intensity is indicated. **g**, Relative luciferase activity of protein-protein interactions was measured. One-way ANOVA with Tukey's test (mean \pm s.d.; $n = 5$, biologically independent samples). Different letters indicate significant difference at $P < 0.05$. **h**, Western blot analysis confirming the expression level of each protein in *N. benthamiana*. **k, n, q**, Punch inoculation of *PICI1*-KO transgenic plants in ZH11 (**k**), Ky-*Pi9* (**n**)

and NIPB (**q**). Blast resistance of WT and two representative transgenic lines, 5 dpi with avirulent strain YN2 (**k, q**) or Guy11 (**n**). TH12 served as a susceptible control (**q**). **r**, Transient assay of cell death in rice protoplast. Relative luciferase activity was measured after transformation with *avrPizt*-mCherry in rice protoplasts derived from ZH11 or *PICII*-KO/ZH11. Empty vector (EV) was used as a negative control. Two-tailed Student's *t*-test (mean ± s.d.; *n* = 9, biologically independent samples). **s**, Western blot analysis confirming expression of proteins in rice protoplasts. Actin (**b, h, j, m, p, s**) or Ponceau S staining (**e**) served as loading control. The asterisks indicate the target proteins. For **c, k, n, q**, data were analysed by two-tailed Student's *t*-test. Fungal growth (mean ± s.d.; *n* = 3, biologically independent samples), lesion lengths or areas (mean ± s.d.; *n* = numbers of biologically independent samples in the graphs). Scale bars, 1 cm. Asterisks represent significant difference (**P* < 0.05, ***P* < 0.01) (**c, k, n, q, r**). Exact *P* values are provided in Supplementary Table 4 (**c, g, k, n, q, r**). Experiments were repeated twice (**b, e, h, j, m, p**) or three times with similar results (**c, d, f, g, k, n, q, r, s**).

[Source data](#)

[Extended Data Fig. 3 PICII is involved in chitin- and flg22-induced PTI responses.](#)

a, g, p, Induction of the PTI-related defence genes *OsPAL1* and *KS4* at 1 h after chitin (**a, g**) or flg22 (**p**) incubation in WT, *PICII*-KO and *PICII*-OE lines in NIL-*Pigm* (**a, p**) or NIPB (**g**) background. Two-tailed Student's *t*-test (mean ± s.d.; *n* = 3, biologically independent samples). **b, c**, Induction of *PICII* in NIL-*Pigm* upon chitin/flg22 treatment (**b**) and during *M. oryzae* TM21 infection (**c**) with water as control. **d, h**, *PICII* was induced upon *M. oryzae* TM21 infection (**d**) and chitin treatment (**h**) with water as control. The asterisk indicates endogenous *PICII* (**d**). The relative *PICII* protein abundance was indicated. **e**, Induction of PTI-related defence genes *KS4* and *OsPAL1* in NIL-*Pigm* leaves during *M. oryzae* TM21 infection. **f, o**, Chitin- (**f**) or flg22- (**o**) induced ROS burst in WT, *PICII*-KO and *PICII*-OE lines in NIPB (**f**) or NIL-*Pigm* (**o**) background. Data are mean ± s.d. (*n* = 9, biologically independent samples). **i**, Disease resistance of *PICII* transgenic plants in NIPB. **j, l**, Schematic of two independent

PICII/OsCERK1 double knockout (**j**) and *PICII-OE/OsCERK1-KO* (**l**) lines in NIPB. **k, m**, Protein levels of PICI1 (**k**) or PICI1-GFP (**m**) in *PICII/OsCERK1-KO* (**k**) or *PICII-OE/OsCERK1-KO* (**m**). **n**, Punch inoculation of *OsCERK1-KO*, *PICII/OsCERK1-KO* and *PICII-OE/OsCERK1-KO* transgenic lines. **q, r**, PICI1 was induced upon flg22 treatment (**q**) and bacterial *Xoo* (strain, PXO99^A) infection (**r**) with water as control. **s–u**, Disease resistance to bacterial *Xoo* in the *PICII* transgenic lines in NIL-*Pigm*. Triangles represent the end of bacterial infection sites. For **i, n, t, u**, data were analysed by two-tailed Student's *t*-test. Lesion lengths (mean ± s.d.; *n* = numbers of biologically independent samples in the graphs), fungal or bacterial growth (mean ± s.d.; *n* = 3, biologically independent samples). For **b, c, e**, data are mean ± s.d. (*n* = 3, biologically independent samples). The rice *ACTIN1* served as an internal control (**a–c, e, g, p**). Actin was detected as a loading control (**d, h, k, m, q, r**). Blast resistance of representative transgenic lines, 5 dpi with virulent strain TH12 (**i, n**). Scale bars, 1 cm (**i, n**) or 5 cm (**s**). Asterisks represent significant difference (**P* < 0.05, ***P* < 0.01) (**a, g, i, n, p, t, u**). Experiments were repeated twice (**k, m**) or three times with similar results (**a–i, n, o–u**).

Source data

Extended Data Fig. 4 PICI1 functions as a deubiquitinase, and interacts with and stabilizes OsMETS1.

a, His-PICI1 and His-PICI1-PPPDE displayed cleavage activity toward K48- and K63-linked but not linear ubiquitin in vitro. His alone served as a control. **b**, Homology model of PICI1-PPPDE domain (cyan) aligned with Desi-1 (PDB ID code 2WP7) (blue) using PyMOL. The putative catalytic dyad residues of H41 and C115 in PICI1 are shown in yellow and the catalytic residues of Desi-1 H38 and C108 are shown in magenta. The boxed section is magnified for close view. **c**, Induction of *KS4*, *PR10* and *PR5* in PICI1-YFP or PICI1^{C115S}-YFP overexpression protoplast after chitin treatment (1 hpi) with the empty vector control. Lowercase letters indicate statistical significance (*P* < 0.05). One-way ANOVA with Tukey's test. Data are mean ± s.d. (*n* = 3, biologically independent samples). Exact *P* values are provided in Supplementary Table 4. **d**, Ubiquitin conjugation in total protein extracts from 2-week-old NIL-*Pigm* and *PICII-KO/NIL-Pigm*,

after spraying TH12 or TM21 with water as control for 36 h. **e**, Morphological phenotype of mature plants of wild-type and *PICII*-KO/NIL-*Pigm*. No obvious change in morphology was observed in the *PICII*-KO plants. Scale bar, 10 cm. **f**, Venn diagram showing the number of total ubiquitinated proteins (blue), PICI1 interacting candidates identified in IP-MS (red), PTI-related ubiquitinated proteins (green) and overlapping proteins. **g**, Protein alignment of OsMETS1 and its homologs in sorghum, soybean, maize, wheat and Arabidopsis (XP_021301657.1, XP_003542326.1, PWZ52049.1, TraesCS4D02G012900.2 and XP_002871787.1) with MegAlign software. **h**, Ethylene biosynthesis pathway in plants. S-AdoMet, S-adenosylmethionine; Hcys, homocysteine; THF, tetrahydrofolate; SAH, S-adenosylhomocysteine; Ado, adenosine; ACS, 1-aminocyclopropane-1-carboxylate synthase; ACC, 1-aminocyclopropane-1-carboxylate. **i**, A SLC assay of OsMETS1 and PICI1 interaction in *N. benthamiana*. PICI2 served as a negative control. Fluorescence signal intensity is indicated. **j**, Western blot confirming protein expression in *N. benthamiana*. **k**, GST-PICI1 pulls down His-OsMETS1 in vitro. **l**, Subcellular localization of OsMETS1-GFP in OsMETS1-GFP/NIL-*Pigm* root. DAPI staining indicates the nucleus. Scale bars, 10 μ m. **m**, Immunodetection of OsMETS1-GFP in transgenic lines, with NIL-*Pigm* as a negative control. **n**, Subcellular localization of OsMETS1 and co-localization of PICI1-YFP/OsMETS1-mCherry in rice protoplasts. Scale bars, 5 μ m. **o**, Western blot confirming expression of OsMETS1-mCherry without free mCherry in rice protoplast. mCherry (lane 1), OsMETS1-mCherry (lane 2). **p**, Western blot confirming expression of OsMETS1-YFP without free YFP in rice protoplast. YFP (lane 1), OsMETS1-YFP (lane 2). **q**, Degradation of OsMETS1-Flag was proteasome-dependent and delayed by PICI1 in cell-free system. **r**, PICI1 stabilizes OsMETS1-Flag in rice plants. The lysates from WT, *PICII*-KO and *PICII*-OE plants were co-incubated with OsMETS1-Flag in the presence or absence of MG132. **s**, Peptide coverage and ubiquitination sites identified in OsMETS1 by tandem mass spectrometry after *M. oryzae* (TH12) or water treatment for 36 h. The various peptides identified by mass spectrometry are shown in blue; the lysine acceptor sites detected in red are numbered below with positions; the four specific deubiquitination sites are denoted with asterisks. **t**, Structure model for OsMETS1. Homology model of the OsMETS1 based on the crystal structure of Arabidopsis Met synthase

(PDB ID code 1U1U) and analysed by PyMOL. The four lysines identified as deubiquitination sites were shown in blue. The other lysine acceptor sites were shown in red. **u**, Cell-free degradation assays show the proteasome-dependent degradation of His-OsMETS1 and delayed degradation of His-OsMETS1 mutant variants. Purified recombinant His-OsMETS1 and its mutants were incubated with protein extracts from non-infected NIPB. His alone served as a negative control. **v**, The OsMETS1 mutants showed higher protein accumulation compared to wild-type OsMETS1 in protoplasts. The YFP tag protein was expressed as an internal control. The transcript levels were determined by semi-quantitative PCR. Actin was detected as a loading control (**d, j, m, o, p**). Ponceau S staining served as loading control (**q, r, u**). The protein abundance was quantified using ImageJ and indicated under lanes (**q, r, u**). Experiments were repeated twice (**j, m, o, p**) or three times with similar results (**a, c, d, i, k, l, n, q, r, u, v**).

Source data

Extended Data Fig. 5 PICI1 directly deubiquitinates and stabilizes OsMETS.

a, Induction of *PR5* and *PR1* in OsMETS1, OsMETS1^{K28R}, OsMETS1^{K196R}, OsMETS1^{K458R} and OsMETS1^{K740R} overexpression protoplasts after chitin treatment (1 hpi). Lowercase letters indicate statistical significance ($P < 0.05$). One-way ANOVA with Tukey's test. Data are mean \pm s.d. ($n = 3$, biologically independent samples). Exact P values are provided in Supplementary Table 4. **b**, Cell-free degradation assays show the delayed degradation of His-OsMETS1^{4KR} mutant variant, compared to His-OsMETS1^{K28R}. Purified recombinant fusion proteins were incubated with non-infected NIPB extracts. **c**, The OsMETS1^{4KR} mutant showed higher protein accumulation compared to wild-type OsMETS1 and OsMETS1^{K28R} in protoplasts. The YFP tag protein was expressed as an internal control. The transcript levels were determined by semi-quantitative PCR. **d**, Stability of the OsMETS1-Flag depends on His-PICI1 but not the His-PICI1^{C115S} in cell-free degradation assay. **e**, Deubiquitination analysis of ubiquitin-modified OsMETS1-GFP, immunoprecipitated from

OsMETS1-GFP/NIL-*Pigm* plants, by incubating with recombinant His-PICI1 or His-PICI1^{C115S} respectively. **f**, Tail-to-head organization of *OsMETS1* and *OsMETS2* genes in rice. **g**, Protein alignment of OsMETS1 and OsMETS2 with MegAlign software. **h**, Co-IP assay of PICI1 with OsMETS2 in rice. PICI2 served as a negative control. **i**, A SLC assay showed that both OsMETS1 and OsMETS2 interacted with PICI1 in *N. benthamiana*. PICI2 was used as a negative control. Fluorescence signal intensity is indicated. **j**, Western blot confirming protein expression in *N. benthamiana*. **k**, Cell-free degradation assay showed the proteasome-dependent degradation of OsMETS2-Flag in *N. benthamiana*. **l**, Cell-free degradation assay showed the delayed degradation of OsMETS2 in presence of recombinant His-PICI1. **m**, Co-IP assay of PICI1 with endogenous OsMETS in rice. The GFP negative control and PICI1-GFP were constitutively expressed in transgenic rice plants. anti-OsMETS, recognizing both OsMETS1 and OsMETS2. **n**, Cell-free degradation assay showed that endogenous OsMETS degradation was delayed by PICI1. **o**, Immunodetection of the recombinant His-OsMETS1 and His-OsMETS2 using the anti-OsMETS antibody. His alone served as a negative control. **p**, A schematic diagram of two independent *OsMETS1*-KO lines in NIL-*Pigm* background. **q**, Protein levels of OsMETS in NIL-*Pigm*, *OsMETS1*-GFP/NIL-*Pigm* and *OsMETS1*-KO/NIL-*Pigm* plants using the anti-OsMETS antibody. The band around 75 kDa indicates the endogenous OsMETS. **r**, Relative transcript levels of *OsMETS* in the wild-type, *PICI1*-KO and *PICI1*-OE lines in NIL-*Pigm* and NIPB background. Data are mean ± s.d. ($n = 3$, biologically independent samples). **s**, Analysis of endogenous OsMETS ubiquitination. OsMETS was immunoprecipitated from wild-type NIL-*Pigm* with anti-OsMETS or anti-IgG as negative control. Short, long = Short or long exposure. The protein abundance was quantified using ImageJ and indicated under lanes (**b**, **d**, **k**, **l**, **n**, **q**). Ponceau S staining (**b**, **d**, **k**, **l**, **n**, **o**) or Actin (**e**, **j**, **q**, **s**) was used as a control for equal loading. The rice *ACTIN1* served as an internal control (**a**, **r**). Experiments were repeated twice (**h**, **j**, **m**, **n**, **o**, **s**) or three times (**a–e**, **i**, **k**, **l**, **q**, **r**) with similar results.

[Source data](#)

[Extended Data Fig. 6 OsMETS functions in rice PTI and ETI.](#)

a, b, c, d, OsMETS is induced upon treatment of chitin (**a**) or flg22 (**b**), and infection of *Xoo* (PXO99^A) (**c**) or *M. oryzae* (TM21) (**d**). **e**, Protein accumulation of OsMETS in NIL-*Pigm*, *PICII*-KO/NIL-*Pigm* and *PICII*-GFP/NIL-*Pigm* in early infection stage (TM21). **f, g**, Decreased ubiquitination of OsMETS1-GFP in PTI (**f**) and ETI (**g**). The protein extracts from OsMETS1-GFP/NIL-*Pigm* plants incubated with chitin for 1 h (**f**) or infected with TH12 for 36 h (**g**), with water as control, was analysed. NIL-*Pigm* served as a negative control. **h**, A schematic diagram of two independent *OsMETS2*-KO lines in NIL-*Pigm*. **i**, Decreased protein levels of OsMETS in *OsMETS2*-KO/NIL-*Pigm* compared with NIL-*Pigm*. **j, l**, Basal blast resistance of NIL-*Pigm*, *OsMETS2*-KO/NIL-*Pigm* (**j**) and *OsMETS2*-GFP/NIL-*Pigm* (**l**) at 5 dpi with punch inoculation (TM21). **k**, Immunodetection of the OsMETS2-GFP in the two independent transgenic lines, with NIL-*Pigm* as a negative control. **m, n**, Disease resistance of WT, *OsMETS1*-KO (**m**), *OsMETS1*-OE (**m**) and *OsMETS2*-KO (**n**) were shown, 5 dpi with TH12, NIPB served as a susceptible control. **o**, Relative transcript levels of *OsMETS1*, *OsMETS2* and *OsMETS* in NIL-*Pigm* and two *OsMETS*-RNAi lines. **p**, Protein levels of OsMETS in NIL-*Pigm* and *OsMETS*-RNAi lines. **q**, Morphological phenotype of NIL-*Pigm* and *OsMETS*-RNAi/NIL-*Pigm* lines at mature stage in the paddy field. **r**, *OsMETS*-RNAi/NIL-*Pigm* plants significantly reduced plant height (upper panel), tiller number (middle panel), grain yield per plant (lower panel). **s**, Relative Met contents in WT and *OsMETS*-RNAi/NIL-*Pigm*. **t**, Increased Met accumulation in NIL-*Pigm* seedling shoots after root incubation with 25 mM Met for 48 h. **u**, Increased blast resistance after Met treatment in NIPB. **v**, Induction of the defence gene *KS4* at 48 h after root dipping with 25 mM Met solution or water in NIL-*Pigm*. **w, y**, Basal blast resistance of Met-treated *OsMETS1*-KO/NIL-*Pigm* at 5 dpi (**w**) or *PICII*-KO/NIL-*Pigm* at 7 dpi (**y**). Met solution (10 mM) or water was supplied from the roots for 48 h and plants were then inoculated with TM21. **x**, Relative Met contents in WT and *PICII*-KO/NIL-*Pigm* lines. For **j, l, m, n**, data were analysed by two-tailed Student's *t*-test. Lesion lengths (mean ± s.d.; *n* = numbers of biologically independent samples in the graphs), fungal growth (mean ± s.d.; *n* = 3, biologically independent samples). For **u, w, y**, one-way ANOVA with Tukey's test. Lesion areas (mean ± s.d.; *n* = numbers of biologically independent samples in the graphs), fungal growth (mean ± s.d.; *n* = 3, biologically independent samples). Lowercase letters indicate

statistical significance ($P < 0.05$). For **r**, **s**, **t**, **x**, data were analysed by two-tailed Student's *t*-test (mean \pm s.d.; n = numbers of biologically independent samples in the graphs). In **o**, **v**, two-tailed Student's *t*-test (mean \pm s.d.; n = 3, biologically independent samples). Asterisks indicate significant differences (* $P < 0.05$, ** $P < 0.01$) (**j**, **l-o**, **r-t**, **v**, **x**). Exact *P* values are provided in Supplementary Table 4 (**j**, **l-o**, **r-y**). Scale bars, 1 cm (**j**, **l-n**, **u**, **w**, **y**), 10 cm (**q**). Relative protein abundance was indicated below (**a-e**, **p**). Actin was detected as a loading control (**a-g**, **i**, **k**, **p**). The rice *ACTIN1* served as an internal control (**o**, **v**). Experiments were independently repeated twice (**i**, **k**, **p**, **r**) or three times (**a-g**, **j**, **l-o**, **s-y**) with similar results.

[Source data](#)

[**Extended Data Fig. 7 PICI1–OsMETS module confers blast resistance through regulating Met–ethylene biosynthesis.**](#)

a, High Met concentrations (10 and 25 mM) inhibit blast fungal growth on medium. The *M. oryzae* isolate TH12 was inoculated on the complete medium (CM) with Met supplement. The colony diameters were measured. One-way ANOVA with Tukey's test (mean \pm s.d.; n = 5, biologically independent samples). **b**, **c**, Ethylene contents in WT and transgenic lines of *OsMETS1* (*OsMETS1*-KO and *OsMETS1*-OE) (**b**) and *PICI1* (*PICI1*-KO and *PICI1*-OE) (**c**) infected by TM21 (2-week-old, 36 hpi). Two-tailed Student's *t*-test (mean \pm s.d.; n = 3, biologically independent experiments). **d**, Induction of *ERF63* and *ERF73* by ethylene treatment in NIL-*Pigm*. Two-week-old seedlings were placed into sealed containers with 10 or 100 ppm (μ L/L) ethylene, or air as control for 3 h to analyse gene expression. **e**, **f**, Suppression of *OsMETS1* (**e**) and *OsMETS2* (**f**) expression by ethylene treatment (100 ppm) in NIL-*Pigm*. **g**, Relative expression of *ERF63* at 1 d treatment with 10 μ M AVG or 20 μ M ACC or water. *ERF63* was chosen as a marker gene for AVG and ACC responses. Two-tailed Student's *t*-test (mean \pm s.d.; n = 3, biologically independent samples). **h**, Effects of AVG and ACC on basal blast resistance in NIPB. Two-week-old seedlings were pre-treated with 10 μ M AVG, or 20 μ M ACC or water for 1 d and then spray-inoculated with rice blast (TH12), 5 dpi. **i**, **j**, Basal blast resistance of ACC-treated *PICI1*-KO/NIL-*Pigm* (**i**) and *OsMETS*-

RNAi/NIL-*Pigm* (**j**) plants. ACC solution (20 µM) or water was supplied from the leaves and inoculated with TM21, 5 dpi. **k, l**, Induction of *OsACS2* in NIL-*Pigm* leaves at early stage by TM21 (**k**, PTI) or late stage by TH12 (**l**, ETI) infection. **m**, Enhanced ethylene production in NIL-*Pigm* after rice blast infection (TH12 or TM21), with water as a control for 36 h. One-way ANOVA with Tukey's test (mean ± s.d.; $n = 3$, biologically independent experiments). **n**, A schematic diagram of *OsACS2*-KO lines in NIL-*Pigm*. **o**, Basal blast resistance in NIL-*Pigm* and *OsACS2*-KO/NIL-*Pigm*, at 5 dpi with punch inoculation (TM21). **p**, Disease resistance of wild type and *OsACS2*-KO lines were shown, 5 dpi with TH12, NIPB served as a susceptible control. **q**, Met enhances rice blast resistance via ethylene biosynthesis. A Met solution (10 mM) or water was supplied from NIL-*Pigm* roots for 48 h and pre-treated with 10 µM AVG for 36 h before spray inoculation with rice blast (TM21). One-way ANOVA with Tukey's test. Lesion areas (mean ± s.d.; $n = 5$, biologically independent samples), fungal growth (mean ± s.d.; $n = 3$, biologically independent samples). **r**, Protein levels of PICI1 in NIL-*Pigm* (left panel) and NIPB (right panel) plants after inoculation with *M. oryzae* at the indicated time points. Actin was detected as a loading control. The protein abundance was quantified using ImageJ and indicated under lanes. **s**, RNA expression of *PICI1* was not significantly affected in late stage by blast infection (TH12 and TM21) in NIL-*Pigm*. For **d, e, f, k, l, s**, data are mean ± s.d. ($n = 3$, biologically independent samples). For **h, i, j, o, p**, data were analysed by two-tailed Student's *t*-test. Lesion lengths or areas (mean ± s.d.; $n = \text{numbers of biologically independent samples in the graphs}$), fungal growth (mean ± s.d.; $n = 3$, biologically independent samples). Lowercase letters indicate statistical significance ($P < 0.05$) (**a, m, q**). Asterisks indicate significant differences (* $P < 0.05$, ** $P < 0.01$) (**b, c, g–j, o, p**). Exact *P* values are provided in Supplementary Table 4 (**a–c, g–j, m, o–q**). The rice *ACTIN1* served as an internal control (**d–g, k, l, s**). Scale bars, 1 cm (**h–j, o–q**). Experiments were repeated twice (**d–g**) three times with similar results (**a–c, h–j, k–m, o–q, r, s**).

Source data

Extended Data Fig. 8 PigmR protects PICI1 from AvrPi9-mediated degradation.

a, His-PICI1 could be pulled down by seven effectors (AvrPWL2, AvrPi9, AvrPikC, AvrPia, AvrPizt, AvrPii and AvrPikD fused with MBP) in vitro. MBP alone served as a negative control. **b**, SLC verification of the interactions between PICI1 and the three blast fungal effectors, AvrPWL2, AvrPi9 and AvrPii in *N. benthamiana*. Note that AvrPikD/Pik-HMA and AvrPii/OsNADP-ME2 served as positive controls. **c**, Relative luciferase activity of protein-protein interactions was measured. Data are mean \pm s.d. ($n = 6$, biologically independent samples). **d**, Western blot confirming protein expression in *N. benthamiana*. Note that 1-12 represented PICI1/AvrBAS4, PICI1/AvrPWL2, PICI1/AvrPi9, PICI1/AvrPita, PICI1/AvrSlp1, PICI1/AvrPikC, PICI1/AvrPia, PICI1/AvrPizt, PICI1/AvrPii, OsNADP-ME2/PICI1, PICI1/AvrPikD and Pik-HMA/AvrPikD, respectively. **e–g**, SLC assays showed that PICI1, but not PICI2, specifically interacted with AvrPWL2 (**e**), AvrPi9 (**f**) and AvrPii (**g**) in *N. benthamiana*. Fluorescence signal intensity is indicated. Protein expression in *N. benthamiana* was detected by western blot (**e–g**, down panel). **h**, Protein levels of endogenous PICI1 in AvrPi9-Flag (AvrPi9-OE), AvrPWL2-Flag (AvrPWL2-OE), AvrPizt-Flag (AvrPizt-OE), and AvrPii-Flag (AvrPii-OE), compared with NIL-*Pigm*. The protein abundance was quantified and indicated under lanes. **i**, Subcellular localization of AvrPi9 and co-localization of PICI1-YFP/AvrPizt-mCherry in rice protoplasts. **j**, Western blot analysis confirming expression of AvrPi9-mCherry or AvrPi9-YFP without free mCherry or YFP in rice protoplasts. **k**, BiFC confirms the PICI1/AvrPi9 interaction in rice protoplasts. PICI2 served as a negative control. **l**, Basal blast resistance of WT and *AvrPi9*-OE lines, 5 dpi with TM21. Scale bars, 1 cm. Data were analysed by two-tailed Student's *t*-test. Lesion lengths (mean \pm s.d.; n = numbers of biologically independent samples in the graphs), fungal growth (mean \pm s.d.; $n = 3$, biologically independent samples). **m**, **n**, Induction of the PTI-related defence genes *OsPAL1* (**m**) and *KS4* (**n**) at 1 h after chitin incubation in WT and *AvrPi9*-OE. Two-tailed Student's *t*-test (mean \pm s.d.; $n = 3$, biologically independent samples). The rice *ACTIN1* served as an internal control. **o**, Cell-free degradation shows AvrPi9^{FL}-mediated PICI1-GFP degradation. Protein extracts were prepared from two-week-old PICI1-GFP/NIL-*Pigm* seedlings and then incubated with AvrPi9^{FL}-mCherry or AvrPizt-mCherry expressed in *N. benthamiana*. **p**, Cell-free degradation shows the AvrPi9^{FL}-mediated PICI1-GFP degradation in NIPB (left panel) and NIL-*Pigm* (right panel)

with or without MG132. Protein extracts were prepared from PICI1-GFP/NIPB and PICI1-GFP/NIL-*Pigm* seedlings and then incubated with mCherry (upper panels) or AvrPi9-mCherry (lower panels) expressed in *N. benthamiana*. FL, full length (**o**, **p**), and we used the Avrs without signal peptide to make the constructs in the experiments unless otherwise indicated. **q**, Immunodetection of two independent transformants of TH12^{AvrPi9} using an anti-S-tag antibody, with its parent TH12 as a negative control. Full length of AvrPi9 tagged by S-tag was transformed into TH12 by *Agrobacterium*-mediated transformation. **r**, Pathogenicity test of the transformants with TH12 as a virulent control. **s**, Protein levels of PICI1 in NIPB (left panel) and NIL-*Pigm* (right panel) after inoculation with TH12 (upper panel) or TH12^{AvrPi9} (lower panel). The protein abundance was quantified using ImageJ and indicated under lanes. Ponceau S staining (**o**–**q**) or Actin (**d**–**h**, **j**, **s**) served as loading control. Scale bars, 5 μm (**i**, **k**). Asterisks represent significant difference (* $P < 0.05$, ** $P < 0.01$) (**l**–**n**). Exact P values are provided in Supplementary Table 4 (**l**–**n**). Experiments were repeated twice (**a**, **d**, **h**, **j**, **q**, **s**) or three times (**b**, **c**, **e**–**g**, **i**, **k**–**p**, **r**) with similar results.

[Source data](#)

[Extended Data Fig. 9 PigmR and Pizt protect PICI1 from Avrs binding in a competitive manner.](#)

a, PICI1-interacting effectors including AvrPi9, AvrPWL2 and AvrPii can not activate PigmR-mediated cell death. Relative luciferase activity was measured after transformation with different effectors in NIPB or NIL-*Pigm*. Empty vector (EV) and MoCDIP1 were used as negative and positive control, respectively. One-way ANOVA with Tukey's test (mean ± s.d.; $n = 3$, biologically independent experiments). Different letters indicate significant difference at $P < 0.05$. **b**, Western blot analysis confirming expression of proteins in rice protoplasts. Note that 1–5 and 6–10 represented MoCDIP1, EV, AvrPi9, AvrPWL2 and AvrPii in NIPB and NIL-*Pigm*, respectively. The asterisks indicate the target proteins. **c**, Increased protein accumulation of PigmR in PigmR-7Myc-His/NIPB after inoculation with TH12 or TH12^{AvrPi9} with water as a control at the indicated time points. **d**, PICI1-PigmR interaction is enhanced by *M. oryzae*

infection. Proteins were prepared from transgenic plants PICI1-GFP/NIL-*Pigm* and PigmR-7Myc-His/NIPB after inoculation with TH12 or water as a control (36 hpi). **e**, A SLC assay shows that the PigmR disrupts PICI1-AvrPi9 interaction in *N. benthamiana*. Relative luciferase activity of protein-protein interactions was measured (lower panel). Two-tailed Student's *t*-test (mean \pm s.d.; $n = 6$, biologically independent samples). The asterisk represents significant difference (** $P < 0.01$). **f**, Western blot confirming protein expression in *N. benthamiana*. Note that left and right represent PICI1/AvrPi9/GFP and PICI1/AvrPi9/PigmR combination, respectively. **g**, PigmR-CC outcompetes AvrPi9 for PICI1 binding in vitro. **h**, PigmR-CC outcompetes AvrPi9 for PICI1 binding in rice protoplasts. **i, j**, SLC assays show that the PigmR-CC also disrupts PICI1/AvrPWL2 (**i**) and PICI1/AvrPii (**j**) interactions in *N. benthamiana*. Protein expression in *N. benthamiana* was detected by western blot (lower panel) (**i, j**). **k, l**, PigmR-CC outcompetes AvrPii (**k**) or AvrPWL2 (**l**) for PICI1 binding in rice protoplasts. The asterisks indicate the target proteins (**l**). **m, n**, Protein accumulation of PICI1 and OsMETS in Ky-Pi9 (**m**) and ZH11 (**n**) plants after inoculation with avirulent strain Guy11 (**m**) or YN2 (**n**) at the indicated time points. **o**, Protein accumulation of PICI1 in ZH11 and *Pizt*-KO/ZH11 plants after inoculation with blast strain CH131 (avirulent to *Pizt*) at the indicated time points. **p**, SLC assays show that the *Pizt*-CC also disrupts PICI1/AvrPi9 (left panel), PICI1/AvrPWL2 (middle panel) and PICI1/AvrPii (right panel) interactions in *N. benthamiana*. Protein expression in *N. benthamiana* was detected by western blot (lower panel). Actin (**b, c, f, i, j, m–o**) or Ponceau S staining (**p**) used as loading control. The protein abundance was quantified using ImageJ and indicated under lanes (**c, m–o**). Fluorescence signal intensity is indicated (**e, i, j, p**). Exact P values are provided in Supplementary Table 4 (**a, e**). Experiments were repeated twice (**b, d, f, g, h, k–p**) or three times (**a, c, e, i, j**) with similar results.

[Source data](#)

[**Extended Data Fig. 10 *PIC11* *Δp* allele shows higher *PIC11* expression and increased basal defence, and acetylation of *PIC11* promotes its degradation.**](#)

a, Twenty-seven SNPs in the *PICII* promoter distribute in different rice varieties based on the 3,000 Rice Genome database³³. SNP = 0 indicates the same promoter sequence of *PICII* with NIPB, which was used as the referee. SNP > 0 indicated at least one base change. **b**, Distribution of the 27 SNPs in *japonica* and *indica* rice based on the 3,000 Rice Genome Project database. **c**, F_{ST} of *PICII* and flanking regions between different rice groups. **d**, Induction of *ProPICII^{ind}-LUC* and *ProPICII^{jp}-LUC* in rice protoplasts by chitin. LUC activity was measured by normalizing to REN signal. **e**, *japonica* rice varieties showed a general higher induction of *PICII* than *indica* varieties by *M. oryzae* cocktail at 3 hpi. In box plots, the centre line represents the median, box edges delimit lower and upper quartiles and whiskers show the highest and lowest data points. Two-tailed Student's *t*-test (mean ± s.d.; *n* = numbers of biologically independent samples in the graphs). **f**, Pathotype test of different *M. oryzae* strains in two-week-old seedlings of Huajingxian74 using spraying inoculation, indicating YN2 and 99-30-1 virulent toward Huajingxian74. **g**, Induction of *PICII* at indicated times by *M. oryzae* infection in Huajingxian74 and CSSL. **h**, Induction of the defence gene *KS4* after blast inoculation in Huajingxian74 and CSSL plants. **i**, Pathotype test of *M. oryzae* isolate and Guy11 (avirulent) and TM21 (virulent) toward *Pi9*-containing rice plant (Ky-*Pi9*) with punch inoculation. **j**, Increased acetylation of PICII after blast infection. The protein extracts from PICI1-GFP/NIL-*Pigm* plants challenged with *M. oryzae* (TM21) or water for 36 h were immunoprecipitated with Ac-K (Acetylated-lysine) or GFP antibody and analysed using anti-GFP or anti-Ac-K antibody, respectively. **k**, A SLC assay shows the reduced interaction of PICI1^{K160Q}/PigmR-CC and PICI1^{K160Q}/OsMETS1 in comparison with PICI1 in *N. benthamiana*. The protein and transcript levels of cLuc-PICI1 and cLuc-PICI1^{K160Q} were determined by immunoblot and semi-quantitative PCR. Actin was detected as a loading control (**j**, **k**). The rice *ACTIN1* served as an internal control (**e**, **g**, **h**). Two-tailed Student's *t*-test (mean ± s.d.; *n* = 3, biologically independent samples) (**d**, **g**, **h**). Asterisks indicate significant differences (**P* < 0.05, ***P* < 0.01) (**d**, **e**, **g**, **h**). Exact *P* values are provided in Supplementary Table 4 (**d**, **e**, **g**, **h**). Scale bars, 1 cm (**f**, **i**). Experiments were repeated twice (**e**, **g**, **h**) or three times (**d**, **f**, **i-k**) with similar results.

Source data

Supplementary information

Supplementary Figure 1

Uncropped blots and gel images.

Reporting Summary

Supplementary Table 1

MS/MS and statistical analysis.

Supplementary Table 2

Natural variation analysis for *PICII*.

Supplementary Table 3

Primers used in this study.

Supplementary Table 4

Statistical summary—a summary of all statistical analysis is provided. Source data for all of the statistical analysis can be found in the Source Data files.

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

[**Source Data Extended Data Fig. 2**](#)

[**Source Data Extended Data Fig. 3**](#)

[**Source Data Extended Data Fig. 4**](#)

[**Source Data Extended Data Fig. 5**](#)

[**Source Data Extended Data Fig. 6**](#)

[**Source Data Extended Data Fig. 7**](#)

[**Source Data Extended Data Fig. 8**](#)

[**Source Data Extended Data Fig. 9**](#)

[**Source Data Extended Data Fig. 10**](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Zhai, K., Liang, D., Li, H. *et al.* NLRs guard metabolism to coordinate pattern- and effector-triggered immunity. *Nature* **601**, 245–251 (2022).
<https://doi.org/10.1038/s41586-021-04219-2>

- Received: 06 October 2020
- Accepted: 09 November 2021
- Published: 15 December 2021

- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04219-2>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04219-2>

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 15 December 2021](#)

Towards the biogeography of prokaryotic genes

- [Luis Pedro Coelho](#) ORCID: [orcid.org/0000-0002-9280-7885^{1,2,3}](https://orcid.org/0000-0002-9280-7885),
- [Renato Alves](#) ORCID: [orcid.org/0000-0002-7212-0234³](https://orcid.org/0000-0002-7212-0234),
- [Álvaro Rodríguez del Río](#) ORCID: [orcid.org/0000-0003-3907-3904⁴](https://orcid.org/0000-0003-3907-3904),
- [Pernille Neve Myers](#) ORCID: [orcid.org/0000-0002-1824-3305⁵](https://orcid.org/0000-0002-1824-3305),
- [Carlos P. Cantalapiedra](#) ORCID: [orcid.org/0000-0001-5263-533X⁴](https://orcid.org/0000-0001-5263-533X),
- [Joaquín Giner-Lamia](#) ORCID: [orcid.org/0000-0003-1553-8295^{4,6}](https://orcid.org/0000-0003-1553-8295),
- [Thomas Sebastian Schmidt](#)³,
- [Daniel R. Mende](#) ORCID: [orcid.org/0000-0001-6831-4557^{3,7}](https://orcid.org/0000-0001-6831-4557),
- [Askarbek Orakov](#) ORCID: [orcid.org/0000-0001-6823-5269³](https://orcid.org/0000-0001-6823-5269),
- [Ivica Letunic](#) ORCID: [orcid.org/0000-0003-3560-4288⁸](https://orcid.org/0000-0003-3560-4288),
- [Falk Hildebrand](#) ORCID: [orcid.org/0000-0002-0078-8948^{3,9,10}](https://orcid.org/0000-0002-0078-8948),
- [Thea Van Rossum](#)³,
- [Sofia K. Forslund](#) ORCID: [orcid.org/0000-0003-4285-6993^{3,11,12}](https://orcid.org/0000-0003-4285-6993),
- [Supriya Khedkar](#) ORCID: [orcid.org/0000-0001-6606-2202³](https://orcid.org/0000-0001-6606-2202),
- [Oleksandr M. Maistrenko](#) ORCID: [orcid.org/0000-0003-1961-7548³](https://orcid.org/0000-0003-1961-7548),
- [Shaojun Pan](#) ORCID: [orcid.org/0000-0002-5270-5614^{1,2}](https://orcid.org/0000-0002-5270-5614),
- [Longhao Jia](#)^{1,2},
- [Pamela Ferretti](#) ORCID: [orcid.org/0000-0002-1707-9013³](https://orcid.org/0000-0002-1707-9013),
- [Shinichi Sunagawa](#) ORCID: [orcid.org/0000-0003-3065-0314^{3,13}](https://orcid.org/0000-0003-3065-0314),
- [Xing-Ming Zhao](#)^{1,2},
- [Henrik Bjørn Nielsen](#) ORCID: [orcid.org/0000-0003-2281-5713¹⁴](https://orcid.org/0000-0003-2281-5713),
- [Jaime Huerta-Cepas](#) ORCID: [orcid.org/0000-0003-4195-5025^{3,4}](https://orcid.org/0000-0003-4195-5025) &
- [Peer Bork](#) ORCID: [orcid.org/0000-0002-2627-833X^{3,15,16,17}](https://orcid.org/0000-0002-2627-833X)

- 11k Accesses
- 345 Altmetric
- [Metrics details](#)

Subjects

- [Computational biology and bioinformatics](#)
- [Microbial ecology](#)

Abstract

Microbial genes encode the majority of the functional repertoire of life on earth. However, despite increasing efforts in metagenomic sequencing of various habitats^{1,2,3}, little is known about the distribution of genes across the global biosphere, with implications for human and planetary health. Here we constructed a non-redundant gene catalogue of 303 million species-level genes (clustered at 95% nucleotide identity) from 13,174 publicly available metagenomes across 14 major habitats and use it to show that most genes are specific to a single habitat. The small fraction of genes found in multiple habitats is enriched in antibiotic-resistance genes and markers for mobile genetic elements. By further clustering these species-level genes into 32 million protein families, we observed that a small fraction of these families contain the majority of the genes (0.6% of families account for 50% of the genes). The majority of species-level genes and protein families are rare. Furthermore, species-level genes, and in particular the rare ones, show low rates of positive (adaptive) selection, supporting a model in which most genetic variability observed within each protein family is neutral or nearly neutral.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Fig. 1: Global Microbial Gene Catalogue, version 1.

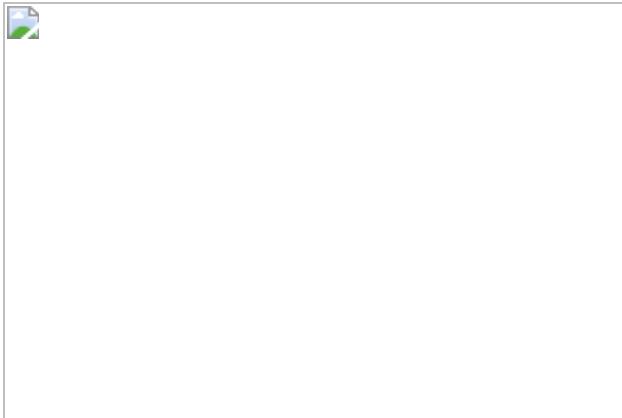


Fig. 2: The number of conspecific genes (gene pool per species) and the functional redundancy in each metagenome show significantly less variation within than between habitats.

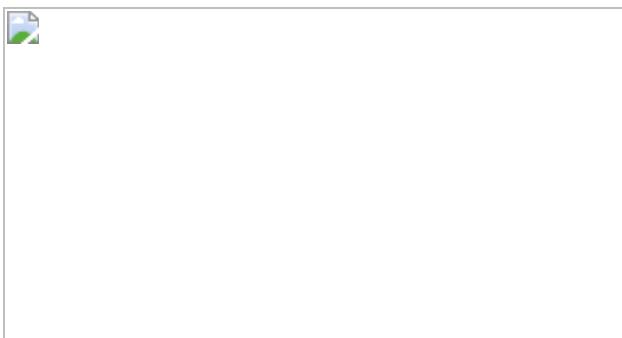


Fig. 3: Most genes are rare.

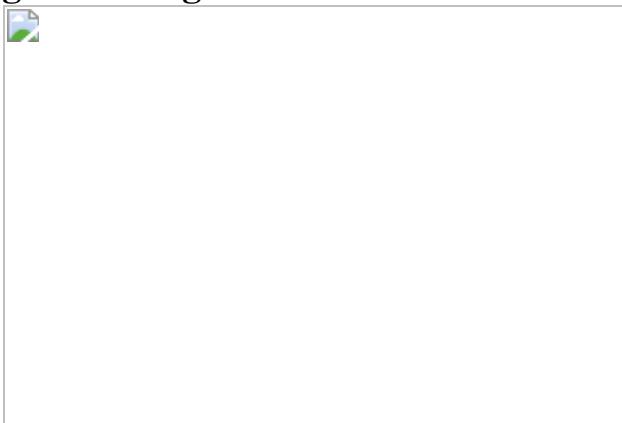
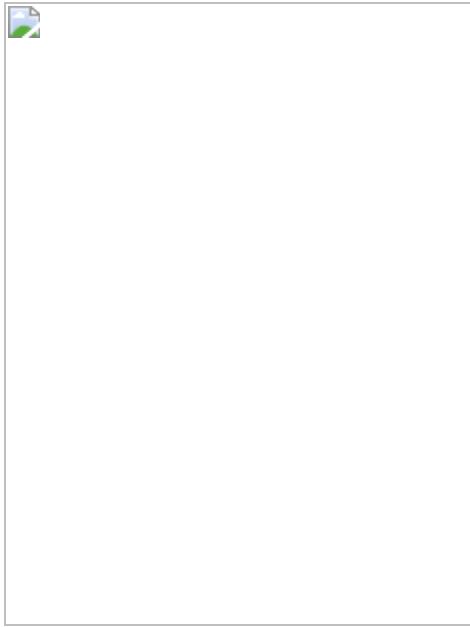


Fig. 4: Rare unigenes are under lower selection pressure.



Data availability

All data analysed during the current study are publicly available. Supplementary Table 1 contains the accession numbers for all the metagenomes used. GMGCv1 is available for download at <https://gmgc.embl.de>. The full catalogue is available for download as are sub-catalogues specialized to individual habitats and the subset derived only from sequenced genomes (which can be further subset to obtain the pangenome of a species of interest). Both the full catalogue and a version containing only complete ORFs are available as they represent different tradeoffs: the complete catalogue achieves higher coverage, while the version with only complete ORFs may be more appropriate for analyses that require the whole gene. Similarly, protein families are available at different amino acid identity thresholds (see ‘Protein family cluster calculation’). In addition to being available for download, the catalogue can be queried with an amino acid sequence. We developed and use a novel k -mer based algorithm (see ‘ k -mer based homology search’) to enable fast queries over the complete 303 million protein database and allow interactive use.

Code availability

The source code implementing the analyses in this manuscript is available on Github (https://github.com/luispedro/Coelho2021_GMGCv1) and is archived at Zenodo (<https://doi.org/10.5281/zenodo.4769556>).

References

1. 1.

Sunagawa, S. et al. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).

2. 2.

Zou, Y. et al. 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nat. Biotechnol.* **37**, 179–185 (2019).

3. 3.

Mohammad, B. F. et al. Structure and function of the global topsoil microbiome. *Nature* **560** 233–237 (2018).

4. 4.

Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).

5. 5.

Xiao, L. et al. A catalog of the mouse gut metagenome. *Nat. Biotechnol.* **33**, 1103–1108 (2015).

6. 6.

Coelho, L. P. et al. Similarity of the dog and human gut microbiomes in gene content and response to diet. *Microbiome* **6**, 72 (2018).

7. 7.

Pasolli, E. et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**, 649–662.e20 (2019).

8. 8.

Partridge, S. R., Kwong, S. M., Firth, N. & Jensen, S. O. Mobile genetic elements associated with antimicrobial resistance. *Clin. Microbiol. Rev.* **31**, (2018).

9. 9.

Mende, D. R. et al. ProGenomes2: An improved database for accurate and consistent habitat, taxonomic and functional annotations of prokaryotic genomes. *Nucleic Acids Res.* **48**, D621–D625 (2020).

10. 10.

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).

11. 11.

Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).

12. 12.

Daniel H. et al. RefSeq: an update on prokaryotic genome annotation and curation. *Nuc. Acids Res.* **46**, D851–D860 (2018).

13. 13.

Mering, C. von et al. Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315**, 1126–1130 (2007).

14. 14.

Richardson, E. J. et al. Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nat. Ecol. Evol.* **2**, 1468–1478 (2018).

15. 15.

Nielsen, H. B. et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).

16. 16.

Mende, D. R., Sunagawa, S., Zeller, G. & Bork, P. Accurate and universal delineation of prokaryotic species. *Nat. Methods* **10**, 881–884 (2013).

17. 17.

Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).

18. 18.

Louca, S. et al. Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* **2**, 936–943 (2018).

19. 19.

Maistrenko, O. M. et al. Disentangling the impact of environmental and phylogenetic constraints on prokaryotic within-species diversity. *ISME J.* **14**, 1247–1259 (2020).

20. 20.

Baumdicker, F., Hess, W. R. & Pfaffelhuber, P. The diversity of a distributed genome in bacterial populations. *Ann. Appl. Probab.* **20**, 1567–1606 (2010).

21. 21.

Sela, I., Wolf, Y. I. & Koonin, E. V. Theory of prokaryotic genome evolution. *Proc. Natl Acad. Sci. USA* **113**, 11399–11407 (2016).

22. 22.

Dandekar, T., Snel, B., Huynen, M. & Bork, P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **23**, 324–328 (1998).

23. 23.

Nei, M., Suzuki, Y. & Nozawa, M. The neutral theory of molecular evolution in the genomic era. *Annu. Rev. Genomics Hum. Genet.* **11**, 265–289 (2010).

24. 24.

Iranzo, J., Cuesta, J. A., Manrubia, S., Katsnelson, M. I. & Koonin, E. V. Disentangling the effects of selection and loss bias on gene dynamics. *Proc. Natl Acad. Sci. USA* **114**, E5616–E5624 (2017).

25. 25.

Wolf, Y. I., Makarova, K. S., Lobkovsky, A. E. & Koonin, E. V. Two fundamentally different classes of microbial genes. *Nat. Microbiol.* **2**, 16208 (2016).

26. 26.

Rasko, D. A. et al. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* **190**, 6881–6893 (2008).

27. 27.

Koskella, B., Hall, L. J. & Metcalf, C. J. E. The microbiome beyond the horizon of ecological and evolutionary theory. *Nat. Ecol. Evol.* **1**,

1606–1615 (2017).

28. 28.

Liu, R. et al. Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nat. Med.* **23**, 859–868 (2017).

29. 29.

Metcalf, J. L. et al. Microbial community assembly and metabolic function during mammalian corpse decomposition. *Science* **351**, 158–162 (2015).

30. 30.

Vincent, C. et al. Bloom and bust: intestinal microbiota dynamics in response to hospital exposures and *Clostridium difficile* colonization or infection. *Microbiome* **4**, 12 (2016).

31. 31.

Zeller, G. et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* **10**, 766 (2014).

32. 32.

Gibson, M. K. et al. Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat. Microbiol.* **1**, 16024 (2016).

33. 33.

Zhang, X. et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat. Med.* **21**, 895–905 (2015).

34. 34.

Brito, I. L. et al. Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435–439 (2016).

35. 35.

Vatanen, T. et al. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* **165**, 842–853 (2016).

36. 36.

Turnbaugh, P. J. et al. The human microbiome project. *Nature* **449**, 804–810 (2007).

37. 37.

Hannigan, G. D. et al. The human skin double-stranded DNA virome: topographical and temporal diversity, genetic enrichment, and dynamic associations with the host microbiome. *MBio* **6**, e01578-15 (2015).

38. 38.

Taft, D. H. et al. Intestinal microbiota of preterm infants differ over time and between hospitals. *Microbiome* **2**, 36 (2014).

39. 39.

Zeevi, D. et al. Personalized nutrition by prediction of glycemic responses. *Cell* **163**, 1079–1094 (2015).

40. 40.

Wilhelm, R. C. et al. Biogeography and organic matter removal shape long-term effects of timber harvesting on forest soil microbial communities. *ISME J.* **11**, 2552–2568 (2017).

41. 41.

Xie, H. et al. Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell Syst.* **3**, 572–

584.e3 (2016).

42. 42.

The MetaSUB International Consortium. The metagenomics and metadesign of the subways and urban biomes (metasub) international consortium inaugural meeting report. *Microbiome* **4**, 24 (2016).

43. 43.

Chatelier, E. L. et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).

44. 44.

Li, J. et al. Gut microbiota dysbiosis contributes to the development of hypertension. *Microbiome* **5**, (2017).

45. 45.

Pehrsson, E. C. et al. Interconnected microbiomes and resistomes in low-income human habitats. *Nature* **533**, 212–216 (2016).

46. 46.

Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).

47. 47.

Feng, Q. et al. Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat. Commun.* **6**, 6528 (2015).

48. 48.

Gu, Y. et al. Analyses of gut microbiota and plasma bile acids enable stratification of patients for antidiabetic treatment. *Nat. Commun.* **8**, 1785 (2017).

49. 49.

Karlsson, F. H. et al. Gut metagenome in european women with normal, impaired and diabetic glucose control. *Nature* **498**, 99–103 (2013).

50. 50.

Yu, J. et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).

51. 51.

Youngster, I. et al. Fecal microbiota transplant for relapsing clostridium difficile infection using a frozen inoculum from unrelated donors: a randomized, open-label, controlled pilot study. *Clin. Infect. Dis.* **58**, 1515–1522 (2014).

52. 52.

Guittar, J., Shade, A. & Litchman, E. Trait-based community assembly and succession of the infant gut microbiome. *Nat. Commun.* **10**, 512 (2019).

53. 53.

Vogtmann, E. et al. Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. *PLoS ONE* **11**, e0155362 (2016).

54. 54.

Chng, K. R. et al. Whole metagenome profiling reveals skin microbiome-dependent susceptibility to atopic dermatitis flare. *Nat Microbiol* **1**, 16106 (2016).

55. 55.

Chu, D. M. et al. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat. Med.* **23**, 314–326 (2017).

56. 56.

Van Rossum, T. et al. Spatiotemporal dynamics of river viruses, bacteria and microeukaryotes. Preprint at <https://doi.org/10.1101/259861> (2018).

57. 57.

Feng, Q. et al. Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. *Sci. Rep.* **6**, 22525 (2016).

58. 58.

Oh, J., Byrd, A. L., Park, M., Kong, H. H. & Segre, J. A. Temporal stability of the human skin microbiome. *Cell* **165**, 854–866 (2016).

59. 59.

Xiao, L. et al. A reference gene catalogue of the pig gut microbiome. *Nat. Microbiol.* **1**, 16161 (2016).

60. 60.

R Core Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, 2014).

61. 61.

Coelho, L. P. et al. NG-meta-profiler: Fast processing of metagenomes using ngless, a domain-specific language. *Microbiome* **7**, 84 (2019).

62. 62.

Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct De Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).

63. 63.

Besemer, J. & Borodovsky, M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res.* **33**, W451–W454 (2005).

64. 64.

Coelho, L. P. Jug: Software for parallel reproducible computation in Python. *J. Open Res. Softw.* **5**, 30 (2017).

65. 65.

Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using diamond. *Nat. Methods* **12**, 59–60 (2015).

66. 66.

Eberhardt, R. Y. et al. AntiFam: A tool to help identify spurious ORFs in protein annotation. *Database* **2012**, bas003 (2012).

67. 67.

Kang, D. et al. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

68. 68.

Li, H. Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. Preprint at <https://arxiv.org/abs/1303.3997> (2013).

69. 69.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

70. 70.

Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).

71. 71.

Zhou, W., Gay, N. & Oh, J. ReprDB and panDB: minimalist databases with maximal microbial representation. *Microbiome* **6**, 15 (2018).

72. 72.

Hingamp, P. et al. Exploring nucleo-cytoplasmic large DNA viruses in tara oceans microbial metagenomes. *ISME J.* **7**, 1678–1695 (2013).

73. 73.

Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

74. 74.

Huerta-Cepas, J. et al. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).

75. 75.

Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).

76. 76.

Smyshlyaev, G., Barabas, O. & Bateman, A. Sequence analysis allows functional annotation of tyrosine recombinases in prokaryotic genomes. *Mol. Syst. Biol.* **17**, e9880 (2021).

77. 77.

Jia, B. et al. CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).

78. 78.

Gibson, M. K., Forsberg, K. J. & Dantas, G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* **9**, 207–216 (2015).

79. 79.

Li, T., Fan, K., Wang, J. & Wang, W. Reduction of protein sequence complexity by residue grouping. *Protein Eng.* **16**, 323–330 (2003).

80. 80.

Zhao, M., Lee, W.-P., Garrison, E. P. & Marth, G. T. SSW library: an SIMD Smith–Waterman C/C++ library for use in genomic applications. *PLoS ONE* **8**, e82138 (2013).

81. 81.

Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2017).

82. 82.

Milanese, A. et al. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat. Commun.* **10**, 1014 (2019).

83. 83.

Salter, S. J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).

84. 84.

Kumar, R., Acharya, V., Singh, D. & Kumar, S. Strategies for high-altitude adaptation revealed from high-quality draft genome of non-violacein producing *Janthinobacterium lividum* ERGS5:01. *Stand. Genomic Sci.* **13**, 11 (2018).

85. 85.

Patijanasoontorn, B. et al. Hospital acquired *Janthinobacterium lividum* septicemia in srinagarind hospital. *J. Med. Assoc. Thai.* **75 Suppl 2**, 6–10 (1992).

86. 86.

Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).

87. 87.

Virtanen, P. et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

88. 88.

Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

89. 89.

Collins, R. E. & Higgs, P. G. Testing the infinitely many genes model for the evolution of the bacterial core genome and pangenome. *Mol. Biol. Evol.* **29**, 3413–3425 (2012).

90. 90.

Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol. Syst. Biol.* **7**, 539 (2011).

91. 91.

Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).

92. 92.

Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).

93. 93.

Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

94. 94.

Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–12 (2006).

95. 95.

Murrell, B. et al. FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. *Mol. Biol. Evol.* **30**, 1196–1205 (2013).

96. 96.

Smith, M. D. et al. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353 (2015).

Washietl, S. et al. RNACode: robust discrimination of coding and noncoding regions in comparative sequence data. *RNA* **17**, 578–594 (2011).

Acknowledgements

Funding was provided by the European Union's Horizon 2020 Research and Innovation Programme (grant 686070: DD-DeCaF to P.B.) and Marie Skłodowska-Curie Actions (grant 713673 to A.R.d.R.), the European Research Council (ERC) MicrobioS (ERC-AdG-669830 to P.B.), JTC project jumpAR (01KI1706 to P.B.), a BMBF Grant (grant 031L0181A: LAMarCK to P.B.), the European Molecular Biology Laboratory (P.B.), the ETH and Helmut Horten Foundation (S.S.), the National Key R&D Program of China (grant 2020YFA0712403 to X.-M.Z.), National Natural Science Foundation of China (grant 61932008 to X.-M.Z.; grant 61772368 to X.-M.Z.; grant 31950410544 to L.P.C.), the Shanghai Municipal Science and Technology Major Project (grant 2018SHZDZX01 to X.-M.Z. and L.P.C.) and Zhangjiang Lab (X.-M.Z. and L.P.C.), the International Development Research Centre (grant 109304, EMBARK under the JPI AMR framework; to L.P.C.), la Caixa Foundation (grant 100010434, fellowship code LCF/BQ/DI18/11660009 to A.R.d.R.), the Severo Ochoa Program for Centres of Excellence in R&D from the Agencia Estatal de Investigación of Spain (grant SEV-2016-0672 (2017–2021) to C.P.C.), the Ministerio de Ciencia, Innovación y Universidades (grant PGC2018-098073-A-I00 MCIU/AEI/FEDER to J.H.-C. and J.G.-L.), the Innovation Fund Denmark (grant 4203-00005B, PNM), the Biotechnology and Biological Sciences research Council (BBSrC) Institute Strategic Programme Gut Microbes and Health BB/r012490/1 and its constituent project BBS/e/F/000Pr10355 (F.H.). R.A. is a member of the Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences. The authors thank the Bork group for helpful discussion, in particular A. Głązek for discussions of algorithm design, J. C. Somody (EMBL) for help with figure design, and A. Fullam (EMBL) for computational assistance in processing the MAGs.

Author information

Affiliations

1. Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

Luis Pedro Coelho, Shaojun Pan, Longhao Jia & Xing-Ming Zhao

2. MOE Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, and MOE Frontiers Center for Brain Science, Shanghai, China

Luis Pedro Coelho, Shaojun Pan, Longhao Jia & Xing-Ming Zhao

3. Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

Luis Pedro Coelho, Renato Alves, Thomas Sebastian Schmidt, Daniel R. Mende, Askarbek Orakov, Falk Hildebrand, Thea Van Rossum, Sofia K. Forslund, Supriya Khedkar, Oleksandr M. Maistrenko, Pamela Ferretti, Shinichi Sunagawa, Jaime Huerta-Cepas & Peer Bork

4. Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC), Madrid, Spain

Álvaro Rodríguez del Río, Carlos P. Cantalapiedra, Joaquín Giner-Lamia & Jaime Huerta-Cepas

5. Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark

Pernille Neve Myers

6. Departamento de Biotecnología-Biología Vegetal, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas,

Universidad Politécnica de Madrid (UPM), Madrid, Spain

Joaquín Giner-Lamia

7. Daniel K. Inouye Center for Microbial Oceanography: Research and Education, University of Hawai‘i at Mānoa, Honolulu, HI, USA

Daniel R. Mende

8. biobytte solutions GmbH, Heidelberg, Germany

Ivica Letunic

9. Earlham Institute, Norwich Research Park, Norwich, UK

Falk Hildebrand

10. Gut Health and Microbes Programme, Quadram Institute, Norwich Research Park, Norwich, UK

Falk Hildebrand

11. Experimental and Clinical Research Center (ECRC), a joint venture of the Max Delbrück Centre (MDC) and Charité University Hospital, Berlin, Germany

Sofia K. Forslund

12. Berlin Initiative of Health, Berlin, Germany

Sofia K. Forslund

13. Department of Biology, Institute of Microbiology and Swiss Institute of Bioinformatics, ETH Zürich, Zürich, Switzerland

Shinichi Sunagawa

14. Clinical Microbiomics A/S, Copenhagen, Denmark

Henrik Bjørn Nielsen

15. Max Delbrück Centre for Molecular Medicine, Berlin, Germany

Peer Bork

16. Yonsei Frontier Lab (YFL), Yonsei University, Seoul, South Korea

Peer Bork

17. Department of Bioinformatics, Biocenter, University of Würzburg,
Würzburg, Germany

Peer Bork

Contributions

The study was conceived and supervised by P.B. and designed by L.P.C., S.S., J.H.-C. and P.B. L.P.C., R.A., A.R.d.R., P.N.M., T.S.S., A.O., F.H., T.V.R., S.K.F., S.K., O.M.M., P.F. and J.H.-C. analysed data. L.P.C., T.S.S., F.H., T.V.R., S.K.F., P.F., J.H.-C. and P.B. drafted the manuscript. L.P.C., R.A., A.R.d.R., C.P.C. and D.R.M. built the unigene, protein clusters and protein family catalogues. L.P.C., R.A., T.S.S., D.R.M., I.L., F.H., S.K.F., S.K. and J.H.-C. annotated the catalogue. A.R.d.R., C.P.C., J.G.-L., O.M.M. and J.H.-C. performed the selection pressure analyses. P.N.M. and H.B.N. built the MGSSs. L.P.C., R.A., I.L., S.P., L.J., X.-M.Z., T.V.R. and J.H.-C. designed and implemented the web resource, including the search algorithms and the associated GMGC-mapper tool. L.P.C., T.S.S., F.H. and O.M.M. annotated metagenomes. T.S.S. and A.O. built the MAGs. All authors contributed to the review of the manuscript before submission for publication and approved the final version.

Corresponding authors

Correspondence to [Luis Pedro Coelho](#), [Jaime Huerta-Cepas](#) or [Peer Bork](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Gene accumulation curves. Legend.

(a) For most (but not all) habitats, unigenes with high prevalence ($\geq 5\%$) have been well-captured, while rare unigenes continue to be found in each new sample. **(b-d)** New unigenes continue to be found in each sample. Each grey line represents a random permutation of the samples, while the solid black line shows the mean over these random permutations. The dotted red line is least-squares fit of Heap's Law ($N = k \cdot \text{sample}^{\alpha}$). In all cases, the parameter fit indicates that the number of genes has not reached saturation. **(e)** The number of assembled/detected genes per sample grows with sequencing depth without a plateau being reached. **(f)** Similarly, the number of detected ORFs per insert grows with sequencing depth.

Extended Data Fig. 2 Identity thresholds and their relationship to taxonomy and function in the GMGCv1.

Legend: **(a)** A 95% nucleotide identity threshold is a proxy for species. Shown is nucleotide identity of closest gene homolog within the same species or within the same genus (excluding within-species comparisons). The threshold used in this work (95%) is marked with a dashed red line. **(b)** Within well-conserved, universal, 40 single-copy orthologues (see Methods), the average pairwise amino acid identity is 49%, albeit with a wide range (27-75%) when considering within-orthologue averages. In dashed red, the thresholds used for building protein families are

highlighted. Boxplots display quartiles and ranges (see Methods). **(c)** Proportion of genes annotated at each taxonomic level.

Extended Data Fig. 3 Short reads map to the GMGCv1 at higher rates (compared to a reference database of reference genomes).

Legend: **(a)** Mapping rates for short reads from metagenomes mapped against the GMGCv1 or the reference genomes in proGenomes2. **(b)** Fraction of short reads from human gut metagenomes mapping to a collection of sequenced genomes and the GMGCv1, per country, **(c)** Same data as **(b)**, aggregated by the World Bank's classification of countries into income groups. In all panels, boxplots show quartiles (including median) and range (except for outliers, see Methods). Blue boxes show mapping rates to proGenomes2, while orange boxes show mapping rates to GMGCv1.

Extended Data Fig. 4 MAGs only capture a small fraction of all genes in a sample.

Legend: Fraction of undetected genes when mapping to only the genes captured by metagenome-assembled genomes (MAGs) across the habitats compared to mapping to the full GMGCv1.

Extended Data Fig. 5 Species and protein cluster sharing between habitats is similar to unigene sharing, but sharing of protein families is more extensive.

Legend: **(a)** The sharing of metagenomic species between habitats mimics unigene sharing. Width of each ribbons represents the number of MGSs shared between the habitats (the largest number shared is between the human and the pig gut, which share 166 MGSs out of 1,908 MGSs in the human gut and 898 in pig gut, respectively). **(b)** Species-level unigene sharing between habitats by fraction of the number of unigenes from each habitat (cf. Fig. 1b, which uses abundance weighting). **(c)** Sharing of protein clusters (90% amino acid identity clusters) between habitats, abundance-weighted. **(d)** Sharing of protein families between habitats,

abundance-weighted. When considering coarser clusterings of sequences, gene sharing between habitats increases, yet we still observed higher rates of sharing between similar habitats and significant fractions of habitat-specific families (e.g., in the marine environment, 31.3% of the genes, by abundance, are in marine-specific protein families).

Extended Data Fig. 6 Antibiotic resistance and mobile genes are more likely to be multi-habitat genes, while most species are found in a single habitat.

Legend: **(a)** Fraction of unigenes within each habitat which are multi-habitat genes (for all unigenes, or when considering only mobile elements or antibiotic resistance genes). **(b)** A total of 7,443 MGSs were built, across all the habitats as species proxies to reliably assess their habitats. Each circle shows the number of metagenomic species for each habitat, x-axis represents the number of genes in the catalogue specific to each habitat, the y-axis represents the number of samples. Note that differing sampling depth and habitat-specific biodiversity impact those numbers.

Extended Data Fig. 7 Determinants of functional community structure.

Legend: **(a)** principal coordinate analysis of all samples by protein family profile and the correlations with taxonomic and protein family richness (after rarefying to 1 million inserts to remove effects of sample depth). **(b)** Hierarchical clustering of the habitats using high-level functional profiles based.

Extended Data Fig. 8 Marine and soil richness patterns are a mixture of sub patterns.

Legend: Conspecific genes per species in marine **(a)** and **(b)** soil sub-habitats. The differences in the marine environment are particularly large when comparing the samples in the photic zones (the shallower, light-accessible, surface and deep-chlorophyll maximum samples) to the non-photic mesopelagic samples (deeper, beyond the reach of sunlight). The

differences in the soil environment follow differences in acidity (with Podzol, Dystric Brunisol and Ultic soils being acidic, while Luvisols are usually neutral or alkaline) and differences in moisture (with Xeralfs being dry in the summer, while Glossudalfs are moist year round).

Extended Data Fig. 9 Most genes are detected only in frequently and rare genes are (on average) present at a lower abundance in metagenomes.

Legend: **(a)** Shown are the percentage of genes detected in at most 1,...,50 metagenomes (out of a total of 13,174). **(b, c)** Histograms of gene prevalence are roughly linear on a log-log scale, as predicted from neutral or nearly-neutral evolution models. Shown are histograms for 90% amino acid identity protein clusters **(b)** and 20% amino acid identity protein families **(c)**, which behave similar to species-level unigenes (see Fig. 3). **(d)** Shown is the percentage of genes in each sample that is composed of rare genes (**Count**) and the total abundance represented by these (**Abundance**). Except for wastewater (likely due to under-sampling), rare genes represent a lower fraction of the abundance than of detection. Boxplots show quartiles (including median drawn as a line) and whiskers show the range of the data excluding outliers, which are shown as extra elements (see Methods).

Extended Data Fig. 10 More abundant and larger protein families are under more intense selection.

Legend: **(a)** dN/dS within each protein family, with protein families split into 5 abundance quintiles, showing a downward trend with abundance (higher negative selection). **(b)** dN/dS within each gene size category, similarly showing a downward trend with size. Categories are defined by increasing size, with each bin representing the same number of unigenes. Boxplots show quartiles and ranges (see Methods).

Supplementary information

Reporting Summary

Supplementary Table 1

List of samples included in the global gene catalogue including their habitat annotation. This table includes the accession ID of the metagenome, the GC fraction, and the size of the input (measured in total basepairs and number of inserts) both prior to (Raw) and post quality filtering (HQ, for high-quality). It also includes the number of ORFs predicted.

Supplementary Table 2

Mapping of our human-readable habitat terms to predefined ontology terms. List of habitat annotations used and their correspondence to EnvO/Uberon terms and NCBI taxonomic ID of the host (for host-associated habitat).

Supplementary Table 3

Statistics of (redundant) ORF predictions. The tables summarize the assembly statistics per habitat (including number of contigs, average contig size, N50) and the ORF stats per habitat, both complete ORFs and fragmentary ORFs.

Supplementary Table 4

Robustness checks for protein family definition and inclusion or exclusion of fragmentary ORFs. The major statements in the manuscript were tested using alternative protein family thresholds as well as including or excluding fragmentary ORFs to ensure robustness.

Supplementary Table 5

List of metagenome-assembled genomes and their statistics. All the assembled bins are included, including those marked as low-quality (following standard criteria, see Methods).

Supplementary Table 6

Pearson correlations between observed gene prevalence and theoretical prediction assuming neutral evolution.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Coelho, L.P., Alves, R., del Río, Á.R. *et al.* Towards the biogeography of prokaryotic genes. *Nature* **601**, 252–256 (2022).
<https://doi.org/10.1038/s41586-021-04233-4>

- Received: 16 June 2019
- Accepted: 12 November 2021
- Published: 15 December 2021
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04233-4>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04233-4>

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 22 December 2021](#)

Non-syntrophic methanogenic hydrocarbon degradation by an archaeal species

- [Zhuo Zhou](#) [ORCID: orcid.org/0000-0001-7014-5748](#)^{1 na1},
- [Cui-jing Zhang](#) [ORCID: orcid.org/0000-0003-0904-5531](#)^{2 na1},
- [Peng-fei Liu](#)^{1,3 na1},
- [Lin Fu](#)^{1 na1},
- [Rafael Laso-Pérez](#) [ORCID: orcid.org/0000-0002-6912-7865](#)^{4,5 nAff7},
- [Lu Yang](#)¹,
- [Li-ping Bai](#)¹,
- [Jiang Li](#)¹,
- [Min Yang](#)¹,
- [Jun-zhang Lin](#)⁶,
- [Wei-dong Wang](#)⁶,
- [Gunter Wegener](#) [ORCID: orcid.org/0000-0002-6819-373X](#)^{4,5},
- [Meng Li](#) [ORCID: orcid.org/0000-0001-8675-0758](#)² &
- [Lei Cheng](#) [ORCID: orcid.org/0000-0003-1178-8190](#)¹

[Nature](#) volume 601, pages 257–262 (2022)

- 9219 Accesses
- 1 Citations
- 228 Altmetric
- [Metrics details](#)

Subjects

- [Archaeal physiology](#)
- [Carbon cycle](#)

Abstract

The methanogenic degradation of oil hydrocarbons can proceed through syntrophic partnerships of hydrocarbon-degrading bacteria and methanogenic archaea^{1,2,3}. However, recent culture-independent studies have suggested that the archaeon '*Candidatus Methanoliparum*' alone can combine the degradation of long-chain alkanes with methanogenesis^{4,5}. Here we cultured *Ca.* *Methanoliparum* from a subsurface oil reservoir. Molecular analyses revealed that *Ca.* *Methanoliparum* contains and overexpresses genes encoding alkyl-coenzyme M reductases and methyl-coenzyme M reductases, the marker genes for archaeal multicarbon alkane and methane metabolism. Incubation experiments with different substrates and mass spectrometric detection of coenzyme-M-bound intermediates confirm that *Ca.* *Methanoliparum* thrives not only on a variety of long-chain alkanes, but also on *n*-alkylcyclohexanes and *n*-alkylbenzenes with long *n*-alkyl ($C_{\geq 13}$) moieties. By contrast, short-chain alkanes (such as ethane to octane) or aromatics with short alkyl chains ($C_{\leq 12}$) were not consumed. The wide distribution of *Ca.* *Methanoliparum*^{4,5,6} in oil-rich environments indicates that this alkylotrophic methanogen may have a crucial role in the transformation of hydrocarbons into methane.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Fig. 1: Methanogenesis in the oily sludge and visualization of microorganisms.

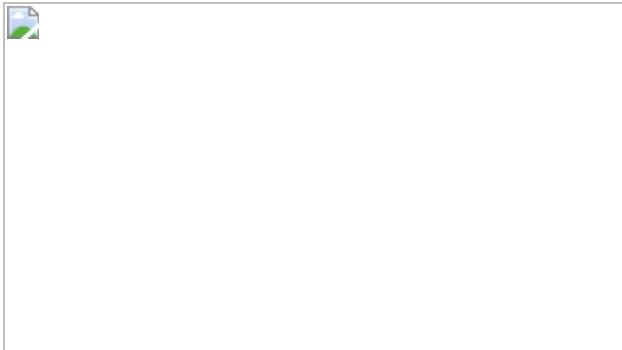


Fig. 2: Methanogenic hexadecane degradation by *Ca. Methanoliparum*.

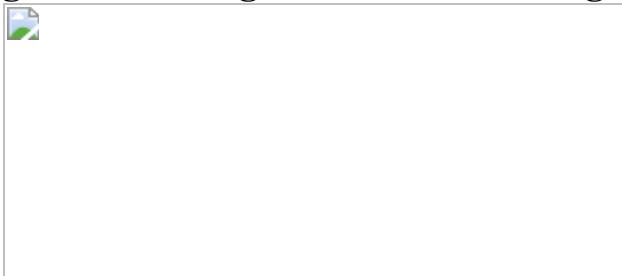


Fig. 3: Hexadecane degradation pathway of *Ca. Methanoliparum*.

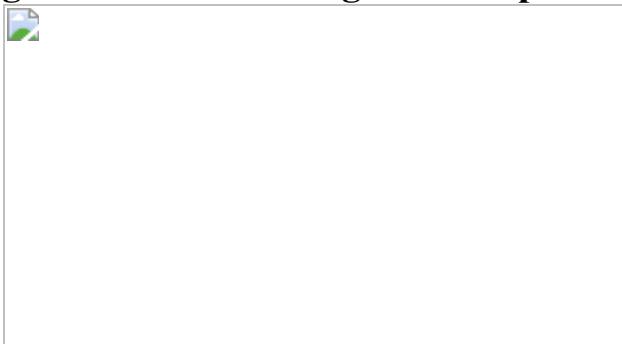


Fig. 4: Identification of the intermediate hexadecyl-CoM.

Data availability

The 16S rRNA gene amplicon sequences, metagenomic and metatranscriptomic data generated in current study are available in the NODE database (<http://www.biosino.org/node/project/detail/OEP001282>).

The data of dereplicated MAGs analysed during the current study are available in the NODE database under the accession numbers OEZ006960 and OEZ007009–OEZ007026. Further details are provided in Supplementary Table 13. All other data are available in the main text or the [Supplementary Information](#).

Code availability

The sources of the code and programs used for analyses are mentioned in the [Methods](#), and are also available at GitHub (https://github.com/liupfskygre/Methanoliparum_MS_code/tree/main).

References

1. 1. Jones, D. M. et al. Crude-oil biodegradation via methanogenesis in subsurface petroleum reservoirs. *Nature* **451**, 176–180 (2008).
2. 2. Zengler, K., Richnow, H. H., Rossello-Mora, R., Michaelis, W. & Widdel, F. Methane formation from long-chain alkanes by anaerobic microorganisms. *Nature* **401**, 266–269 (1999).
3. 3. Dolfing, J., Larter, S. R. & Head, I. M. Thermodynamic constraints on methanogenic crude oil biodegradation. *ISME J.* **2**, 442–452 (2008).
4. 4. Laso Pérez, R. et al. Anaerobic degradation of non-methane alkanes by “*Candidatus Methanoliparia*” in hydrocarbon seeps of the Gulf of Mexico. *mBio* **10**, e01814-19 (2019).
5. 5.

Borrel, G. et al. Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. *Nat. Microbiol.* **4**, 603–613 (2019).

6. 6.

Cheng, L. et al. Progressive degradation of crude oil *n*-alkanes coupled to methane production under mesophilic and thermophilic conditions. *PLoS ONE* **9**, e113253 (2014).

7. 7.

Head, I. M., Jones, D. M. & Röling, W. F. M. Marine microorganisms make a meal of oil. *Nat. Rev. Microbiol.* **4**, 173–182 (2006).

8. 8.

Van Hamme, J. D., Singh, A. & Ward, O. P. Recent advances in petroleum microbiology. *Microbiol. Mol. Biol. Rev.* **67**, 503–549 (2003).

9. 9.

Aitken, C. M., Jones, D. M. & Larter, S. R. Anaerobic hydrocarbon biodegradation in deep subsurface oil reservoirs. *Nature* **431**, 291–294 (2004).

10. 10.

Head, I. M., Jones, D. M. & Larter, S. R. Biological activity in the deep subsurface and the origin of heavy oil. *Nature* **426**, 344–352 (2003).

11. 11.

Gieg, L. M., Fowler, S. J. & Berdugo-Clavijo, C. Syntrophic biodegradation of hydrocarbon contaminants. *Curr. Opin. Biotechnol.* **27**, 21–29 (2014).

12. 12.

Rabus, R. et al. Anaerobic microbial degradation of hydrocarbons: from enzymatic reactions to the environment. *J. Mol. Microbiol. Biotechnol.* **26**, 5–28 (2016).

13. 13.

Fowler, S. J., Dong, X., Sensen, C. W., Suflita, J. M. & Gieg, L. M. Methanogenic toluene metabolism: community structure and intermediates. *Environ. Microbiol.* **14**, 754–764 (2012).

14. 14.

Thauer, R. K. Methyl (alkyl)-coenzyme M reductases: nickel F-430-containing enzymes involved in anaerobic methane formation and in anaerobic oxidation of methane or of short chain alkanes. *Biochemistry* **58**, 5198–5220 (2019).

15. 15.

Hahn, C. J. et al. “*Candidatus Ethanoperedens*”, a thermophilic genus of Archaea mediating the anaerobic oxidation of ethane. *mBio* **11**, e00600-20 (2020).

16. 16.

Laso-Pérez, R. et al. Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* **539**, 396–401 (2016).

17. 17.

Chen, S.-C. et al. Anaerobic oxidation of ethane by archaea from a marine hydrocarbon seep. *Nature* **568**, 108–111 (2019).

18. 18.

Wang, Y., Wegener, G., Hou, J., Wang, F. & Xiao, X. Expanding anaerobic alkane metabolism in the domain of Archaea. *Nat. Microbiol.* **4**, 595–602 (2019).

19. 19.

Wang, Y., Wegener, G., Ruff, S. E. & Wang, F. Methyl/alkyl-coenzyme M reductase-based anaerobic alkane oxidation in Archaea. *Environ. Microbiol.* **23**, 530–541 (2020).

20. 20.

Boyd, J. A. et al. Divergent methyl-coenzyme M reductase genes in a deep-subseafloor Archaeoglobi. *ISME J.* **13**, 1269–1279 (2019).

21. 21.

Baker, B. J. et al. Diversity, ecology and evolution of Archaea. *Nat. Microbiol.* **5**, 887–900 (2020).

22. 22.

Seitz, K. W. et al. Asgard archaea capable of anaerobic hydrocarbon cycling. *Nat. Commun.* **10**, 1822 (2019).

23. 23.

Cheng, L. et al. Preferential degradation of long-chain alkyl substituted hydrocarbons in heavy oil under methanogenic conditions. *Org. Geochem.* **138**, 103927 (2019).

24. 24.

Oldenburg, T. B. P. et al. The controls on the composition of biodegraded oils in the deep subsurface—part 4. Destruction and production of high molecular weight non-hydrocarbon species and destruction of aromatic hydrocarbons during progressive in-reservoir biodegradation. *Org. Geochem.* **114**, 57–80 (2017).

25. 25.

Cheng, L. et al. DNA-SIP reveals that *Syntrophaceae* play an important role in methanogenic hexadecane degradation. *PLoS ONE* **8**,

e66784 (2013).

26. 26.

Liu, Y.-F. et al. Anaerobic hydrocarbon degradation in candidate phylum ‘Atribacteria’ (JS1) inferred from genomics. *ISME J.* **13**, 2377–2390 (2019).

27. 27.

Liu, Y.-F. et al. Anaerobic degradation of paraffins by thermophilic Actinobacteria under methanogenic conditions. *Environ. Sci. Technol.* **54**, 10610–10620 (2020).

28. 28.

Breese, K., Boll, M., Alt-Mörbe, J., Schägger, H. & Fuchs, G. Genes coding for the benzoyl-CoA pathway of anaerobic aromatic metabolism in the bacterium *Thauera aromatica*. *Eur. J. Biochem.* **256**, 148–154 (1998).

29. 29.

Egland, P. G., Pelletier, D. A., Dispensa, M., Gibson, J. & Harwood, C. S. A cluster of bacterial genes for anaerobic benzene ring biodegradation. *Proc. Natl Acad. Sci. USA* **94**, 6484–6489 (1997).

30. 30.

Borrel, G. et al. Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC Genom.* **15**, 679 (2014).

31. 31.

Lyu, Z., Shao, N., Akinyemi, T. & Whitman, W. B. Methanogenesis. *Curr. Biol.* **28**, R727–R732 (2018).

32. 32.

Ferry, J. G. & Lessner, D. J. Methanogenesis in marine sediments. *Ann. N. Y. Acad. Sci.* **1125**, 147–157 (2008).

33. 33.

Thauer, R. K., Kaster, A.-K., Seedorf, H., Buckel, W. & Hedderich, R. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* **6**, 579–591 (2008).

34. 34.

Mayumi, D. et al. Methane production from coal by a single methanogen. *Science* **354**, 222–225 (2016).

35. 35.

Suflita, J. M., Davidova, I. A., Gieg, L. M., Nanny, M. & Prince, R. C. in *Studies in Surface Science and Catalysis* Vol. 151 (eds Vazquez-Duhalt, R. & Quintero-Ramirez, R.) 283–305 (Elsevier, 2004).

36. 36.

Bryant, M. Commentary on the Hungate technique for culture of anaerobic bacteria. *Am. J. Clin. Nutr.* **25**, 1324–1328 (1972).

37. 37.

Friedrich, W., Antje, B. & Ralf, R. in *The Prokaryotes: Ecophysiology and Biochemistry* Vol. 2 (eds Martin Dworkin et al.) 1028–1049 (Springer, 2006).

38. 38.

Aydin, O. & Yassikaya, M. Y. Validity and reliability analysis of the plotdigitizer software program for data extraction from single-case graphs. *Perspect. Behav. Sci.* (2021).

39. 39.

Dolfing, J. & Mulder, J.-W. Comparison of methane production rate and coenzyme F₄₂₀ content of methanogenic consortia in anaerobic granular sludge. *Appl. Environ. Microbiol.* **49**, 1142–1145 (1985).

40. 40.

Cheng, L., Dai, L., Li, X., Zhang, H. & Lu, Y. Isolation and characterization of *Methanothermobacter crinale* sp. nov, a novel hydrogenotrophic methanogen from the Shengli oil field. *Appl. Environ. Microbiol.* **77**, 5212–5219 (2011).

41. 41.

Ma, T.-T. et al. Coexistence and competition of sulfate-reducing and methanogenic populations in an anaerobic hexadecane-degrading culture. *Biotechnol. Biofuels* **10**, 207 (2017).

42. 42.

Stumm, W. & Morgan, J. J. *Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters* (Wiley, 1996).

43. 43.

Deines, P., Langmuir, D. & Harmon, R. S. Stable carbon isotope ratios and the existence of a gas phase in the evolution of carbonate ground waters. *Geochim. Cosmochim. Acta* **38**, 1147–1164 (1974).

44. 44.

Pernthaler, A., Pernthaler, J. & Amann, R. Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* **68**, 3094–3101 (2002).

45. 45.

Amann, R. I. et al. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56**, 1919–1925 (1990).

46. 46.

Daims, H., Brühl, A., Amann, R., Schleifer, K.-H. & Wagner, M. The domain-specific probe EUB338 is insufficient for the detection of all bacteria: development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.* **22**, 434–444 (1999).

47. 47.

Stahl, D. A. in *Nucleic Acid Techniques in Bacterial Systematics* 205–248 (1991).

48. 48.

Pernthaler, A., Preston, C. M., Pernthaler, J., DeLong, E. F. & Amann, R. Comparison of fluorescently labeled oligonucleotide and polynucleotide probes for the detection of pelagic marine bacteria and archaea. *Appl. Environ. Microbiol.* **68**, 661–667 (2002).

49. 49.

Sofie, T. et al. Comparative evaluation of four bacteria-specific primer pairs for 16S rRNA gene surveys. *Front. Microbiol.* **8**, 494 (2017).

50. 50.

Wei, S. et al. Comparative evaluation of three archaeal primer pairs for exploring archaeal communities in deep-sea sediments and permafrost soils. *Extremophiles* **23**, 747–757 (2019).

51. 51.

Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

52. 52.

Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).

53. 53.

Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).

54. 54.

Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).

55. 55.

Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

56. 56.

Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).

57. 57.

Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

58. 58.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered

from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

59. 59.

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2019).

60. 60.

Parks, D. H. et al. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* **38**, 1079–1086 (2020).

61. 61.

Shaffer, M. et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* **48**, 8883–8900 (2020).

62. 62.

Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).

63. 63.

Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).

64. 64.

Yoon, S. H., Ha, S. M., Lim, J., Kwon, S. & Chun, J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* **110**, 1281–1286 (2017).

65. 65.

Qin, Q.-L. et al. A proposed genus boundary for the prokaryotes based on genomic insights. *J. Bacteriol.* **196**, 2210–2215 (2014).

66. 66.

Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).

67. 67.

Eddy, S. R. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. *PLoS Comput. Biol.* **4**, e1000069 (2008).

68. 68.

Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

69. 69.

Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

70. 70.

Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).

71. 71.

Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

72. 72.

Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829 (2012).

73. 73.

Ludwig, W. et al. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).

74. 74.

Mendler, K. et al. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res.* **47**, 4442–4448 (2019).

75. 75.

Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

76. 76.

Lane, D. J. *16S/23S rRNA Sequencing* 205–248 (John Wiley & Sons, 1991).

77. 77.

Selvaraj, V. A.-O. et al. Development of a duplex droplet digital PCR assay for absolute quantitative detection of "Candidatus Liberibacter asiaticus". *PLoS ONE* **13**, e0197184 (2018).

78. 78.

Peng, J., Lü, Z., Rui, J. & Lu, Y. Dynamics of the methanogenic archaeal community during plant residue decomposition in an anoxic rice field soil. *Appl. Environ. Microbiol.* **74**, 2894–2901 (2008).

79. 79.

Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).

80. 80.

Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).

81. 81.

Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).

82. 82.

Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

83. 83.

Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

84. 84.

Quinlan, A. R. BEDTools: the Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinform.* **47**, 11.12.1–11.12.34 (2014).

85. 85.

Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997> (2013).

86. 86.

Wickham, H. *ggplot2*. *Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185 (2011).

87. 87.

RCore Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020);
<http://www.R-project.org/>

Acknowledgements

We thank A. Oren (The Hebrew University of Jerusalem) for discussing the naming of the different *Ca. Methanoliparum* species; R. Conrad and W. B. Whitman for discussing the manuscript; K. Wrighton for providing access to the server Zenith; Q. Yuan, Y. Liu, J. Pan, M.-w. Cai and Y.-n. Tang for assisting in data analysis; L.-r. Dai, D. Zhang and L. Li for assisting in cultivation and experiments; and Z. Zhou for technical support. This study was supported by National Natural Science Foundation of China (nos 92051108, 91851105, 41802179, 31970066, 31570009 and 31970105), Agricultural Science and Technology Innovation Project of the Chinese Academy of Agriculture Science (no. CAAS-ASTIP-2016-BIOMA), the Innovation Team Project of Universities in Guangdong Province (no. 2020KCXTD023) and the Shenzhen Science and Technology Program (no. JCYJ20200109105010363), the Fundamental Research Funds for the Central Universities (LZUJBKY-2021-KB16), the Central Public-interest Scientific Institution Basal Research Fund (Y2021PT02, Y2021XK06). R.L.-P. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC-2077-390741603) via Excellence Chair Victoria Orphan. G.W. was funded by DFG under Germany's Excellence Strategy-EXC-2077-390741603 and the Max Planck Society.

Author information

Author notes

1. Rafael Laso-Pérez

Present address: Systems Biology Department, Centro Nacional de Biología (CNB-CSIC), Madrid, Spain

2. These authors contributed equally: Zhuo Zhou, Cui-jing Zhang, Peng-fei Liu, Lin Fu

Affiliations

1. Key Laboratory of Development and Application of Rural Renewable Energy, Biogas Institute of Ministry of Agriculture and Rural Affairs, Chengdu, China

Zhuo Zhou, Peng-fei Liu, Lin Fu, Lu Yang, Li-ping Bai, Jiang Li, Min Yang & Lei Cheng

2. Archaeal Biology Center, Institute for Advanced Study, Shenzhen University, Shenzhen, China

Cui-jing Zhang & Meng Li

3. Center for The Pan-Third Pole Environment, Lanzhou University, Lanzhou, China

Peng-fei Liu

4. MARUM, Center for Marine Environmental Sciences, University Bremen, Bremen, Germany

Rafael Laso-Pérez & Gunter Wegener

5. Max Planck Institute for Marine Microbiology, Bremen, Germany

Rafael Laso-Pérez & Gunter Wegener

6. Key Laboratory of Microbial Enhanced Oil Recovery, SINOPEC, Dongying, China

Jun-zhang Lin & Wei-dong Wang

Contributions

L.C. and M.L. initiated the study. L.C., M.L., G.W. and P.-f.L. designed research. J.-z.L., W.-d.W. and Z.Z. collected the oily sludge samples. Z.Z., J.L., M.Y. and L.C. conducted cultivation experiments. Z.Z. and L.Y. performed oil analysis. C.-j.Z., P.-f.L., Z.Z., R.L.-P. and M.L. performed all bioinformatics analyses. R.L.-P. and L.C. designed CARD-FISH probes, and R.L.-P. performed CARD-FISH and cell visualization. L.F., L.C. and L.-p.B. performed metabolite analyses. P.-f.L., R.L.-P., G.W., M.L. and L.C. analysed data and wrote the manuscript with contributions from all of the co-authors.

Corresponding authors

Correspondence to [Gunter Wegener](#), [Meng Li](#) or [Lei Cheng](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Guillaume Borrel, Rudolph Thauer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Molecular characterization of the methanogenic oily sludge incubated at different temperatures.

a, Accumulation of methane in the headspace of treatments at different temperatures over an incubation time of 301 days. The estimates of reported methane production rates base on the time interval for the formation of 5% and 90% of the maximum methane formation. **b**, Mass spectrometric analysis of extracted residual oil for *n*-alkanes $m/z = 85$, *n*-alkylcyclohexanes $m/z = 82$, *n*-alkylbenzenes $m/z = 92$. Exemplary data of the 55 °C culture is presented in Figs. 1b–1d. Data shown are mean ± standard deviation ($n = 3$ biologically independent replicates). **c** and **d**, Archaeal and bacterial community structure revealed by amplicon sequencing in the different temperature treatments after 204 days of incubation, respectively. Only families with relative abundances $\geq 1\%$ are shown. “Other” indicates the sum of groups with relative abundance $< 1\%$. Data shown are mean - standard deviation ($n = 3$ biologically independent replicates).

Extended Data Fig. 2 Epifluorescence micrographs of different community members of the oily sludge.

a-c, Visualization of archaea (green) and bacteria (red). **d-f**, Visualization of ‘*Ca. Methanoliparum*’ (green) and archaea (red). Hybridization of ‘*Ca. Methanoliparum*’ with the general archaeal probe and the specific DC06-660Mlp probe. The vast majority of archaea hybridized also with the probe for ‘*Ca. Methanoliparum*’. **g-i**, Visualization of ‘*Ca. Methanoliparum*’ (green) and bacteria (red). Oligonucleotide probes were ARCH-915 for archaea, EUB388 I-III for bacteria and DC06-660Mlp for ‘*Ca. Methanoliparum*’. Three representative recorded images from $n = 3$ independent samples (**a-i**, 9 rows of images in total) of one culture are shown. Scale bars in all images are 10 μm .

Extended Data Fig. 3 Phylogenetic analyses of MAGs and 16S rRNA gene sequences of ‘*Ca. Methanoliparia*’.

a, Phylogenomic analyses of ‘*Ca. Methanoliparia*’ MAGs based on the concatenated alignments of 16 ribosomal proteins⁶⁷. Bootstrap values > 0.95 are marked with grey dots, ‘*Ca. Bathyarchaeota*’ set as outgroup. The maximum-likelihood tree was constructed by using the IQ-TREE software with the parameters ‘-m WAG -bb 1000’. **b**, Phylogenetic analysis of 16S rRNA gene sequences retrieved from all ‘*Ca. Methanoliparia*’ MAGs. For MAG- derived sequences source information is given: i.e., T55 indicates temperature of the culture (55 °C) and after the MAGs (bin) number the substrate used is indicated (e.g., *n*-hexadecane). The asterisk (*) marking ‘*Ca. M. whitmanii*’ sequence identifiers indicates 16S rRNA genes that were truncated during assembly. In these cases, the longest partial sequence was used for the phylogenetic analyses. The 16S rRNA gene sequences were added to the consensus tree with ‘quick add’ option, thus no bootstrap values are available.

Extended Data Fig. 4 Identities between ‘*Ca. Methanoliparum*’ clusters and phylogenetic analysis of their AcrA and McrA protein sequences.

a, Identities of the 16S rRNA gene. **b**, Genome based average Amino Acid Identity (AAI). **c**, Genome based Average Nucleotide Identity (ANI). **d**, Identity based on the percentage of conserved proteins (POCP). All matrices consistently showed that all ‘*Ca. Methanoliparia*’ MAGs from this study grouped into four species-level clusters within the genus ‘*Ca. Methanoliparum*’. In the box plots the central line represents the median; the lower and upper box limits correspond to the 25th and 75th percentiles, respectively; Numbers represent the times of pairwise comparisons of MAGs between two groups. Cluster 1 (C1): ‘*Ca. M. thermophilum*’; Clusters 2 (C2): ‘*Ca. M. widdelii*’; Cluster 3 (C3): ‘*Ca. M. whitmanii*’; Cluster 4 (C4): ‘*Ca. M. zhangii*’. Mv indicates the genomes of the sister marine clade ‘*Ca. Methanoliviera*’. **e**, Maximum-likelihood tree of the protein sequences of AcrA and McrA present in ‘*Ca. Methanoliparum*’ MAGs retrieved in the present studied. Different colours indicate the different ‘*Ca. Methanoliparum*’ species. Numbers in parenthesis indicate the number of *acrA/mcrA* sequences detected in the different metagenomes. In each MAG, maximum one *acr* and one *mcr* were detected. Trees were

constructed by using IQ-TREE with the parameters ‘-m WAG, -bb 1000’, with bootstrap values >0.95 shown in grey dots.

Extended Data Fig. 5 Gene clusters associated with the alkane degradation and methanogenesis pathways detected in the representative MAGs of the four ‘Ca. Methanoliparum’ species.

Several copies of *fadD* and ACADM were detected and only copies with the highest transcript abundances are shown. In orange, alkane activation and conversion to a fatty acid; in blue, beta oxidation pathway and in red, the ACS/CODH complex and the methanogenesis pathway. Details of all copies are included in Supplementary Table 6.

Extended Data Fig. 6 Relative transcript abundances of alkane-degrading and methane-producing pathways coding genes.

The colour code shows the $\log_2(\text{FPKM})$ values of each gene. For enzymes or subunits with several putative coding genes, only the ones with the highest level of $\log_2(\text{FPKM})$ are shown here. Two samples were taken for cultures with *n*-hexadecane addition (Hex.) at day 31 and 55, while sampling at one time point (day 55) with 3 replicates (designated as r1-r3) was performed for control cultures without *n*-hexadecane amendment (Con.). Grey cells indicate that the corresponding genes were not found in the MAGs. Details of all copies are included in Supplementary Table 7.

Extended Data Fig. 7 Identification of coenzyme M derivatives in cultures by HPLC-MS/MS based on the corresponding retention times.

a and **b**, hexadecyl-CoM and the corresponding 3 characterized fragments (in blue) in cell extracts from cultures with hexadecane ($C_{16}H_{34}$) addition. **c** and **d**, eicosyl-CoM and 3 characterized fragments (in blue) in cell extracts from cultures with eicosane ($C_{20}H_{42}$) additions. Standard appears in black primary anions and second anions (produced by fragmentation) detected in

hexadecane and eicosane cultures showed the same retention time as the synthetic standards of hexadecyl-CoM and eicosyl-CoM, respectively.

Extended Data Fig. 8 Identification of coenzyme M derivatives in cultures incubated with specific hydrocarbons.

a, Scheme for the activation of long-chain alkanes and alkyl-substituted compounds as CoM thioethers in ACR, and their expected fragmentation patterns. The residual ‘R-’ describes a methyl-, cyclohexane- or aromatic unit with an alkyl chain C_nH_{2n+1} for $n \geq 13$. Dash arrows and numbers above indicate the fragmentation positions. **b** and **c**, QE Plus-Orbitrap MS analyses of cultures supplemented with eicosane resulted in a mass peak of eicosyl-CoM ($C_{20}H_{41}-SC_2H_4SO_3^-$ at $m/z = 421.28162$ and the fragments eicosyl-thiol ($C_{20}H_{33}S^-$, $m/z = 313.29373$), ethenesulfonate ($C_2H_3SO_3^-$, $m/z = 106.98092$) and bisulfite (HSO_3^- , $m/z = 80.96519$). All peaks match those of an eicosyl-CoM standard. **d-i**, QE Plus-Orbitrap MS analyses of cultures supplemented with a mixture of *n*-docosane ($C_{22}H_{46}$), *n*-hexadecyl benzene ($C_{22}H_{38}$) and *n*-hexadecyl cyclohexane ($C_{22}H_{44}$) as substrates, and detection of **d** and **e** docosyl-CoM ($C_{24}H_{49}S_2O_3^-$, $m/z = 449.31134$) with the fragment $C_{22}H_{45}S^-$ ($m/z = 341.32495$); of **f** and **g** *n*-hexadecyl benzene coenzyme M ($C_{24}H_{41}S_2O_3^-$, $m/z = 441.25064$) with the predicted fragment $C_{22}H_{37}S^-$ ($m/z = 333.26212$) and of **h** and **i** *n*-hexadecyl cyclohexane CoM ($C_{24}H_{47}S_2O_3^-$, $m/z = 447.29730$) with the fragment $C_{22}H_{43}S^-$ ($m/z = 339.30939$). The mass error for all mass peaks shown here are < 5 p.p.m.

Extended Data Fig. 9 Semi-continuous cultivation of the ‘Ca. Methanoliparum’ cultures at 55 °C.

Microorganisms were cultured using a mixture of *n*-docosane, *n*-hexadecyl benzene, *n*-hexadecyl cyclohexane as substrate. The culture was transferred when 15 to 20 mmol of methane were formed, and 30% to 50% of the culture were transferred. Displayed are transfers 3 to 6. **a**, Methane formation in the headspace. Grey arrows indicate transfer events. **b** and **c**, Abundance of 16S rRNA gene of ‘*Ca. Methanoliparum*’ and bacteria as

determined by qPCR, respectively. **d**, Relative abundance of main archaeal groups determined by 16S rRNA gene sequencing with archaeal primer set Arch519F/Arch915R.

Extended Data Fig. 10 Proposed metabolic pathway and related gene clusters for benzene-CoA degradation in ‘Ca. Methanoliparum’.

a, Gene clusters found in the four representative MAGs with potential for benzoyl-CoA degradation. Numbers in the gene clusters indicate kilobases. **b**, Annotations and Locus tag for the corresponding genes shown in panel **a** that are found in the representative MAG of ‘*Ca. M. thermophilum*’ (XY_C20_T55_P2_bin.5 of Cluster 1). **c**, Proposed pathway for the degradation of benzoyl-CoA based on the pairwise comparison of the candidate genes of ‘*Ca. Methanoliparum*’ (red) with the genes involved in benzoyl-CoA degradation in the model organisms *Thauera aromatica* (green) and *Rhodopseudomonas palustris* (blue). The letters for candidate genes of ‘*Ca. Methanoliparum*’ refer to the letters indicated in the panel **a** (see Supplementary Table 10 for more details).

Supplementary information

Supplementary Information

Supplementary Figs. 1–6 and the legends for Supplementary Tables 1–13.

Reporting Summary

Supplementary Tables

Supplementary Tables 1–13.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Zhou, Z., Zhang, Cj., Liu, Pf. *et al.* Non-syntrophic methanogenic hydrocarbon degradation by an archaeal species. *Nature* **601**, 257–262 (2022). <https://doi.org/10.1038/s41586-021-04235-2>

- Received: 21 December 2020
- Accepted: 10 November 2021
- Published: 22 December 2021
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04235-2>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [**A microbe that uses crude oil to make methane**](#)
 - Guillaume Borrel

Nature (2022)

A microbe that uses crude oil to make methane

- Guillaume Borrel

News & Views 22 Dec 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04235-2>

| [Section menu](#) | [Main menu](#) |

- Article
- Open Access
- [Published: 22 December 2021](#)

Cancer risk across mammals

- [Orsolya Vincze](#) [ORCID: orcid.org/0000-0001-5789-2124](#)^{1,2,3,4},
- [Fernando Colchero](#) [ORCID: orcid.org/0000-0001-8613-4568](#)^{5,6,7},
- [Jean-Francois Lemaitre](#)⁸,
- [Dalia A. Conde](#) [ORCID: orcid.org/0000-0002-7923-8163](#)^{6,7,9},
- [Samuel Pavard](#)¹⁰,
- [Margaux Bieuville](#)¹⁰,
- [Araxi O. Urrutia](#)^{11,12},
- [Beata Ujvari](#) [ORCID: orcid.org/0000-0003-2391-2988](#)¹³,
- [Amy M. Boddy](#)¹⁴,
- [Carlo C. Maley](#) [ORCID: orcid.org/0000-0002-0745-7076](#)¹⁵,
- [Frédéric Thomas](#) [ORCID: orcid.org/0000-0003-2238-1978](#)¹ &
- [Mathieu Giraudeau](#)^{1,2}

[Nature](#) volume 601, pages 263–267 (2022)

- 37k Accesses
- 913 Altmetric
- [Metrics details](#)

Subjects

- [Cancer](#)
- [Evolution](#)
- [Zoology](#)

Abstract

Cancer is a ubiquitous disease of metazoans, predicted to disproportionately affect larger, long-lived organisms owing to their greater number of cell divisions, and thus increased probability of somatic mutations^{1,2}. While elevated cancer risk with larger

body size and/or longevity has been documented within species^{3,4,5}, Peto's paradox indicates the apparent lack of such an association among taxa⁶. Yet, unequivocal empirical evidence for Peto's paradox is lacking, stemming from the difficulty of estimating cancer risk in non-model species. Here we build and analyse a database on cancer-related mortality using data on adult zoo mammals (110,148 individuals, 191 species) and map age-controlled cancer mortality to the mammalian tree of life. We demonstrate the universality and high frequency of oncogenic phenomena in mammals and reveal substantial differences in cancer mortality across major mammalian orders. We show that the phylogenetic distribution of cancer mortality is associated with diet, with carnivorous mammals (especially mammal-consuming ones) facing the highest cancer-related mortality. Moreover, we provide unequivocal evidence for the body size and longevity components of Peto's paradox by showing that cancer mortality risk is largely independent of both body mass and adult life expectancy across species. These results highlight the key role of life-history evolution in shaping cancer resistance and provide major advancements in the quest for natural anticancer defences.

[Download PDF](#)

Main

Complex multicellular organisms are built of millions to quadrillions of cells, ultimately all being derived from a single cell, the zygote. During the course of the organisms' lifetime and owing to various mutational processes, cell lineages tend to accumulate mutations^{7,8}. While the majority of mutations are harmless, some enable cells to escape cell cycle control, to grow and proliferate uncontrollably, resulting in cancer^{9,10}. Cancer is a multistage process, where a set of mutations is required for both initiation and malignant progression. Given that every cell division carries a risk of generating mutations, organisms with large bodies (composed of more cells) and extended longevities (with a longer time to accumulate mutations) should be more likely to develop cancer^{1,6,11,12}. Indeed, within humans^{3,11} and dogs⁴, larger individuals are more likely to develop cancer than smaller ones. Similarly, increasing age is one of the most potent carcinogenic factors in species in which cancer aetiology is well studied. Yet, while current evidence suggests that large body size and extended longevity result in increased cancer risk within species, this relationship may not hold across taxa¹³.

Limited data available so far indicate that vertebrates do not face clear size-dependent cancer risks despite their size and longevity varying by orders of magnitude. This poses a logical challenge, first formulated by Sir Richard Peto^{6,14}. He noted that although mice have approximately 1,000 times fewer cells and >30 times shorter lifespans than humans, their risk of carcinogenesis is not markedly different (coined as

Peto's paradox)⁵. Peto's paradox is an evolutionary conundrum that has puzzled the scientific community and has led to lively debate regarding the evolution of anticancer mechanisms. It is often postulated that natural selection on large size or extended longevity is inherently inseparable from the evolution of anticancer defences.

Knowledge gained from investigating Peto's paradox might thus largely contribute to our knowledge on natural anticancer mechanisms that could potentially be harnessed for medical use. Further, understanding cross-species variation of cancer vulnerability is an important next step in animal health and welfare. While a few studies aimed to establish cross-species variation in cancer risk^{15,16}, most estimates and analyses have considerable limitations. These include small cross- or within-species sample sizes^{1,6,17}, lacking information on the age distribution of cancer^{15,16,17}, data heterogeneity (for example, biases due to domestication¹⁷ or combining data from multiple taxa^{17,18}) or lack of control for phylogenetic relatedness among species¹⁷. Moreover, the effect of longevity was generally tested using the much-debated metric of maximum reported lifespan^{15,16,17}. Nonetheless, cancer prevalence (the parameter exclusively considered by earlier studies^{15,16,17}) is expected to correlate with life expectancy (the average time lived by individuals in the population of interest), not maximum lifespan potential (that veryfew individuals achieve), making these analyses inherently flawed.

To characterize cancer incidence in a homogeneous sample and across a wide taxonomic range, here we used the Zoological Information Management System (ZIMS), managed by Species360 (a non-profit organization custodian of zoo and aquarium data)¹⁹. We assembled information on 110,148 adult non-domesticated mammals distributed over 191 species, including data on their age, sex, dead/alive status and postmortem pathological records for 11,840 individuals. Cancer is registered in this database only for deceased animals and only if the inspecting veterinary pathologist considered it to be a factor that contributed to the individual's death. First, to characterize species longevities, we used survival modelling ($n = 110,148$) and calculated species-specific adult life expectancies, representing average adult longevity in our sample^{20,21}. Second, to estimate cancer mortality risk, we used two metrics, both estimating the proportion of individuals dying of cancer. We first calculated a simple measure of cancer mortality risk (hereafter CMR; that is, the ratio between the number of cancer-related deaths and the total number of individuals whose postmortem pathological records were entered in the database, $n = 11,840$), a measure adopted by earlier comparative studies^{15,17}. This measure relies solely on dead individuals, ignoring the incomplete records of live animals, potentially introducing bias in cancer mortality estimates. Therefore, we also calculated the cumulative incidence of cancer mortality (hereafter ICM), a metric of cancer mortality risk eliminating potential biases due to disregarding left-truncation (that is, cancer before individuals enter the study) and right-censoring (individuals alive, thus with unknown fate at data extraction). Using these two metrics, we explored the

phylogenetic distribution of cancer-related mortality across mammals. We then investigated Peto's paradox and tested whether cancer mortality risk is associated with body size or the mean number of years lived by adults in the explored populations (that is, adult life expectancy).

Cancer across the mammalian phylogeny

CMR was highly variable among species, ranging from 0% (in 47 species out of 191) to 57.14% in the kowari (*Dasyuroides byrnei*). CMR exceeded 10% in 41 species (21.5% of all species inspected), indicating that the oncogenic process is a prevailing source of mortality of many mammalian species distributed along the phylogeny, at least in managed populations (Fig. 1). ICM showed strong consistency with CMR (Pearson correlation test, $r=0.89$, $t=25.14$, $df=170$, $P<0.0001$). Nonetheless, all models were performed using both metrics to test for consistency in the results.

Fig. 1: Distribution of cancer mortality risk across the mammalian phylogeny.



a, CMR in various mammals (scale to bar plots is provided on the left of the graph). **b**, Violin plots indicating order differences in CMR in orders with a minimum of two species represented. Solid black lines represent order medians. Animal silhouettes used to visually represent mammalian orders were downloaded from PhyloPic (<http://www.phylopic.org>).

Cross-species variation in cancer risk showed strong phylogenetic signal²² (CMR: $n=191$, $\lambda=0.87$, $P<0.0001$; ICM: $n=172$, $\lambda=0.69$, $P<0.0001$). To explore this, we

compared cancer risk among mammalian orders represented by at least two species using linear regressions. Results indicated that the phylogenetic signal was mostly driven by cancer mortality risk in Carnivora, which was significantly higher than in Primates or in Artiodactyla (Extended Data Table 1, Fig. 1 and Extended Data Fig. 1). Both cancer mortality risk metrics indicated that Artiodactyla is the least cancer-prone mammalian order, despite the frequency of large-bodied species in this group (Extended Data Table 1).

High cancer risk in managed populations of Carnivora has previously been reported^{23,24}. Possible explanations include the use of hormonal contraception (for example, progestins) and pregnancy postponement in zoo carnivores, both being significant risk factors for certain cancers in humans as well as non-domestic felids^{24,25,26}. Nonetheless, if contraception was the key factor driving elevated cancer risk in Carnivora, a significant sex bias in cancer risk would be expected in this group, because hormonal contraception is usually administered to females. To test for sex bias in cancer risk across Carnivora, we estimated sex-specific CMR and ICM (only species with a minimum of ten males and ten females with available postmortem pathological records: $n = 36$ and $n = 30$ species, respectively). Pairwise comparison between sexes revealed no sex bias in either measure of cancer mortality risk (phylogenetic paired *t*-tests, CMR: $t = 0.52$, $df = 33$, $P = 0.6061$; ICM: $t = -0.6815$, $df = 27$, $P = 0.5014$) (ref. 23). Therefore, the generally high cancer risk in Carnivora is unlikely to be driven solely by the carcinogenic effects of reproductive management in zoo populations.

A high-fat, low-fibre diet, a known risk factor for carcinogenesis, has also been suggested to explain the elevated cancer risk in Carnivora^{26,27}. Moreover, carnivores are on the top of the food chain, exposing them to bio-magnified effects of carcinogenic compounds²⁸, such as pollutants²⁶. Importantly, the consumption of raw meat can also expose carnivores to pathogens that can drive oncogenic transformation²⁹. For instance, in humans it was estimated that 10–20% of all cancers are of viral origin³⁰. While this figure is unknown in any other animal species²⁹, it is arguable that raw meat consumption might exacerbate the spread of carcinogenic pathogens³¹. Exploring the association between diet and cancer risk could help to disentangle the influences of these risk factors.

To explore the link between carnivorous diet and cancer risk, we collected data on the species' natural diet (that is, consumption of animals, including invertebrates or vertebrates, and specifically of fish, reptiles, birds and mammals) from the literature³². Phylogenetic generalized least-squares (PGLS) regressions controlling for differences in longevity and body mass run separately for each diet item (Fig. 2, Extended Data Figs. 2 and 3 and Extended Data Table 2) indicated that species with animal-based diets have comparable cancer mortality risks (both CMR and ICM) to species that

rarely or never consume animals. Nonetheless, consumption of vertebrate but not invertebrate prey was associated with increased cancer mortality risk. Specifically, mammals frequently consuming mammalian prey had significantly higher cancer mortality risk compared to mammals that rarely or never consume other mammals. Similar differences could not be detected in the case of fish, reptile or bird prey frequencies. These results indicate that a carnivorous diet has significant costs in terms of heightened oncogenic predisposition across mammals, particularly for diets high in mammalian prey. The nonsignificant association between cancer mortality risk and diet content of invertebrate, fish, reptile or bird indicates bio-magnification as a less likely source of elevated cancer risk among Carnivora. Nonetheless, the limited number of species primarily consuming these preys in our sample does not allow for definitive conclusions regarding this hypothesis. By contrast, the result that mammals consuming other mammals appear to have the highest cancer risk of all diet categories is consistent with a pathogenic origin of elevated cancer mortality risk among Carnivora. Host jumping of pathogens is most likely to occur in the case of phylogenetic proximity between the reservoir prey and the predator species³³, making a mammal-to-mammal transmission the most likely host jump scenario. These results suggest that pathogen-driven oncogenesis might have a considerable role in shaping cancer mortality risk in mammals and urges the search for pathogens in various cancer types, while considering the notorious difficulty of proving oncogenic properties of pathogens³⁰. Alternatively, high cancer risk in carnivorous animals might be related to their low microbiome diversity³⁴, limited physical exercise under human care or other aspects of their physiology. Nonetheless, a lack of bias in these results, caused by potential alterations in the diet of housed carnivores, should be confirmed by studying natural populations. Importantly, these results probably reflect a complex, maybe indirect evolutionary link between diet and cancer vulnerability; therefore, the effect of meat consumption on cancer risk should be interpreted with caution.

Fig. 2: Cancer mortality risk in mammals as a function of animal content in diet.



Violin plots show CMR as a function of seven diet items, each coded as rarely/never occurring in the diet or representing the primary/secondary food item of the species. Medians are marked with solid black lines. *P* values indicate pairwise differences as indicated by models presented in Extended Data Table 2 that also control for body mass, life expectancy and phylogeny.

Test of Peto's paradox

Owing to the large number of zero cancer mortality risk estimates and thus non-Gaussian distributions, cancer risks were analysed using zero-inflated phylogenetic models ([Methods](#)), as a function of sample size, body mass and life expectancy. The probability of detecting at least one individual with cancer in a species increased steeply with increasing number of individuals with available postmortem pathological records (Extended Data Table 3). In fact, cancer was detected in at least one individual in almost all species with more than 82 individual pathological records available (Extended Data Figs. 4 and 5). Exceptions were the blackbuck (*Antilope cervicapra*) and the Patagonian mara (*Dolichotis patagonum*), where no cancer was detected despite postmortem pathological records being available for 196 and 213 individuals, respectively. This highlights the quasi-universality of oncogenic phenomena across mammals, illustrating that with adequate sampling, cancer is likely to be detected in all mammals. Our results further emphasize the fact that some members of the order Artiodactyla, besides rodents, are particularly cancer-resistant. Rodents have long been subject to scrutiny in the search for natural cancer resistance mechanisms³⁵, owing to notoriously low cancer incidence in some species³⁶. Nonetheless, cancer mortality risk in our dataset was lowest among ruminants, complying with rare cancer case reporting in this taxonomic group^{15,37}. This indicates that other mammalian groups, especially Artiodactyla, might serve as informative model organisms in cancer research.

The probability of detecting cancer (CMR: $n = 188$; ICM: $n = 141$) (Extended Data Table 3 and Extended Data Figs. 4 and 5), as well as non-zero cancer mortality risk (CMR: $n = 141$; ICM: $n = 128$) (Extended Data Table 3 and Fig. 3), tended to decrease with larger body masses and to increase with longer life expectancies. These effects were not significant, and were consistent between the two cancer mortality metrics. These associations were also largely independent of each other, and were similar in single-predictor models (Supplementary Table 3). The effect of body mass on cancer risk was even slightly negative, but the models indicated only a 2.8–2.9% decrease in cancer risk for a doubling of the body weight (for example, CMR changes from 3.82% (1 kg) to 3.71% (2 kg), or from 2.86% (1,024 kg) to 2.78% (2,048 kg), respectively, predictions obtained from the model presented in Extended Data Table 3, with an arbitrary life expectancy of 27 years, Extended Data Fig. 6). Additionally, body mass accounted for only 0.78% of the cross-species variance in CMR (that is, partial coefficient of determination). Similarly, cancer mortality risk only minimally and

nonsignificantly increased with higher life expectancy, indicating a 24.7–25.2% increase in cancer risk for a doubling of the adult life expectancy (for example, CMR changes from 0.89% (1 year) to 1.18% (2 years), or from 2.80% (16 years) to 3.72% (32 years), respectively, predictions obtained from the model presented in Extended Data Table 3 with an arbitrary body mass of 10 kg, Extended Data Fig. 6). Adult life expectancy accounted for only 2.94% of the variance observed in CMR (that is, partial coefficient of determination). Overall, these results provide the largest-scale and most robust support to the body size and life expectancy components of Peto's paradox in mammals. They suggest that lifespan extension and larger body size jointly evolved with better anticancer mechanisms across mammals.

Fig. 3: Association between cancer mortality risk and body mass or adult life expectancy across mammals.

 figure3

a, b, Non-zero CMR plotted against body mass (**a**) or adult life expectancies (**b**). Slopes were obtained from the PGLS model presented in Extended Data Table 3a. Points are proportional to the log number of individuals with available postmortem pathological records.

Since the first indication of species differences in cancer predisposition, an intense search has been conducted to identify mechanisms explaining cancer resistance in certain species, mostly rodents³⁵, and very large animals¹⁷. Although these studies demonstrated key species-specific anticancer mechanisms^{17,28,35}, a considerable gap remains in our knowledge on the taxonomic and phylogenetic diversity and distribution of tumour-suppressing mechanisms. Our results provide a solid foundation for future studies scrutinizing these questions, by providing information on the generality and frequency of oncogenic phenomena across the mammalian phylogeny.

We also highlight the exceptional resources provided by zoos for studies of cancer in wildlife^{[35,36](#)}.

Our study indicates that death due to oncogenic phenomena is frequent and taxonomically widespread in mammals. In some species more than 20–40% of the managed adult population die of cancer-related pathologies. This estimate is staggering, especially knowing that cancer incidences estimated here are conservative ([Methods](#)). This observation urges the extensive exploration of cancer in wildlife, especially in the context of recent environmental perturbations^{[38](#)}, as serious threats to animal welfare^{[29](#)}.

Methods

Documenting cancer in wild animals is extremely challenging in most cases owing to the lack of information on the age of individuals, the difficulty retrieving the bodies for necropsy and the likelihood of cancer negatively influencing survival before cancer itself could be detected. Although data on cancer incidence from wild populations would be indispensable to describe natural incidences of malignancies, such data, especially with corresponding ages and demographic histories, are unfortunately still far from our reach. Therefore, to estimate cancer mortality risk, we used data provided by Species360 and the Zoological Information Management System (ZIMS, Data Use Approval Number 73836), an international non-profit organization that maintains a real-time and centralized database of animals under human care (regrouping information from over 1,200 zoos worldwide). Although we recognize that the interpretation of data gathered on zoo animals requires caution, owing to strong human control on the diet, health, mortality factors, environment or standard biological functions of the animals, zoos provide exceptionally high data resolution on the demography and cause of death for a wide range of species. Here we rely on the high probability of body retrieval of deceased zoo animals and the necropsy routinely performed on most of them (unless found in an advanced stage of decomposition), aiming to identify the most likely pathology causing the death of the animal. These examinations are likely to reveal most solid tumours, but (although possible) benign tumours, liquid tumours (for example, leukaemia) or early-stage cancers are unlikely to be recorded here, either owing to their diagnostic difficulty or their perceived limited role in contributing to the death of the animal.

Specifically, here we use the husbandry module of ZIMS, providing information on birth, death, sex and pre-defined categories of pathological findings, including neoplasia (by definition tumours that contributed to the death, albeit with no option to specify the cancer type or other details). No statistical methods were used to predetermine sample size, but to minimize bias caused by potential temporal heterogeneity in data management practices and necropsy record-keeping^{[15](#)}, here we

focused on individuals alive or born after 1 January 2010 (data extraction: 30 May 2020). This sample was then used to characterize species-specific life expectancies and cancer incidence, but only after the exclusion of data that did not fulfil a series of criteria, to ensure the highest and most homogeneous data quality possible. First, cancer is an age-related disease that rarely occurs in juveniles, and pediatric cancers are usually medically distinct from adults' cancers. As such, infant mortality differences observed across species can significantly confound cancer incidence estimates. Therefore, we gathered sex-specific or species-specific (wherever the former was not available) ages at sexual maturity and we considered individuals for analyses only if they reached maturity before or during our sampling period. For individuals of unknown sex (about 12% of all individuals in the raw data extract), the maximum of ages at sexual maturity of males and females was used as an inclusion age threshold. Sex-specific age at sexual maturity for each species was obtained from Conde et al.¹⁹ or from published literature resources (see data sources in <https://github.com/OrsolyaVincze/VinczeEtal2021Nature/blob/main/SupplementaryData.xls>). Second, given that age is a key predictor of cancer emergence, we considered only individuals for which date of birth was recorded precisely or within a narrow (maximum 30 days) time interval. Third, we considered only species in which postmortem pathological records were available for at least 20 adult individuals, irrespective of the cause of death (for example, infection, accident, geriatric disease and so on). Nonetheless, models presented were performed with increased thresholds of 40, 60, 80 and 100 individuals to check for consistency in the results ([Supplementary Table 2](#)). Fourth, given that the process of domestication is widely regarded as a major contributing factor to inbreeding depression and higher incidence of cancer³⁹, we excluded all species that were subject to domestication as well as their wild ancestors (taxa excluded owing to being subject to domestication are listed in [Supplementary Table 1](#)). Following these restrictions, data extraction on age and cause of death resulted in information for 110,148 (62,556 live and 47,592 dead) individuals ($n = 191$ species). For calculation of ICM, we included only species in which survival is correctly estimated until old ages (that is, data allowing the estimation of age-specific survival until the age at which only 10% of individuals are surviving, $n = 172$ species). While these restrictions removed multiple sources of bias in our cancer mortality risk estimates, we cannot exclude the possibility that some species (for example, more charismatic ones) are subject to more frequent or more detailed necropsies. Nonetheless, our statistical approach, especially the complete case analysis, is largely insensitive to such biases, as individuals not having available postmortem diagnostic records are considered censored (see below). Also, while the depth of necropsy might vary slightly among species, neoplasia that had a significant contribution to the death of the animals (the focus of our study) are generally detected even at gross necropsies. Additionally, large species are considered of key importance for zoos, also reflected by the fact that the proportion of dead individuals with postmortem pathological records is larger in larger species (Pearson's correlation: $r =$

0.24 , $t = 3.35$, $df = 189$, $P = 0.001$). Accordingly, if charisma had a role in cancer detection, we would expect a larger cancer risk in large mammals, opposite to the (nonsignificant) negative body mass effect in our models. Consequently, we believe that charisma is unlikely to represent a major source of bias in our analysis.

Estimation of adult life expectancy

As we have no reason to believe that censored individuals would not have the same prospect of survival as those who continue to be followed, we estimate adult life expectancy from age-specific survival estimated using the Kaplan–Meier procedure (using the `survfit` function in the R package `survival`⁴⁰). Individuals older than their age at sexual maturity on 1 January 2010 were left-truncated at their age at this date; individuals reaching sexual maturity after this date were left-truncated at their age at sexual maturity. Individuals still alive at the time of data extraction were considered right-censored (samples per species varied from 42 to 5,816 individuals), while known fate individuals were assigned as dead ($n = 47,592$), irrespective of whether their cause of death was specified or not.

Estimation of ICM

ICM was calculated using a competing risk approach, based on the cumulative hazard of cancer-related deaths and survival probability of the species under human care. First, age-specific survival S_x , at age x , was estimated from KM analysis as above. However, here we performed a complete-case analysis, using only 11,840 individuals for which the cause of death was specified together with right-censored survivors. Complete-case analysis assumes that missingness in the cause of failure is random, but we had no reason to believe that this was not the case in our dataset. Postmortem examinations are routinely carried out on most recovered bodies in zoos, and once examinations are performed the results are equally likely to be entered in the database irrespective of the pathologies identified. ICM estimates were thus based on $n = 74,396$ individuals, $n = 179$ species. Second, the cancer mortality hazard h_x^c was estimated using a KMx1 analysis where only deaths by cancer were incorporated as a death event. ICM is then such that

$$\text{\$}\$\{\text{\rm\{I\}}\}\{\text{\rm\{C\}}\}\{\text{\rm\{M\}}\}\text{=}\text{\rm\mathop{\sum }}\limits_{\text{x=\alpha }}^{\text{x=\infty }}\{\text{\rm\{S\}}\}_{\text{x}}\{\text{\rm\{h\}}\}_{\text{x}}^{\text{c}}\{\text{\rm\{c\}}\}\text{\$}\$$$

where α is the age at sexual maturity. The only difference with classic estimation is that we extracted h_x^c (and S_x) for each time unit with discrete jumps (and falls) at event times at age t and with $h_{t \leq x < t+1}^c = 0$ (S_x constant) between these events; instead of estimating discrete hazard $\{(h)_t^c\} = \{d\}_t / \{n\}_t$ on

these time intervals (where d_t^c is the number of deaths by cancer within the interval and n_t is the number of survivors at the beginning of the interval). We chose this method to reflect true variation in the data for the interspecific comparison where species differ greatly in the number of events and time interval between these (sometimes a third of the organism's adult lifespan), a situation rarely met when comparing human groups.

Covariates and statistical analyses

For each species, we obtained sex-specific adult body mass data from Species360's ZIMS (see <https://github.com/OrsolyaVincze/VinczeEtal2021Nature/blob/main/SupplementaryData.xls>). Species-specific body mass was calculated as the average of all body mass measurements recorded in the ZIMS database in adults, while species-specific values were obtained by averaging the body masses of males and females. These were calculated only in species for which there were at least 100 adult body mass records; otherwise, body mass was taken from the literature and database review by Conde et al.¹⁹. We verified that there was a one-to-one correspondence in the body mass information for species with records in both datasets.

Diet information was obtained from a global diet dataset for terrestrial mammals³², providing information on diet composition at four hierarchical levels of food items (never consumed, occasionally consumed, secondary food item, primary food item). We collected information on animal content in diet, as well as subcategories of this diet class, namely invertebrate or vertebrate consumption, as well as specifically fish, reptile, bird and mammal consumption. Given that most food items had few species in the intermediate levels (occasional consumption and secondary food item), we re-categorized the diet variables at two levels: never/rarely consumed or representing the primary/secondary food item of the species. The effect of diet was tested in PGLS regressions using only species with non-zero cancer mortality risks. Models were run separately for each food item that were entered to a base model including body mass and adult life expectancy as covariates. Results are shown in Extended Data Table 2.

To account for statistical non-independence due to phylogenetic relationships, we obtained a sample of 1,000, equally plausible phylogenetic trees, from the posterior distribution published by Upham et al.⁴¹, covering 5,911 species. We then obtained a rooted consensus tree using the sumtrees Python library⁴². Two species recently raised to species level were manually added to the tree as sister taxa of the species it was recently separated from (that is, *Cervus canadensis* to *Cervus elaphus* and *Gazella marica* to *Gazella subgutturosa*). Phylogenetic signal of cancer risk was assessed using the function phylosig from the R package phytools²². Partial coefficients of

determinations were calculated using the function R2.pred from the R package rr2 (ref. [43](#)), based on models presented in Extended Data Table [3](#).

Models of cancer risk testing Peto's paradox were performed using zero-inflated logistic models, which allow us to make inferences on the probability of detecting at least one cancer case in the species and, given that cancer was detected, inferences on the CMR or ICM. Therefore, the first part of this consisted of a phylogenetic binomial regression (using the function binaryPGLMM, in the R package ape^{[44](#)}), where the dependent variable explained the presence of zeros and non-zeros in CMR or ICM. This model contained the log number of deceased individuals with available postmortem pathological records as an explanatory variable, due to the higher probability of detecting cancer with a higher number of dead individuals inspected. Additionally, the model contained body mass and adult life expectancy as covariates. The second part of the model consisted of a PGLS regression that investigated variance only in non-zero cancer risks. ICM and CMR were logit-transformed in all PGLS models as recommended when analysing proportions^{[45](#)}. These models were weighted by log number of deceased individuals with available postmortem pathological records, as the precision of cancer mortality risk estimates is expected to increase with the number of dead individuals subject to postmortem examination, but it is not expected to explain bias in the estimation of the dependent variable in any particular direction (as in the case of the binomial models). These models also contained body mass and adult life expectancy as explanatory variables. Given the expected additive effect of body mass and longevity, the interaction between body mass and longevity metrics was also tested in all four models (binomial and logistic regressions for CMR and ICM), but these interactions did not increase model fit in any case and are therefore not presented. Both models were controlled for phylogenetic relatedness among species, where the scaling parameter of phylogenetic dependence (that is, $s^2/\text{Pagel's } \lambda$ in PGLMMs and PGLSs respectively) was set to the most appropriate values assessed by likelihood ratio statics in each model separately. PGLS models in which Pagel's λ converged to negative values were refitted with Pagel's λ fixed at 0. Three species (*Lagurus lagurus*, *Cricetus cricetus* and *Dasyuroides byrnei*) had been removed from the latter models, due to their high leverage caused by their very low adult life expectancy compared to the rest of the species and therefore concerns of strong influence of these points over model fit. Nonetheless, all models were performed using the entire dataset, and the results were highly consistent with and without the exclusions ([Supplementary Table 4](#) and Extended Data Fig. [7](#)).

Order differences in cancer incidence were tested using standard linear regressions, built using only taxonomic orders in which at least two species had their cancer incidence estimated. The model contained CMR or ICM (non-transformed) as dependent variables and order as the sole explanatory factor. Pairwise order differences were assessed using the R package emmeans^{[46](#)}. All analysis were

performed in the R Statistical and Programming Environment, version 4.0.4 (ref. [47](#)). Cancer mortality risks were transformed to percentages in figures and in the analysis performed on order differences (Extended Data Table [1](#)), for easier interpretation. Models presented in Extended Data Tables [2](#) and [3](#) and [Supplementary Tables 2–4](#) are based on probabilities.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Data used for the analysis presented in the paper are available at <https://github.com/OrsolyaVincze/VinczeEtal2021Nature/blob/main/SupplementaryData.xls>. Raw data used to estimate cancer risk (Species360 Data Use Approval Number 73836) cannot be publicly shared, as Species360 is the custodian (not the owner) of their members' data. Raw data are accessible through Research Request applications (form available at https://docs.google.com/forms/d/1znoy62VEkDlhAp_0RfEvF7Zsx03g4W5AlppJHqo3_WQ/viewform?edit_requested=true&pli=1). Research Requests are reviewed by both the Species360 Research Committee and their Board of Trustees every four months. The Board of Trustees makes the final decision on data sharing, based on recommendations by the Research Committee. Once Species360 grants access to data, they are intended only for and restricted to use in the project they were approved for and for a single publication. The researcher cannot use them for other projects, publications and/or purposes, nor can the researcher share the data with third parties. For any other inquiries, all of the details for the submission of research requests to Species360 can be found at https://conservation.species360.org/wp-content/uploads/2020/08/Species360-Sharing-Data-v3-3_komprimeret.pdf. Any email communications should be directed to support@species360.org.

Code availability

Data and R code needed to reproduce the analysis are publicly available at <https://github.com/OrsolyaVincze/VinczeEtal2021Nature>.

References

1. 1.

Leroi, A. M., Koufopanou, V. & Burt, A. Cancer selection. *Nat. Rev. Cancer* **3**, 226–231 (2003).

2. 2.

Armitage, P. & Doll, R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* **8**, 1–12 (1954).

3. 3.

Wirén, S. et al. Pooled cohort study on height and risk of cancer and cancer death. *Cancer Causes Control* **25**, 151–159 (2014).

4. 4.

Fleming, J. M., Creevy, K. E. & Promislow, D. E. L. Mortality in North American dogs from 1984 to 2004: an investigation into age-, size-, and breed-related causes of death. *J. Vet. Intern. Med.* **25**, 187–198 (2011).

5. 5.

Nunney, L. Lineage selection and the evolution of multistage carcinogenesis. *Proc. R. Soc. B* **266**, 493–498 (1999).

6. 6.

Peto, R. in *Origins of Human Cancer* Vol 45 (eds. Hiatt, H. et al.) 1403–1428 (Cold Spring Harbor Laboratory, 1977).

7. 7.

Couzin-Frankel, J. The bad luck of cancer. *Science* **347**, 12 (2015).

8. 8.

Chatterjee, N. & Walker, G. C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* **58**, 235–263 (2017).

9. 9.

Nunney, L. The real war on cancer: the evolutionary dynamics of cancer suppression. *Evol. Appl.* **6**, 11–19 (2013).

10. 10.

Ujvari, B., Roche, B. & Thomas, F. *Ecology and Evolution of Cancer* (Academic, 2017).

11. 11.

Nunney, L. Size matters: height, cell number and a person's risk of cancer. *Proc. R. Soc. B* **285**, 20181743 (2018).

12. 12.

Caulin, A. F. & Maley, C. C. Peto's paradox: evolution's prescription for cancer prevention. *Trends Ecol. Evol.* **26**, 175–182 (2011).

13. 13.

Nunney, L., Maley, C. C., Breen, M., Hochberg, M. E. & Schiffman, J. D. Peto's paradox and the promise of comparative oncology. *Philos. Trans. R. Soc. B* **370**, 20140177 (2015).

14. 14.

Peto, R. Epidemiology, multistage models, and short-term mutagenicity tests. *Int. J. Epidemiol.* **45**, 621–637 (2016).

15. 15.

Boddy, A. M. et al. Lifetime cancer prevalence and life history traits in mammals. *Evol. Med. Public Health* **2020**, 187–195 (2020).

16. 16.

Møller, A. P., Erritzøe, J. & Soler, J. J. Life history, immunity, Peto's paradox and tumours in birds. *J. Evol. Biol.* **30**, 960–967 (2017).

17. 17.

Abegglen, L. M. et al. Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. *JAMA* **314**, 1850–1860 (2015).

18. 18.

Tollis, M. et al. Elephant genomes reveal accelerated evolution in mechanisms underlying disease defenses. *Mol. Biol. Evol.* **38**, 3606–3620 (2021).

19. 19.

Conde, D. A. et al. Data gaps and opportunities for comparative and conservation biology. *Proc. Natl Acad. Sci. USA* **116**, 9658–9664 (2019).

20. 20.

Ronget, V. & Gaillard, J. M. Assessing ageing patterns for comparative analyses of mortality curves: going beyond the use of maximum longevity. *Funct. Ecol.* **34**, 65–75 (2020).

21. 21.

Tidière, M. et al. Comparative analyses of longevity and senescence reveal variable survival benefits of living in zoos across mammals. *Sci. Rep.* **6**, 36361 (2016).

22. 22.

Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

23. 23.

Moresco, A. et al. Taxonomic distribution of neoplasia among non-domestic felid species under managed care. *Animals* **10**, 2376 (2020).

24. 24.

Moresco, A. *The Pro-carcinogenic Effects of Progestogens on Carnivore Target Tissues* (2009).

25. 25.

Harrenstien, L. A. et al. Mammary cancer in captive wild felids and risk factors for its development: a retrospective study of the clinical behavior of 31 cases. *J. Zoo Wildl. Med.* **27**, 468–476 (1996).

26. 26.

Munson, L. & Moresco, A. Comparative pathology of mammary gland cancers in domestic and wild animals. *Breast Disease* **28**, 7–21 (2007).

27. 27.

Chao, A. et al. Meat consumption and risk of colorectal cancer. *JAMA* **293**, 172–182 (2005).

28. 28.

Kelly, B. C., Ikonomou, M. G., Blair, J. D., Morin, A. E. & Gobas, F. A. P. C. Food web-specific biomagnification of persistent organic pollutants. *Science* **317**, 236–239 (2007).

29. 29.

Pesavento, P. A., Agnew, D., Keel, M. K. & Woolard, K. D. Cancer in wildlife: patterns of emergence. *Nat. Rev. Cancer* **18**, 646–661 (2018).

30. 30.

Bogolyubova, A. V. Human oncogenic viruses: old facts and new hypotheses. *Mol. Biol.* **53**, 767–775 (2019).

31. 31.

Khatami, A. et al. Bovine leukemia virus (BLV) and risk of breast cancer: a systematic review and meta-analysis of case-control studies. *Infect. Agents Cancer* **15**, 48 (2020).

32. 32.

Kissling, W. D. et al. Establishing macroecological trait datasets: digitalization, extrapolation, and validation of diet preferences in terrestrial mammals worldwide. *Ecol. Evol.* **4**, 2913–2930 (2014).

33. 33.

Olival, K. J. et al. Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).

34. 34.

Ley, R. E. et al. Evolution of mammals and their gut microbes. *Science* **320**, 1647–1651 (2008).

35. 35.

Seluanov, A., Gladyshev, V. N., Vijg, J. & Gorbunova, V. Mechanisms of cancer resistance in long-lived mammals. *Nat. Rev. Cancer* **18**, 433–441 (2018).

36. 36.

Herrera-Álvarez, S., Karlsson, E., Ryder, O. A., Lindblad-Toh, K. & Crawford, A. J. How to make a rodent giant: genomic basis and tradeoffs of gigantism in the capybara, the world's largest rodent. *Mol. Biol. Evol.* **38**, 1715–1730 (2021).

37. 37.

Albuquerque, T. A. F., Drummond do Val, L., Doherty, A. & de Magalhães, J. P. From humans to hydra: patterns of cancer across the tree of life. *Biol. Rev.* **93**, 1715–1734 (2018).

38. 38.

Giraudeau, M., Sepp, T., Ujvari, B., Ewald, P. W. & Thomas, F. Human activities might influence oncogenic processes in wild animal populations. *Nat. Ecol. Evol.* **2**, 1065–1070 (2018).

39. 39.

Thomas, F. et al. Rare and unique adaptations to cancer in domesticated species: an untapped resource? *Evol. Appl.* 12920 (2020).

40. 40.

Therneau, T. M. & Lumley, T. Package ‘survival’. *CRAN* (2014).

41. 41.

Upham, N. S., Esselstyn, J. A. & Jetz, W. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol.* **17**, e3000494 (2019).

42. 42.

Sukumaran, J. & Holder, M. T. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* **26**, 1569–1571 (2010).

43. 43.

Ives, A. R. R^2 s for correlated data: phylogenetic models, LMMs, and GLMMs. *Syst. Biol.* **68**, 234–251 (2018).

44. 44.

Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

45. 45.

Warton, D. I. & Hui, F. K. C. The arcsine is asinine: the analysis of proportions in ecology. *Ecology* **92**, 3–10 (2011).

46. 46.

Lenth R. V. emmeans: estimated marginal means, aka least-squares means, <https://cran.r-project.org/package=emmeans> (2021).

47. 47.

R Core Team. *R: A Language and Environment for Statistical Computing*, <http://www.R-project.org/> (R Foundation for Statistical Computing, 2021).

Acknowledgements

We are grateful to P. Bustamante and T. Székely for constructive criticism on an earlier version of the manuscript, and R. Thompson and A. Teare for invaluable explanations of ZIMS data. We are grateful to more than 1,200 zoo and aquarium members of Species360 that record data in ZIMS, making this study possible. O.V. was financed by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and by the New National Excellence Programme of the Hungarian Ministry of Innovation and Technology. F.T. is supported by the MAVA Foundation, the ANR TRANSCAN (ANR-18-CE35-0009) and a CNRS International Associated Laboratory Grant. C.C.M. and A.M.B. were supported in part by NIH grant U54 CA217376. C.C.M. was also in part supported by NIH grant U2C CA233254, as well as CDMRP Breast Cancer Research Program Award BC132057 and the Arizona Biomedical Research Commission grant ADHS18-198847. D.A.C. was financed by the Species360 CSA sponsors: Copenhagen Zoo, Wildlife Reserves of Singapore and the World Association of Zoos and Aquariums. The findings, opinions and recommendations expressed here are those of the authors and not necessarily those of the universities where the research was performed or the United States National Institutes of Health.

Author information

Affiliations

1. CREEC/CANECEV, MIVEGEC (CREES), University of Montpellier, CNRS, IRD, Montpellier, France

Orsolya Vincze, Frédéric Thomas & Mathieu Giraudeau

2. Littoral, Environnement et Sociétés (LIENSs), UMR 7266 CNRS-La Rochelle Université, La Rochelle, France

Orsolya Vincze & Mathieu Giraudeau

3. Institute of Aquatic Ecology, Centre for Ecological Research, Debrecen, Hungary

Orsolya Vincze

4. Evolutionary Ecology Group, Hungarian Department of Biology and Ecology, Babeş-Bolyai University, Cluj-Napoca, Romania

Orsolya Vincze

5. Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark

Fernando Colchero

6. Interdisciplinary Centre on Population Dynamics, University of Southern Denmark, Odense, Denmark

Fernando Colchero & Dalia A. Conde

7. Species360 Conservation Science Alliance, Bloomington, MN, USA

Fernando Colchero & Dalia A. Conde

8. Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon, Université Lyon 1; CNRS, UMR5558, Villeurbanne, France

Jean-Francois Lemaître

9. Department of Biology, University of Southern Denmark, Odense, Denmark

Dalia A. Conde

10. Eco-Anthropologie (EA), Muséum National d'Histoire Naturelle, CNRS,
Université de Paris, Musée de l'Homme, Paris, France

Samuel Pavard & Margaux Bieuville

11. Instituto de Ecología, UNAM, Mexico City, Mexico

Araxi O. Urrutia

12. Milner Centre for Evolution, Department of Biology and Biochemistry,
University of Bath, Bath, UK

Araxi O. Urrutia

13. Centre for Integrative Ecology, School of Life and Environmental Sciences,
Deakin University, Geelong, Victoria, Australia

Beata Ujvari

14. Department of Anthropology, University of California Santa Barbara, Santa
Barbara, CA, USA

Amy M. Boddy

15. Arizona Cancer Evolution Center, Biodesign Institute and School of Life
Sciences, Arizona State University, Tempe, AZ, USA

Carlo C. Maley

Contributions

O.V. and M.G. contributed to the study conception; O.V., F.C., M.B., M.G., D.A.C., J.-F.L. and S.P. contributed to data collection and interpretation. O.V. performed statistical analysis with significant contribution from F.C. and S.P.; O.V. led the writing of the manuscript; all authors contributed to the interpretation of the results and writing of the manuscript. All authors read and approved the final manuscript.

Corresponding author

Correspondence to [Orsolya Vincze](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks Oliver Ryder and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

[Extended Data Fig. 1 Phylogenetic distribution and order differences of ICM.](#)

a, Phylogenetic distribution of ICM (%). **b**, Violin plots indicating order differences in ICM (%) in orders with minimum two species assessed. Solid black lines indicate order medians. Animal silhouettes used to visually represent mammalian orders were downloaded from PhyloPic (<http://www.phylopic.org>).

[Extended Data Fig. 2 ICM in function of diet animal content.](#)

Violin plots indicate ICM in function of diet animal content. Diet is characterized by variables on three taxonomic levels: I. animal content (including any vertebrate or invertebrate prey); II. invertebrate or vertebrate prey; and III. within vertebrates fish, reptile, bird and mammal preys. Each variable is coded as a two-level factor: rarely/never occurring in diet and representing the primary/secondary food item of the species. Plots indicate the range and distribution of cancer risks in each category, where the width of each curve corresponds with the approximate frequency of data points in each region. Medians are marked with solid black lines. P values indicate pairwise differences as indicated by models presented in Extended Data Table 2 that also control for body mass, life expectancy and phylogeny.

[Extended Data Fig. 3 Predicted cancer mortality risk in function of animal content in diet.](#)

Values represent estimated marginal means of **a**, CMR or **b**, ICM in function of diet animal content, based on models presented in Extended Data Table 2. Diet is characterized by variables on three taxonomic levels: I. animal content (including any vertebrate or invertebrate prey); II. invertebrate or vertebrate prey; and III. within vertebrates fish, reptile, bird and mammal preys. Each variable is coded as a two-level factor: rarely/never occurring in diet and representing the primary/secondary food item of the species. P-values shown were obtained from models presented in Extended Data Table 2.

Extended Data Fig. 4 Association between occurrence of non-zero CMR and a, body mass or b, adult life expectancy across species.

Occurrence of cancer in each species is plotted in function of the number of deceased individuals for which cause of death was known. Predictions were obtained for two scenarios, one for low and one for high **a**, body masses and **b**, adult life expectancies respectively. Random noise was added to cancer occurrence to facilitate the visualization of overlapping points. Predictions and associated 95% confidence intervals were obtained from a binomial GLM without phylogenetic control.

Extended Data Fig. 5 Association between ICM and a,c, body mass or b,d, adult life expectancy.

Occurrence of cancer is plotted in function of the number of deceased individuals for which cause of death was known and predictions were obtained for two scenarios, one for low and high **a**, body masses and **b**, adult life expectancies respectively. Random noise was added to cancer occurrence to facilitate the visualization of overlapping points. Predictions and associated 95% confidence intervals were obtained from a binomial GLM without phylogenetic control. Non-zero cancer mortality risks were plotted against **c**, body mass or **d**, adult life expectancies. Slopes were obtained from the PGLS model presented in Extended Data Table 3b. Points are proportional to the log number of individuals with known cause of death.

Extended Data Fig. 6 Predicted CMR at various hypothetical life expectancies for small, medium, and large-bodies species.

CMR was predicted based on the logistic model presented in Extended Data Table 3a for a series of hypothetical adult life expectancies, ranging from one to 70 years (x-axis). Predictions were obtained for three body masses, corresponding to a small, medium and a large bodied mammal in our dataset. True life expectancies of the three species are marked with red stars. Distribution of life expectancies across the species set of the model is shown by a histogram. Vertical dashed line marks maximum life-

expectancy in our dataset. *Jaculus jaculus* silhouette by Maija Karala. The other animal silhouettes used to visually represent mammalian orders were downloaded from PhyloPic (<http://www.phylopic.org>).

Extended Data Fig. 7 Association between cancer mortality risk and body mass as well as adult life expectancy across 191 mammal species.

These analysis include the three species excluded in some models due to their high leverages. Plots show **a-d**, CMR or **e-h**, ICM. Occurrence of cancer in each species is plotted in function of the number of individuals with post-mortem pathological records. Predictions were obtained for two scenarios: for small or large **a,e**, body masses and low or high **b,f**, adult life expectancies respectively. Random noise was added to cancer occurrence to facilitate the visualization of overlapping points. Predictions and associated 95% confidence intervals were obtained from a binomial GLM without phylogenetic control. Non-zero cancer mortality risks were plotted against **c,g**, body mass or **d,h**, adult life expectancies. Slopes were obtained from the PGLS model presented in [Supplementary Table 4](#). Points are proportional to the log number of individuals with known cause of death.

Extended Data Table 1 Order differences in cancer mortality risk

Extended Data Table 2 Association between cancer risk and diet animal content

Extended Data Table 3 Results of phylogenetic models exploring variation in of cancer mortality risk

Supplementary information

Supplementary Tables

This file contains Supplementary Tables 1–4, which include a list of species excluded from the analyses due to being subject to domestication (Table 1), and sensitivity analyses performed to test the effect of within-species sample size (Table 2), test for collinearity issues (Table 3) and confirm the consistency of the results when high-leverage species are included in the analyses (Table 4).

Reporting Summary

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the

original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Reprints and Permissions

About this article

Cite this article

Vincze, O., Colchero, F., Lemaître, JF. *et al.* Cancer risk across mammals. *Nature* **601**, 263–267 (2022). <https://doi.org/10.1038/s41586-021-04224-5>

- Received: 26 April 2021
- Accepted: 09 November 2021
- Published: 22 December 2021
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04224-5>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04224-5>

- Article
- Open Access
- [Published: 22 December 2021](#)

HELQ is a dual-function DSB repair enzyme modulated by RPA and RAD51

- [Roopesh Anand¹](#),
[Erika Buechelmaier^{2,3}](#),
- [Ondrej Belan¹](#),
- [Matthew Newton¹](#),
- [Aleksandra Vancevska¹](#),
- [Artur Kaczmarczyk^{4,5}](#),
- [Tohru Takaki¹](#),
- [David S. Rueda](#) [ORCID: orcid.org/0000-0003-4657-6323^{4,5}](#),
- [Simon N. Powell](#) [ORCID: orcid.org/0000-0002-8183-4765²](#) &
- [Simon J. Boulton](#) [ORCID: orcid.org/0000-0001-6936-6834¹](#)

[Nature](#) volume 601, pages 268–273 (2022)

- 10k Accesses
- 43 Altmetric
- [Metrics details](#)

Subjects

- [Double-strand DNA breaks](#)
- [Enzyme mechanisms](#)
- [Single-molecule biophysics](#)

Abstract

DNA double-stranded breaks (DSBs) are deleterious lesions, and their incorrect repair can drive cancer development¹. HELQ is a superfamily 2 helicase with 3' to 5' polarity, and its disruption in mice confers germ cells loss, infertility and increased

predisposition to ovarian and pituitary tumours^{2,3,4}. At the cellular level, defects in HELQ result in hypersensitivity to cisplatin and mitomycin C, and persistence of RAD51 foci after DNA damage^{3,5}. Notably, HELQ binds to RPA and the RAD51-paralogue BCDX2 complex, but the relevance of these interactions and how HELQ functions in DSB repair remains unclear^{3,5,6}. Here we show that HELQ helicase activity and a previously unappreciated DNA strand annealing function are differentially regulated by RPA and RAD51. Using biochemistry analyses and single-molecule imaging, we establish that RAD51 forms a complex with and strongly stimulates HELQ as it translocates during DNA unwinding. By contrast, RPA inhibits DNA unwinding by HELQ but strongly stimulates DNA strand annealing.

Mechanistically, we show that HELQ possesses an intrinsic ability to capture RPA-bound DNA strands and then displace RPA to facilitate annealing of complementary sequences. Finally, we show that HELQ deficiency in cells compromises single-strand annealing and microhomology-mediated end-joining pathways and leads to bias towards long-tract gene conversion tracts during homologous recombination. Thus, our results implicate HELQ in multiple arms of DSB repair through co-factor-dependent modulation of intrinsic translocase and DNA strand annealing activities.

[Download PDF](#)

Main

To investigate the functions of HELQ, we purified recombinant human HELQ from insect cells (Extended Data Fig. 1a), which efficiently unwound substrates containing 3' overhangs or a D-loop (Fig. 1a, b and Extended Data Fig. 1b–d). As previously reported, HELQ prefers to unwind single and double-stranded DNA junctions and therefore showed greater unwinding of 3' overhangs and Y-structures than 3' lagging strand forks and D-loops⁷. However, at higher concentrations of HELQ, no unwound product was observed, especially for substrates containing 3' overhangs (described below; Extended Data Fig. 1e). HELQ showed no unwinding with ATPγS, a poorly hydrolysable ATP analogue, and failed to unwind dsDNA and 5' overhang substrates (Extended Data Fig. 1f–h). The helicase-dead mutant of HELQ (HELQ(K365M)) lacked DNA-unwinding activity and showed slightly increased binding to ssDNA and dsDNA compared with the wild-type (WT) protein (Extended Data Fig. 1a, i–m).

Fig. 1: RAD51 forms a co-complex with HELQ and stimulates HELQ unwinding activity.

 **figure1**

a, Representative gel of the DNA unwinding assay with the indicated concentrations of HELQ with 3' overhang substrate. The asterisk indicates the position of fluorescein isothiocyanate (FITC) labelling at 5' end of the oligo. The products were resolved on a 10% native polyacrylamide gel. **b**, Quantification of the experiments such as shown in **a** and Extended Data Fig. [1b–d](#). HELQ concentrations of only 1–90 nM are shown. $n = 4$ independent experiments. Data are mean \pm s.e.m. **c**, Representative gel of DNA unwinding of 3' overhang substrate with HELQ and the indicated concentrations of RAD51 or RecA. **d**, **e**, Quantification of the experiments shown in **c** and Extended Data Fig. [2c,d](#) for RAD51 (**d**) and RecA (**e**). $n = 3$ (3' overhang), $n = 4$ (Y-structure), $n = 3$ (D-loop) and $n = 3$ (lagging strand fork) independent experiments. Data are mean \pm s.e.m. **f**, Representative gel of the DNA unwinding assay of 3' overhang substrate with the indicated concentrations of HELQ in the absence and presence of RPA (20 nM). **g**, Quantification of the experiments shown in **f**. $n = 3$ independent experiments. Data are mean \pm s.e.m. **h**, Schematics of the experimental set-up of the optical tweezer (C-Trap) system to observe DNA unwinding. These experiments were performed at room temperature. **i–k**, Bead centre displacement measured between the traps as a function of time with 25 nM RAD51 (**i**), 50 nM HELQ (**j**), and 25 nM RAD51 and 50 nM HELQ (**k**). The traces represent individual DNA molecules ($n \geq 4$). **l**, Representative kymographs of single Alx–RAD51 binding events on gapped DNA in the presence or absence of 50 nM HELQ or HELQ(K365M). Unidirectional movement of Alx–RAD51 indicates translocation of Alx–RAD51–HELQ complex. Scale bars, 60 s (horizontal), 10 μm (vertical, left), 5 μm (vertical, right).

RAD51 stimulates HELQ unwinding activity

In vivo studies have shown that HELQ-deficient cells exhibit persistent RAD51 foci after DNA damage^{3,8}. Furthermore, HELQ-1 from *Caenorhabditis elegans* interacts with RAD-51 (ref. ⁸). HELQ and human RAD51, purified from *Escherichia coli* (Extended Data Fig. ^{2a}), also interact directly (Extended Data Fig. ^{2b}). In unwinding assays, RAD51 strongly stimulated HELQ helicase activity with all of the tested substrates, whereas bacterial RecA—an orthologue of RAD51—did not stimulate HELQ even at higher concentrations (Fig. ^{1c–e} and Extended Data Fig. ^{2c–g}). We next purified and tested the BRC4 peptide (single BRCA2 BRC repeat), which prevents RAD51 from binding to DNA (Extended Data Fig. ^{2h–j}). The BRC4 peptide did not inhibit stimulation of HELQ by RAD51 (Extended Data Fig. ^{2k–n}), indicating that RAD51 DNA binding is not required for HELQ stimulation and excludes DNA sequestration as a possible mechanism for stimulation by RAD51. We observed that, at a higher concentration of RAD51 (that is, 120 nM), HELQ unwinding activity is inhibited (Fig. ^{1c, d}). To investigate this, we tested the BRC4 peptide with an excess RAD51 and found that inhibiting RAD51 DNA binding also rescued DNA unwinding by HELQ (Extended Data Fig. ^{2o, p}). Furthermore, by measuring the kinetics of DNA unwinding by HELQ, we found that addition of RAD51 resulted in a concentration-dependent increase in the HELQ DNA unwinding rate, whereas addition of RecA had no effect (Extended Data Fig. ^{2q–s}). In cells, ssDNA generated during DNA processing is bound by RPA. To mimic these conditions, we purified fluorescently tagged human RPA-mRFP1 from *E. coli* (Extended Data Fig. ^{3a}). Addition of RPA inhibited DNA unwinding by HELQ, especially for 3' overhang substrates (Fig. ^{1f, g} and Extended Data Fig. ^{3b, c}). At lower concentrations that were insufficient to cover the entire ssDNA region, RPA still inhibited HELQ unwinding of 3' overhang substrates (Extended Data Fig. ^{3d, e}). Despite the inhibitory effect of RPA, RAD51 still stimulated HELQ helicase activity in the presence of RPA (Extended Data Fig. ^{3f–i}).

Visualization of HELQ DNA unwinding

To better understand HELQ stimulation by RAD51, we used an optical tweezer set-up combined with microfluidics and confocal microscopy (C-TRAP) for single-molecule imaging (SMI) analysis. As shown in Fig ^{1h}, a single dsDNA molecule (λ -DNA) containing a ssDNA gap⁹ was tethered between two optically trapped beads and held at constant force (50 pN) to prevent the reannealing of unwound DNA. After addition of HELQ, DNA unwinding was observed as an increase in the distance between the beads due to the expansion of the ssDNA region. Neither RAD51 alone nor HELQ(K365M) showed evidence of unwinding (Fig. ¹ⁱ and Extended Data Fig. ^{4a}). Combining WT HELQ and RAD51 resulted in a considerable increase in overall DNA

unwinding, whereas no such stimulation was observed with HELQ(K365M) (Fig. 1j, k and Extended Data Fig. 4b). Within unwinding traces for individual DNA molecules, rapid unwinding bursts interspersed by pauses can be distinguished (Extended Data Fig. 4c–e) and corresponded to a mean rate of $3.3 \pm 0.4 \text{ nm s}^{-1}$ (mean \pm s.e.m Extended Data Fig. 4f). In the presence of RAD51, an increased number of molecules showed greater unwinding rates (Extended Data Fig. 4g). To directly visualize RAD51 during DNA unwinding with HELQ, mutant RAD51(C319S) was purified and efficiently labelled with Alexa Fluor 488 C5 maleimide dye (Alx–RAD51) (Extended Data Fig. 4h). Whereas Alx–RAD51 alone displayed mostly static binding traces with occasional diffusing species, addition of HELQ showed unidirectionally translocating traces indicating active movement of an Alx–RAD51–HELQ complex along the ssDNA backbone (Fig. 1l and Extended Data Fig. 4i). After analysis, we found that HELQ with RAD51 translocates fastest at the rate of $14 \pm 5 \text{ nm s}^{-1}$ in gapped substrate (Extended Data Fig. 4j, k). By contrast, HELQ(K365M) retained the ability to bind to RAD51 but showed no translocation with only static or diffusing traces. Together, these results indicate that RAD51 and HELQ form a complex that unwinds DNA at a rate of approximately threefold faster compared with HELQ alone.

DNA strand annealing by HELQ

As shown above, a strong reduction in unwound product was observed at higher concentrations of HELQ (Fig. 1a, b and Extended Data Fig. 1b–e); we reasoned that this could result from the reannealing of the unwound product. Indeed, reactions containing an unlabelled ‘cold’ oligonucleotide (oligo) yielded an increase in unwound product with excess HELQ (Extended Data Fig. 5a (compare lanes 3 and 4 with 7 and 8)). Kinetic analysis also showed that HELQ initially unwinds the substrate but then reanneals it back together at later time points (Extended Data Fig. 5b). Prompted by this, we directly tested HELQ for DNA strand annealing activity without and with an excess of RPA needed for 100% ssDNA coverage (theoretically, 16 nM RPA covers 10 nM ssDNA). We found that HELQ efficiently anneals complementary DNA strands either without or with RPA (Fig. 2a, b). At lower concentrations of HELQ, RPA stimulated HELQ DNA annealing activity by around twofold. However, at higher concentrations, HELQ showed greater DNA annealing activity in the absence of RPA. This raised the possibility that RPA aids HELQ loading on ssDNA when HELQ is present in limiting amounts. Titration experiments confirmed that stoichiometric levels of RPA are sufficient to stimulate HELQ annealing activity (Fig. 2c, d). Notably, HELQ could still anneal complementary DNA strands in the presence of excess RPA (Extended Data Fig. 5c, d).

Fig. 2: HELQ strips RPA from and anneals ssDNA.

 **figure2**

a, Representative gel of DNA annealing assay with the indicated concentrations of HELQ in the absence and presence of RPA (40 nM). The black and blue colours of substrate represent complementary DNA strands. The asterisk indicates the position of FITC labelling at the 5' end of the oligo. The products were resolved with 10% native

polyacrylamide gel. **b**, Quantification of experiments such as shown in **a**. $n = 6$ independent experiments. Data are mean \pm s.e.m. **c**, Representative gel of the DNA annealing assay with HELQ (10 nM) and the indicated concentrations of RPA. **d**, Quantification of the experiments shown in **c**. $n = 3$ independent experiments. Data are mean \pm s.e.m. **e**, Representative gel of the DNA annealing assay with the indicated concentrations of HELQ and RPA (40 nM), in the absence and presence of ATP. **f**, Quantification of the experiments shown in **e**. $n = 3$ independent experiments. Data are mean \pm s.e.m. **g**, Representative gel of DNA annealing assay with HELQ(K365M) (10 nM) and the indicated concentrations of RPA. **h**, Quantification of the experiments shown in **g**. $n = 4$ independent experiments. Data are mean \pm s.e.m. **i**, Schematics of the experimental set-up of the optical tweezer (C-Trap) system to observe RPA–eGFP stripping from ssDNA. These experiments were performed at room temperature. **j**, Representative kymographs of stripping of RPA–eGFP prebound to gapped DNA in the presence or absence of 50 nM HELQ or HELQ(K365M). Unidirectional stripping of RPA–eGFP with HELQ WT indicates 3' to 5' translocation of HELQ. **k**, Removal of RPA–eGFP measured from gapped ssDNA as function of time in indicated conditions. The traces represent individual DNA molecules ($n \geq 3$). Scale bars, 2 min (horizontal), 2 μ m (vertical).

We next tested the requirement of ATP binding and hydrolysis for DNA annealing by HELQ. Surprisingly, in the presence of RPA, HELQ showed no DNA annealing without ATP, whereas ATP became dispensable when RPA was excluded from the reaction (Fig. [2e,f](#) and Extended Data Fig. [5e,f](#)). Even in the absence of RPA, ATP stimulated the DNA annealing activity of HELQ (Extended Data Fig. [5e,f](#)). HELQ also failed completely to anneal DNA with ATP γ S in the presence of RPA (Extended Data Fig. [5g](#)). Collectively, these data suggest that HELQ possesses intrinsic DNA annealing activity that requires ATP binding and hydrolysis when ssDNA is coated with RPA. We next tested the helicase-inactive HELQ(K365M) mutant for DNA annealing activity and found that HELQ(K365M) is defective for DNA annealing in the presence of excess RPA but could anneal ssDNAs when RPA is excluded from reactions (Extended Data Fig. [5h,i](#)). A titration experiment showed that, in contrast to the WT, HELQ(K365M) becomes progressively impaired by increasing concentrations of RPA (Fig. [2g,h](#) and Extended Data Fig. [5j,k](#)). RPA also failed to stimulate HELQ(K365M) (Fig. [2g,h](#)). We also tested *E. coli* SSB protein and found that it only weakly stimulates HELQ annealing activity (Extended Data Fig. [5l,m](#)). The N-terminal fragment of HELQ was previously shown to displace RPA from ssDNA^{[6](#)}. However, full-length HELQ was not analysed for such activity. Thus, we directly tested RPA displacement from ssDNA during DNA strand annealing by omitting the deproteination step. We observed that HELQ can strip an excess of RPA from ssDNA, which occurred coincidentally with the appearance of the annealed products (Extended Data Fig. [6a,b](#)). To directly visualize RPA stripping from ssDNA by HELQ, we measured RPA–eGFP displacement by SMI analysis and found that WT HELQ ($k =$

$0.136 \pm 0.008 \text{ min}^{-1}$) could efficiently strip RPA from ssDNA but HELQ(K365M) ($k = 0.017 \pm 0.004 \text{ min}^{-1}$) could not (Fig. 2*i–k*). Using a single-molecule Förster resonance energy transfer (FRET)-based assay (Extended Data Fig. 6*c–j*), we observed concentration-dependant RPA stripping by WT HELQ, followed by rebinding of RPA (Extended Data Fig. 6*k–n*). The RPA rebinding is independent of HELQ concentration, indicating a constant transition rate from the free (t_{on}) to bound (t_{off}) state at various HELQ concentrations (Extended Data Fig. 6*o*). HELQ(K365M) did not show any RPA stripping (Extended Data Fig. 6*p*), indicating that active RPA stripping has a critical role in HELQ-mediated DNA annealing. Finally, we found that the addition of RAD51 had no effect on HELQ-dependent DNA annealing activity (Extended Data Fig. 6*q,r*). To study whether, like RPA, HELQ can also strip RAD51 from ssDNA, we directly tested RAD51 removal from ssDNA and dsDNA. HELQ did not remove RAD51 from either ssDNA or dsDNA in bulk assays (Extended Data Fig. 7*a–d*) and only weakly displaced RAD51 from ssDNA in our SMI setup (Extended Data Fig. 7*e–g*). This suggests that HELQ is unlikely to disrupt fully formed RAD51 nucleofilaments but might remove RAD51 that is bound to short-resected DNA, as is present during microhomology-mediated end joining (MMEJ).

HELQ captures RPA–ssDNA

RAD52 has been shown to possess robust DNA strand annealing activity and has a central role in single-strand annealing (SSA) repair of DSBs^{10,11,12,13,14}. To investigate the mechanistic basis of the strand annealing ability of HELQ, we modelled our experiments on RAD52 annealing activity. Using optical tweezers, it was previously shown that RAD52 can *trans*-capture labelled oligos at multiple sites along λ -DNA, independent of DNA sequence¹⁵. Using a similar set-up, we tested the ability of HELQ to capture a Cy3-labelled 80-mer oligo ssDNA in the presence of RPA–eGFP (around 100% coverage; Fig 3*a* and Extended Data Fig. 8*a*). HELQ facilitated the capture of λ 4 oligo^{16,17} at multiple sites (Fig. 3*b,c*). Notably, in contrast to its annealing or unwinding activity, HELQ(K365M) exhibited efficient DNA capture activity (Fig. 3*b,c*). We also analysed the dwell times of captured oligos and found that HELQ(K365M) showed moderately increased dwell times compared with the WT (Fig. 3*d,e*), most likely due to slightly higher DNA binding of HELQ(K365M). We also tested a 79-nucleotide T-homopolymer (dT79) and obtained similar capture frequencies as with the λ 4 oligo (Fig. 3*c*).

Fig. 3: HELQ exhibits sequence-independent RPA-coated DNA capture activity.

 **figure3**

a, Schematics of the experimental optical tweezer set-up for observing the capture of Cy3-labelled DNA oligos. **b**, Kymographs showing the capture of targeted Cy3-labelled 80-mer oligo (λ 4) in *trans* with HELQ and HELQ(K365M) in the presence of RPA–eGFP at multiple positions of RPA–eGFP-coated λ -ssDNA (top). KM, HELQ(K365M). Kymographs showing the capture of non-targeted Cy3-labelled dT 79 homopolymer in *trans* with HELQ in the presence of RPA–eGFP at multiple positions of RPA–eGFP-coated λ -ssDNA (bottom). Scale bars, 60 s (horizontal), 10 μ m (vertical). **c**, Quantification of the experiments shown in **b**. Each datapoint represents a single DNA molecule. Data are mean \pm s.d. Statistical analysis was performed using two-sided Mann–Whitney *U*-tests. No adjustments were made for multiple comparisons. **d**, The dwell times of captured λ 4 by HELQ in the presence of RPA–eGFP. $n = 61$. The black line represents the exponential fit. $Tau = 134$ s. **e**, The dwell times of captured λ 4 by HELQ(K365M) in the presence of RPA–eGFP. $n = 87$. The black line represents the exponential fit. $Tau = 179$ s. **f**, The experimental conditions for the optical tweezer experiment in **g**. **g**, Kymograph showing the Cy3– λ 4 oligo captured on RPA–eGFP-coated λ -ssDNA in the presence of HELQ and RPA–eGFP after stretching of tethered λ -ssDNA by a gradual increase of force. Scale bars, 60 s (horizontal), 10 μ m (vertical). **h**, Schematic of the bulk capture assay. **i**, Native gel showing the capture assay with the indicated concentrations of HELQ and RPA (82 nM). B indicates biotin at the 3' end of B-dT43 oligo. The experiment was performed twice with similar results.

HELQ showed capture of both homologous (λ 4) and T-homopolymer oligos at various sites on λ -DNA. To determine whether microhomologies present in ~50 kb λ -DNA could explain the capture pattern of both oligos, we performed forced-stretching

experiments as shown in Fig. 3f. As DNA starts to melt at forces of >60 pN, we reasoned that, if HELQ oligo capture involves base-pairing interactions, short microhomologies should dissociate faster than the ones with longer homology. However, even at very high force (90–100 pN), all oligos remained engaged with ssDNA, irrespective of position (Fig. 3g). We next performed pulling experiments¹⁸ and observed characteristic force spikes when beads were pulled apart at a low force (10–15 pN) (Extended Data Fig. 8b,c). These spikes correspond to the disruption of HELQ complexes capturing RPA-coated ssλ-DNA in *cis*. To examine this further, we developed a bulk capture assay, in which we attempted to pull out labelled non-complementary DNA (Cy3-dT79) with a biotinylated dT43 oligo with HELQ. We found that both WT HELQ and the HELQ(K365M) mutant could capture non-complementary oligos bound to RPA (Fig. 3h,i and Extended Data Fig. 8d). Thus, HELQ and HELQ(K365M) can both capture DNA strands independent of sequence, probably through DNA tethering, but only the WT can anneal RPA-coated complementary DNA strands through RPA stripping. Interestingly, yeast Rad52, when bound to RPA-coated ssDNA clusters, can capture additional free RPA in pre-existing Rad52–RPA–ssDNA clusters¹⁹. This activity was postulated to be important for second-end capture during homologous recombination (HR).

HELQ functions in SSA and MMEJ

To extend our findings with HELQ to DSB repair *in vivo*, we first confirmed that deletion of *HELQ* or protein depletion in cells inhibits HR (Extended Data Fig. 9a–e). As DNA annealing is required for SSA repair, we investigated a potential role for HELQ in this process. Strikingly, HELQ depletion or *HELQ* deletion impaired SSA repair of an integrated SSA reporter (SA-GFP; Fig. 4a,b). Although depletion of the HR factor, BRCA2, increases SSA repair, this was strongly reduced by HELQ depletion (Fig. 4c and Extended Data Fig. 9f). Consistent with an epistatic role in SSA, co-depletion of RAD52 and HELQ did not further decrease SSA repair compared with individual depletions (Extended Data Fig. 9g). We also assessed whether HELQ functions in alternative end joining repair, which also involves an annealing step. Using cells containing both EJ-RFP and DR-GFP reporter systems for simultaneous detection of alternative end joining and HR, respectively, we observed a significant reduction in both DSB repair pathways after HELQ depletion (Fig. 4d,e). Alternative end joining encompasses broad non-nonhomologous end joining (NHEJ) repair events including MMEJ. To specifically study the role of HELQ in MMEJ, we used a Cas9-mediated DSB repair assay in which the fate of DSB repair by MMEJ, NHEJ or single-stranded templated repair (SSTR) can be determined²⁰ (Fig. 4f). The loss of *HELQ* resulted in a slight increase in NHEJ, whereas MMEJ was significantly impaired and SSTR was completely abolished (Fig. 4g). Interestingly, DNA annealing is important for SSTR^{21,22}. A role for HELQ in DNA end resection could potentially

explain defects in HR, SSA and MMEJ. However, we did not observe any reduction in RPA or RAD51 (refs. [3,23,24](#)) focus formation after irradiation or camptothecin treatment in *HELQ*^{-/-} or depleted cells, excluding a role for HELQ in resection (Extended Data Fig. [9h-l](#)). Impaired DNA strand annealing during second-end capture in DSB repair or failure to capture the repaired strand in synthesis-dependent strand annealing can result in a shift towards long-tract gene conversion (LTGC)[25,26,27,28,29](#). Using the same reporter system, we found that, similar to RAD52, HELQ deficiency results in a strong and moderate reduction in short-tract gene conversion (STGC) and LTGC, respectively, with the LTGC/total gene conversion (GC) ratio showing a significant bias towards LTGC (Fig. [4h-k](#)). Finally, co-depletion of HELQ with RAD52 showed a further bias towards LTGC, implying that they have roles in parallel pathways during GC (Fig. [4k](#)).

Fig. 4: HELQ functions in DSB repair, SSA and MMEJ.



a, Schematic of the SA-GFP reporter assay. **b**, I-SceI-induced SSA frequency in U2OS-SA *HELQ*^{-/-} cells and WT cells treated with the indicated short interfering RNA (siRNA). $n = 17$ (non-targeting siRNA (siNT)/WT), $n = 7$ (siHELQ 1), $n = 12$ (siHELQ_2), $n = 10$ (siHELQ 3) and $n = 5$ (*HELQ*^{-/-}) independent experiments. Data are mean \pm s.e.m. **c**, SSA frequency in U2OS-SA cells treated with the indicated siRNA. $n = 6$ independent experiments. Data are mean \pm s.e.m. **d**, Schematic of the EJ-RFP reporter assay. **e**, I-SceI-induced mutagenic end-joining (mEJ-RFP) and HR (DR-GFP) frequencies in U2OS-EJDR cells treated with the indicated siRNA. $n = 7$ (siNT),

$n = 4$ (siHELQ 1), $n = 2$ (siHELQ 2) and $n = 3$ (siHELQ 3) independent experiments. Data are mean \pm s.e.m. **f**, Schematic of Cas9-mediated DSB repair assay. **g**, Cas9-mediated DSB repair assay showing the frequencies of NHEJ, MMEJ and SSTR repair in *HELQ*^{-/-} and WT U2OS-DR or U2OS-SA cells. $n = 5$ (U2OS-DR WT and DR *HELQ*^{-/-}) and $n = 4$ (U2OS-SA and SA *HELQ*^{-/-}) independent experiments. Data are mean \pm s.e.m. **h**, Schematic of the RFP-SCR reporter assay. Tr GFP, truncated GFP. **i**, I-SceI-induced STGC frequency in U2OS-RFP-SCR cells treated with the indicated siRNA. $n = 13$ (siNT), $n = 11$ (siHELQ 1), $n = 12$ (siRAD52) and $n = 6$ (siHELQ 1/siRAD52) independent experiments. Data are mean \pm s.e.m. **j**, LTGC frequency as shown in **i**. $n = 13$ (siNT), $n = 12$ (siHELQ 1), $n = 11$ (siRAD52) and $n = 6$ (siHELQ 1/siRAD52) independent experiments. Data are mean \pm s.e.m. **k**, The ratio of LTGC/total GC from the experiments in **i** and **j**. $n = 14$ (siNT), $n = 12$ (siHELQ 1), $n = 12$ (siRAD52) and $n = 6$ (siHELQ 1/siRAD52) independent experiments. Data are mean \pm s.e.m. For **b**, **c**, **e**, **g**, **i–k**, statistical analysis was performed using two-tailed paired *t*-tests as compared with siNT, unless otherwise indicated.

In summary, our study implicates HELQ in several distinct DSB repair pathways, including HR, SSA and MMEJ, casting light on its role in genome stability and tumour avoidance. As these repair pathways each require DNA annealing steps, we propose that HELQ functions in these pathways through its intrinsic ability to capture RPA-bound ssDNA and then displace RPA to facilitate annealing of complementary DNA strands. HELQ alone or in complex with RAD51 might unwind D-loops before annealing complementary strands (Extended Data Fig. 10). The bias towards LTGC events after HELQ depletion is consistent with a role for DNA annealing by HELQ during second-strand capture and/or synthesis-dependent strand annealing during HR, which may explain why HR is reduced in HELQ-deficient cells. Finally, our finding that HELQ is epistatic with RAD52 for SSA but additive for GC during HR is surprising, as this implicates two DNA strand annealing enzymes in several distinct DSB repair processes in cells.

Methods

Preparation of expression vectors

Custom HELQ ORF was purchased from GeneArt and used as a template during PCR to prepare plasmid (pFastbac1) compatible for bacmid preparation for expression in insect cells. To prepare the MBP-HELQ-Flag construct, HELQ was amplified by PCR using the primers HELQ_F and HELQ_FLAG_R. The amplified insert was digested with BamHI and XmaI and inserted into the pFastbac1 vector containing MBP (maltose binding protein tag; previously inserted using the EcoRV and HindIII restriction sites). The resulting construct was pFB-MBP-HELQ-FLAG. To prepare helicase-dead HELQ(K365M), pFB-MBP-HELQ-Flag was mutagenized with the

primers HELQ_K365M_F and HELQ_K365M_R using the Q5 site-directed mutagenesis kit according to the manufacturer's instructions. The constructs for pET11c-RAD51 and, RPA-eGFP (pMM801) and RPA-mRFP1 (pMM802) were gifts from L. Krejci and M. Modesti, respectively.

Recombinant protein purification

To express proteins in insect cells, bacmid, primary and secondary baculoviruses were prepared according to the manufacturer's instructions (Bac-to-bac system, Life technologies). To express recombinant MBP-HELQ-FLAG, *Spodoptera frugiperda* (Sf9) insect cells were seeded at 500,000 cells per ml and, after around 24 h, cells were infected with MBP-HELQ-Flag baculovirus. The infected cells were incubated at 27 °C for 56 h with continuous agitation. Cells were collected by centrifugation at 500g for 10 min and washed once with 1× PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄). The collected pellets were snap-frozen in liquid nitrogen and stored at -80 °C until further use. All subsequent steps were performed either on ice or at 4 °C. The cells pellets were resuspended in 3 volumes of lysis buffer containing 50 mM Tris-HCl pH 7.5, 1 mM ethylenediaminetetraacetic acid (EDTA), protease inhibitor cocktail tablets (Roche), 30 µg ml⁻¹ leupeptin (Merck), 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM dithiothreitol (DTT), 0.1% NP-40 substitute (NP-40) and incubated for 15 min with continuous agitation. Next, 50% glycerol and 5 M NaCl were added sequentially to the final concentrations of 16.7% and 310 mM, respectively, and the suspension was further incubated for 30 min with continuous agitation. The suspension was centrifuged at ~48,000g for 30 min to obtain the soluble extract. The amylose resin (NEB) was pre-equilibrated with amylose wash buffer I (50 mM Tris-HCl pH 7.5, 1 mM 2-mercaptoethanol (β-ME), 1 M NaCl, 1 mM PMSF, 10% glycerol and 0.1% NP-40) and added to 50 ml tubes containing soluble extract. These tubes were subsequently incubated for 1 h with continuous rotation. After incubation, the resin was washed batchwise four times by centrifugation at 2,000g for two min and twice on column with amylose wash buffer I. Resin was washed twice more on column with amylose wash buffer II (same as wash buffer I but with 0.5 mM β-ME and 0.8 M NaCl). Protein was eluted from the resin using amylose elution buffer (same as amylose wash buffer II supplemented with 10 mM maltose) and total protein was estimated using the Bradford assay. To remove the MBP tag, 1/8 (w/w) of PreScission protease to the total protein was added to amylose eluate and incubated for 2 h without rotation but with gentle agitation at regular intervals. The Flag resin (anti-Flag M2 resin, Sigma-Aldrich), which was pre-equilibrated with Flag wash buffer (50 mM Tris-HCl pH 7.5, 0.8 M NaCl, 1 mM PMSF, 10% glycerol), was added to amylose eluate containing PreScission protease and incubated for 2 h with continuous rotation. Flag resin was collected directly on column and washed six times with Flag wash buffer. The protein was eluted from Flag resin with Flag elution buffer (50 mM Tris-HCl pH 7.5, 0.5 mM β-ME, 150 mM NaCl, 1 mM PMSF, 10% glycerol,

150 ng μl^{-1} 3 \times Flag peptide (Sigma-Aldrich)), aliquoted, frozen in liquid nitrogen and stored at -80°C . The same purification procedure was used to purify HELQ(K365M).

Recombinant human RAD51 was purified as described previously with a few modifications³⁰. The pET11c-RAD51 expression vector was transformed into *E. coli* BLR(DE3)pLysS cells and subsequent culture containing ampicillin (100 mg l^{-1}) and chloramphenicol (33 mg l^{-1}) was grown to an optical density (OD) at 600 nm of 0.7. RAD51 expression was induced with 1 mM isopropyl β -d-1-thiogalactopyranoside (IPTG) at 37°C for 3–4 h. All of the subsequent steps were performed either on ice or at 4°C . Cells were collected by centrifugation at 5,000g. Cell pellets were resuspended in cell breakage buffer (50 mM Tris-HCl pH 7.5, 10% sucrose, 0.5 mM EDTA, 1 M KCl, 1 tablet per 50 ml of protease inhibitor cocktail tablets (Roche), 1 mM PMSF, 1 mM DTT and 0.01% NP-40), sonicated and centrifuged at 100,000g for 1 h. To precipitate RAD51, 0.242 g ml^{-1} ammonium sulphate was mixed with clarified supernatant and centrifuged for 20 min at 10,000g. The pellet was resuspended with buffer K (20 mM K_2HPO_4 pH 7.5, 10% glycerol, 0.5 mM EDTA, 1 mM DTT, 0.01% NP-40) and loaded onto the Q Sepharose Fast flow column (Cytiva), pre-equilibrated with K buffer-low (K buffer supplemented with 175 mM KCl). The column was washed extensively with K buffer-low and protein was subsequently eluted with a KCl gradient using K buffer-high (K buffer supplemented with 0.6 M KCl). The eluted fractions containing RAD51 were pooled and diluted with 6 volumes of dilution buffer (25 mM Tris HCl pH 7.5, 0.5 mM EDTA, 1 mM DTT, 0.01% NP-40). The diluted sample was loaded onto the HiTrap Heparin HP affinity column (Cytiva), which was pre-equilibrated with buffer H without glycerol (25 mM Tris HCl pH 7.5, 0.5 mM EDTA, 1 mM DTT, 0.01% NP-40, 150 mM KCl) and washed with buffer H containing 10% glycerol. Protein was eluted in buffer H with a KCl gradient (0.1 M to 1 M KCl) and the fractions containing RAD51 were pooled and dialysed in buffer H without glycerol. The dialysed sample was loaded onto the Mono Q 5/50 GL column (Cytiva), equilibrated with buffer Q (25 mM Tris HCl pH 7.5, 0.5 mM EDTA, 1 mM DTT, 0.01% NP-40, 100 mM KCl, 10% glycerol) and the column was further washed with buffer Q containing 50 mM KCl but lacking glycerol. RAD51 was eluted from the Mono Q column with a KCl gradient (0.05 M to 1 M) in buffer Q lacking glycerol. The eluted fractions containing RAD51 were pooled and further concentrated with the Vivaspin Centrifugal Concentrator (30 kDa molecular weight cut-off (MWCO)). Glycerol was added to the concentrated sample to a final concentration of 10%. Finally, the samples were aliquoted, frozen in liquid nitrogen and stored at -80°C . The RAD51(C319S) mutant was purified using same procedure. RPA-mRFP1 and RPA-eGFP were purified as described previously³¹. Recombinant RecA (M0249) and ET SSB (M2401) were commercially purchased from NEB, England.

To purify GST and GST-BRC4, the BRC4 construct was cloned into the pGEX6P-1 vector containing a GST tag using BamHI and EcoRI restriction sites. The pGEX6P-1

and pGEX6P-1–BRC4 constructs were transformed into BL21 DE3 cells and plated onto an agar plate containing ampicillin ($100\text{ }\mu\text{g ml}^{-1}$). A single colony was isolated and inoculated into 6 ml preculture overnight. The next day, 4.5 ml preculture medium was added to 450 ml LB medium containing ampicillin ($100\text{ }\mu\text{g ml}^{-1}$), and the OD at 600nm was monitored at regular intervals. Proteins were induced with 1 mM IPTG at 0.6 OD and cultures were incubated for 4 h. The cell pellets were collected, washed with cold PBS and stored at $-80\text{ }^{\circ}\text{C}$. For purification, the cell pellets were lysed by sonication in PBS and the samples were centrifuged at 75,600g for 30 min at $4\text{ }^{\circ}\text{C}$. Next, the supernatant was collected and incubated with 1.2 ml glutathione resin for 1 h at $4\text{ }^{\circ}\text{C}$. The resin was washed three times with PBS and the proteins were eluted with elution buffer (20 mM Tris-HCl pH 7.5, 20 mM glutathione). The eluted proteins were aliquoted, snap-frozen in liquid nitrogen and stored at $-80\text{ }^{\circ}\text{C}$.

Preparation of labelled proteins

The RAD51(C319S) variant was expressed and purified as described earlier for WT RAD51 (refs. [30,32](#)). After purification, the protein was fluorescently labelled with Alexa Fluor 488 C₅ maleimide (Thermo Fisher Scientific, A10254) according to previously described protocol^{[31](#)}. Labelled protein was purified away from the free dye using the Zeba column gel filtration system (0.5 ml resin, 50 kDa MWCO). The protein concentration was estimated by Coomassie staining and dye concentration was measured spectrophotometrically. The presence of minimum free dye concentration was assessed using SDS-PAGE on labelled proteins. The protein to dye concentration ratio was consistently 0.9–1.0. D-loop formation of labelled RAD51 was tested and gave yields comparable to unlabelled WT RAD51 protein, consistent with previous reports^{[32](#)}. RPA–eGFP and RPA–mRFP1 were expressed and purified as described previously^{[31](#)}. DNA binding of labelled RPA was tested. All RPA-fusion proteins displayed similar ssDNA affinities within nanomolar K_d range.

Preparation of DNA substrates and oligonucleotides used for in vitro analysis

All DNA oligonucleotides used in the in vitro analysis were commercially synthesized and purchased from Merck Life Sciences. To prepare various substrates used in this study, when needed, combination(s) of DNA oligonucleotides were annealed together by mixing and heating them at $95\text{ }^{\circ}\text{C}$ for 3 min, followed by gradual cooling of the samples overnight. The names and sequences of oligos used were as follow: oligo 1 (5' FITC-AGCTACCATGCCTGC ACGAATTAAAGCAATTGTA ATCATGGTCATAGCT), oligo 2 (5'-AGCTATGACCATGATT ACGAATTGCTTAATTCTGTGCAGGCATGGTAGCT, oligo 4 (AATTCGTGCAGGCATGGTAGCT), oligo 5 (AGCTATGACCATG

ATTACGAATTGCTT), oligo 6
 (AGCTATGACCATGATTACGAATTGCTTGGAAATCCTGACGAACGTAG), oligo
 23 (5'-FITC-GACGCTGCCGAATTCTACCAAG TGCCTTGCTAGGACATCTTGCC
 CACCTGCAGGTTCACCC), oligo 22 (GGGTGAACCTGCAGGTGGG CG
 GCTGCTCATCGTAGGTTAGTTGGTAGAATTGGCAGCGTC), oligo 21
 (TAAGAGCAAGATGTTCTATAAAA GATGTCCTAGCAAGGCAC), oligo 20
 (TATAGAACATCTTGCTCTTA); oligo F (5'-FAM-AGCTACCATGCCTGCACG
 AATTAAGCAATTGCTAA TCATGGTCATAGCT) and oligo R
 (AATTCGTGCAGGCATGGTAGCT-ROX-3'). FITC, FAM and ROX indicate the
 position of fluorescein isothiocyanate, 6-carboxyfluorescein and rhodamine,
 respectively in oligos above. The combinations of oligos were annealed together to
 prepare 3' overhang (oligo 1 + oligo 4), 5' overhang (oligo 1 + oligo 5), dsDNA (oligo
 1 + oligo 2), Y structure (oligo 1 + oligo 6), lagging strand fork (oligo 1 + oligo 4 +
 oligo 6), D-loop (oligo 23 + oligo 22, oligo 21 + oligo 20) and 3' overhang used for
 quenching kinetic assay (oligo F + oligo R). Oligo 1 was used as the ssDNA substrate.
 The additional oligos used were as follows; λ4 (5'-Cy3-CCTGAACGACCAAG
 GCGTCTCGTTCATC TATCGGATCGCCACACTCA
 CAACAATGAGTGGCAGATAT AGCCTGGTGGTTC), dT79 (5'-Cy3-
 TTT
 TTT
 (TT-3'-Bio), where Bio
 indicates the position of biotin in the oligo sequence.

DNA unwinding assay

The unwinding assays were performed in 15 µl helicase buffer containing 25 mM Tris-HCl pH 8.0, 2 mM ATP, 2 mM MgCl₂, 1 mM DTT, 50 mM NaCl, 0.1 mg ml⁻¹ bovine serum albumin (BSA, New England Biolabs), 1 mM PEP (phosphoenolpyruvate, Sigma-Aldrich), 10 U ml⁻¹ pyruvate kinase (Sigma-Aldrich) and 5'-end-FITC-labelled 25 nM DNA substrate (in molecules). All of the steps except the assembling reactions and protein addition were performed in the dark. The reactions were assembled on ice and recombinant proteins were added, mixed and incubated at 37 °C for 30 min. The reactions were stopped with 5 µl of 2% stop solution (0.2% SDS, 30% glycerol, 150 mM EDTA, bromophenol blue) and 1 µl proteinase K (Roche, 18.4 mg ml⁻¹) and incubated for 10 min at 37 °C. To prevent reannealing, 2% stop solution was supplemented with tenfold excess of unlabelled oligos with the same sequence as the FITC-labelled oligo. The products were resolved by 10% native polyacrylamide gel (19:1 acrylamide-bisacrylamide, Bio-Rad) using Mini-Protean Tetra Cell electrophoresis system (Bio-Rad) at 100 V for 1 h. The gels were directly imaged in ChemiDoc MP imaging system.

Quenching-based kinetic assay for DNA unwinding

These assays were performed in 60 μ l helicase buffer with 20 nM DNA substrate (in molecules). The oligo F (49-mer) in DNA substrate was labelled with 6-fluorescein amidite (6-FAM) at the 5' end whereas oligo R (22-mer) was labelled at 3' end with rhodamine. The reactions were assembled on ice in 96-microwell plate and the recombinant proteins were directly added to their respective wells. The microplate was transferred to a microplate reader (CLARIOstar, BMG Labtech) at 37 °C and 6-FAM intensity was continuously monitored at every 30 s for 60 min. The final values were plotted as graphs using GraphPad PRISM.

Electrophoretic mobility shift assay

EMSA reactions (15 μ l) were performed in binding buffer containing 25 mM Tris-HCl pH 8.0, 2 mM ATP, 2 mM MgCl₂, 1 mM DTT, 50 mM NaCl, 0.1 mg ml⁻¹ BSA (New England Biolabs) and 5'-end-FITC-labelled 25 nM DNA substrate (in molecules). All of the steps except for the assembling reactions and protein addition were performed in the dark. The reactions were assembled on ice and recombinant proteins were added to reactions, mixed and incubated for 10 min at 37 °C in the dark. Reactions were supplemented with 5 μ l of EMSA loading buffer (50% glycerol, bromophenol blue) and resolved with 6% native TBE polyacrylamide gel (19:1 acrylamide-bisacrylamide, Bio-Rad) using the Mini-Protean Tetra Cell electrophoresis system (Bio-Rad) at 80 V for 45 min on ice. Finally, gels were imaged using the ChemiDoc MP imaging system. Scans of the gels are provided in the [Supplementary Information](#).

RPA stripping gel-based assay

The stripping assay was performed as described for EMSA except that the products were resolved at room temperature and longer 6% TBE native gel was used.

DNA annealing assay

DNA annealing assays were performed in 15 μ l annealing buffer containing 25 mM Tris-HCl pH 8.0, 2 mM ATP, 2 mM MgCl₂, 1 mM DTT, 50 mM NaCl and 0.1 mg ml⁻¹ BSA (New England Biolabs). All of the steps except for the assembling reactions and protein addition were performed in the dark. For DNA substrate, 10 nM (in molecules) complementary oligos (5'-FITC-oligo 1 and oligo 2) were separately incubated in 7.5 μ l annealing buffer and as indicated, with or without RPA on ice for 2 min. Recombinant proteins were added to FITC-oligo 1 reactions (7.5 μ l) on ice, immediately followed by the addition of oligo 2 reactions (7.5 μ l). Reactions were incubated for 10 min at 37 °C. The final concentration of both individual oligo and annealed dsDNA product was 5 nM. The reactions were stopped with 5 μ l of 2% stop solution (0.2% SDS, 30% glycerol, 150 mM EDTA, bromophenol blue) and 1 μ l

proteinase K (Roche, 18.4 mg ml⁻¹) and incubated for 10 min at 37 °C. To prevent the detection of spontaneous annealing during deproteination, 25-fold excess of unlabelled oligo 1 to FITC-oligo 1 was included in the 2% stop solution. The products were resolved and imaged identically as described for the unwinding assays.

Interaction assay

To study the interactions between HELQ and RAD51, MBP-HELQ-Flag baculovirus was expressed in 300 ml insect cells, and soluble extract from the collected pellet was prepared as described for the protein purification procedure. Reagent volumes for the preparation of the soluble extract were scaled down accordingly. The soluble extract was divided equally into two parts and incubated with amylose (E8021, NEB) and anti-Flag M2 resin (A2220, Sigma-Aldrich) for 1 h at 4 °C. Next, amylose resin and anti-Flag M2 resin were washed with wash buffer I (50 mM Tris-HCl pH 8.0, 1 mM DTT, 310 mM NaCl, 10% glycerol, 1 mM PMSF). Both resins were divided into 50 µl volumes in separate microtubes. 4 µg recombinant RAD51 was added to all except for one tube for each resin and incubated for 1 h at 4 °C. Resins were washed with wash buffer II (the same as wash buffer I but containing 100 mM NaCl). Proteins were eluted from resin in 1× SDS buffer by boiling at 95 °C for 4 min. The eluate was separated by 4–12% native SDS-PAGE gel (NuPAGE Bis-Tris, Invitrogen) and stained with instant blue Coomassie protein stain (Abcam).

DNA capture assay

The capture assays were performed in 20 µl DNA annealing buffer supplemented with 0.05% Tween-20. The reactions were assembled on ice and, where indicated, 82 nM RPA, 10 nM biotinylated dT43 (bio-dT43) and 10 nM 3'Cy3-dT79 were added to reactions. Next, HELQ was added to reactions as indicated. Reactions were mixed and incubated at 37 °C for 8 min in the dark. To pull-down bio-dT43, magnetic streptavidin beads were washed twice with PBS-0.1% Tween-20 (Dynabeads M-280, Thermo Fisher Scientific) and 5 µl of beads was added to each reaction. Reactions were further incubated for 4 min in the dark at room temperature and then washed twice with 80 µl washing buffer (25 mM Tris-HCl pH 8.0, 2 mM ATP, 2 mM MgCl₂, 1 mM DTT, 100 mM NaCl, 0.5 mg ml⁻¹ bovine serum albumin, NP-40) on a magnetic rack. Finally, the beads were resuspended in 30 µl loading buffer (7.5 µl 2% stop solution and 22.5 µl washing buffer) and boiled at 95 °C for 4 min. The samples were centrifuged at high speed for 1 min and 25 µl volume sample was loaded immediately on 10% native polyacrylamide gel and run as described for the unwinding assay. The gels were directly imaged in the ChemiDoc MP imaging system (Bio-Rad).

Substrate and flow cell preparation for SMI analysis

Experiments were performed using the commercially available C-trap (LUMICKS) set-up. Protein channels of the microfluidics chip were first passivated with BSA (0.1% (w/v) in PBS) and Pluronics F128 (0.5% (w/v) in PBS), with a minimum 500 μ l of both flowed through before use. Biotinylated ssDNA precursor was prepared as described previously³³. To generate gapped λ DNA, a protocol described previously was used⁹. In brief, biotinylated hairpin oligonucleotides were annealed to λ -dsDNA ends and ligated³⁴. *Streptococcus pyogenes* Cas9 D10A nickase (IDT) bound to previously described¹⁶ guide RNAs was subsequently used to generate targeted DNA nicks. The reaction was then stored at 4 °C and directly diluted in PBS on the day of the experiment. DNA was captured between 4.5 μ m SPHERO Streptavidin-coated polystyrene beads at 0.005% (w/v) using the laminar flow cell, stretched and held at forces of 100 pN (for ssDNA) or 65 pN (λ -gDNA 4/5) until the strands were fully melted. The presence of ssDNA and/or a ssDNA gap was verified by comparison with a built-in freely joined chain model. For confocal imaging, three excitation wavelengths were used—488 nm for eGFP and Alexa Fluor 488, 532 nm for Cy3 and 638 nm for Cy5, with emission detected in three channels with blue filter 512/25 nm, green filter 585/75 nm and red filter 640 LP.

Single-molecule DNA unwinding assay

For all the unwinding assays, the λ -gDNA 4/5 construct was held at a constant force of 50 pN. Beads and DNA were kept in PBS during the experiment, while DNA was melted in 0.5× NTM buffer (25 mM Tris-HCl pH 7.5, 50 mM NaCl, 0.5 mM MgCl₂) supplemented with 0.2 mg ml⁻¹ BSA and 1 mM DTT. HELQ (50 nM) and/or 25 nM RAD51(A488) were flowed into the system in 1× HELQ buffer (25 mM Tris-HCl pH 8.0, 2 mM MgCl₂, 50 mM NaCl) supplemented with 2 mM ATP, 0.2 mg ml⁻¹ BSA and 1 mM DTT. Unwinding was monitored by the change in the distance between the beads over time. To directly image fluorescent RAD51, the following image acquisition set-up was used: 4 μ W blue laser power, 0.1 ms px⁻¹ dwell time, 100 nm pixel size, 1 s interframe wait time.

SMI-based RPA/RAD51 stripping

Using optical tweezers in the stripping assays, the λ -gDNA 4/5 construct was held at a distance corresponding to a force of 10 pN after melting. The beads and DNA were kept in PBS during the experiment (microfluidic channels 1 and 2), while DNA was melted in PBS (microfluidic channel 2) and subsequently incubated with 5 nM RPA–eGFP or a mixture of 100 nM RAD51 and 100 nM Alx–RAD51 in 1× HELQ buffer (25 mM Tris-HCl pH 8.0, 2 mM MgCl₂, 50 mM NaCl), 0.2 mg ml⁻¹ BSA and 1 mM DTT in channel 3. Once RPA–eGFP or Alx–RAD51 were assembled on λ -gDNA 4/5 (after ~2 min of incubation) beads with DNA were moved to channel 4 containing 50

nM HELQ in 1× HELQ buffer (25 mM Tris-HCl pH 8.0, 2 mM MgCl₂, 50 mM NaCl) supplemented with 2 mM ATP, 0.2 mg ml⁻¹ BSA and 1 mM DTT. RPA-eGFP or Alx-RAD51 signal disappearance was monitored over time. Image acquisition setup was performed as follows: 1.6 µW blue laser power, 0.1 ms px⁻¹ dwell time, 100 nm pixel size, 1 s interframe wait time.

Single-molecule DNA oligonucleotide capture assay

Image processing and data analysis

Real-time force, distance and fluorescence data were exported from Bluelake HDF5 files and analysed using custom scripts in the Pylake Python package. Force was downsampled to 3 Hz for plotting. The worm-like chain model for λ -dsDNA was used as a reference for force–extension curve comparison. Bead distance–time traces were processed in GraphPad Prism 7. First derivative and smoothing of the traces were performed to extract individual unwinding stroke rates. Unwinding stroke rate distribution was analysed in GraphPad Prism 7 by fitting a single or double Gaussian curve. Dwell times and binding frequencies were estimated in Fiji. Dwell-time frequency distribution was analysed in GraphPad Prism 7. Mann–Whitney *U*-tests were used to assess statistical significance of the data where appropriate.

For the position, velocity and mean square displacement (MSD) analysis (Extended Data Fig. 4*i–k*), we used a custom-made single-particle tracking algorithm in Python (<https://github.com/singlemoleculegroup>). The sub-pixel position of the fluorescent particle in each frame of the kymograph was calculated by fitting the signal intensity

of a three-frame moving window with a 1D Gaussian function (linetime = 0.997 s, 100 nm px⁻¹).

For the obtained trajectories, the MSD was calculated using the following equation:

$$\text{MSD} = \frac{1}{N-n} \sum_{i=1}^{N-n} \frac{(X_{i+n} - X_i)^2}{t^{\alpha}}$$

where N is the total number of frames in the kymograph, n is the number of frames within a moving window (lag time) from which the square displacement was calculated (ranging from 1 to $N-1$) and X_i is the molecule position along the DNA in time.

To evaluate whether a trajectory represents random walk or directed motion, the MSD dependency was fitted with a power law. A particle exhibits free or constrained diffusion with rate D when the MSD scales with an exponent $\alpha \leq 1$. When $\alpha > 1$, the process is characterized as superdiffusive motion (for example, unidirectional walk).

To estimate the average velocity of the translocating molecule, the total route of the molecule (a sum of frame-to-frame displacements) was divided by the total trajectory time. Here, every trajectory was smoothed using the Savitzky–Golay filter (smoothing factor = 51) to eliminate tracking inaccuracies and the molecule's thermal fluctuations.

To estimate the loop size formed by the HELQ–RPA–DNA complex, the contour length after each unfolding event in the force–distance curve was fitted by the worm-like chain model. The difference between contour lengths of the neighbouring events corresponds with the loop size.

RPA stripping using single-molecule FRET

Flow chambers were prepared as described previously^{1,2}. Quartz slides and coverslips were passivated with polyethylene glycol (5% biotinylated) and flow chambers were constructed using double-sided sticky tape and sealed with epoxy. 5'-biotin- and internal amino linker-modified DNA oligonucleotides were labelled with Cy3-NHS ester and HPLC purified as previously described³. DNA (6 pM) was immobilized through biotin–streptavidin interactions. All of the experiments were performed in the standard HELQ buffer with addition of the PCA/PCD oxygen scavenger system with 5 mM PCA, 100 nM PCD and saturating Trolox. The flow chambers were imaged on a home-built, prism-based total internal reflection microscope with a 532 nm excitation laser (~3.8 mW), and images were acquired on an EM-CCD camera (Andor) with a 30 ms exposure time. FRET efficiencies were calculated from integrated donor (I_D) and acceptor (I_A) intensities as $\text{FRET} = I_A/(I_D + I_A)$ (refs. ^{1,3}). The images and data were

analysed using custom IDL, MATLAB and R scripts, which are available on request. FRET efficiency histograms were constructed by averaging the first 10 frames of each trajectory, with bins of 0.1. The dwell times of the free (high FRET) and bound (low FRET) states were measured, and dwell-time histograms were plotted. These were fit with single exponential fits to obtain average dwell times.

Cell culture

The U2OS human osteosarcoma cell line was grown in DMEM supplemented with 10% bovine growth serum, 2 mM l-glutamine, 100 µg ml⁻¹ streptomycin and 100 U ml⁻¹ penicillin. U2OS-EJDR cells were cultured in DMEM supplemented with 10% tetracycline-free fetal bovine serum, 2 mM l-glutamine, 100 µg ml⁻¹ streptomycin and 100 U ml⁻¹ penicillin. U2OS-DR cells contain a stably integrated DR-GFP reporter to measure DSB repair by HR as previously described³⁵. U2OS-SA cells contain a stably integrated SA-GFP reporter to measure DSB repair by SSA as previously described³⁶. U2OS-DR cells containing a stably integrated EJ-RFP reporter to measure DSB repair by mutagenic end-joining constitute the U2OS-EJDR cell line as previously described³⁷. U2OS-RFP-SCR cells contain a stably integrated RFP-SCR reporter for quantifying STGC and LTGC in HR as previously described³⁸.

U2OS-DR *HELQ*^{-/-} and U2OS-SA *HELQ*^{-/-} cells were generated using the CRISPR–Cas9 system. Knockouts were verified using Sanger sequencing and immunoprecipitation/western blot.

siRNA

The following siRNA oligonucleotides were used to transiently deplete *HELQ*: HELQ 1 (Qiagen FlexiTube siRNA, SI00435449); HELQ 2 (Horizon siGENOME SMARTpool, M-015379-01-0005); HELQ 3 (HelQ_M²³, CAAAGGAAGATTCCTCCAACCTAAA).

RAD52 was depleted using the On-Target plus SMART pool siRNA L-011760-00-0005 (Horizon). *BRCA2* was depleted using the On-Target plus SMART pool siRNA L-003462-00-0005 (Horizon). The On-Target plus non-targeting siRNA pool was used for non-targeting controls (D-001810-10-05, Horizon).

DSB repair assays

Cells (0.25×10^6) were reverse-transfected with 30 pmol siRNA using Lipofectamine RNAiMAX Reagent (Invitrogen) according to the manufacturer's instructions. After 48 h, cells were transfected with 2 µg of pCMV-ISceI-3×NLS or pCMV 3×NLS empty

vector and 30 pmol siRNA using Lipofectamine 2000 (Invitrogen). Cells were collected for analysis by flow cytometry at 72 h using the LSR Fortessa instrument (BD Biosciences). For each experiment, the percentage of GFP- or RFP-positive cells in the empty vector control was subtracted from the I-SceI-transfected cells. Data from each reporter assay represent the mean \pm s.e.m. of at least three independent experiments, and statistical analysis was performed using two-tailed paired *t*-tests. The Cas9-mediated DSB repair assay was used to introduce a Cas9-mediated site-specific DSB and break repair outcomes were detected using next-generation sequencing as previously described³⁹. In brief, cells were transfected with AAVS1 T2 CRISPR in pX330 (Addgene plasmid, 72833) using Lipofectamine 3000 (Invitrogen) and collected for genomic DNA extraction after 48 h. To measure SSTR, 1 μ l of a 10 μ M 121 bp donor oligonucleotide with three substitutions (purchased from IDT) was co-transfected with the CRISPR plasmid. A 201 bp PCR amplicon covering the expected Cas9 break site was sent for next-generation sequencing and reads were analysed for insertions, deletions and substitutions. NHEJ is defined as 1–5 bp deletion, MMEJ as >5 bp deletion with at least 2 bp microhomology, and SSTR as the introduction of three 1 bp substitutions. Data represent the mean \pm s.e.m. of at least four independent experiments, and statistical analysis was performed using two-tailed paired *t*-tests.

Immunoprecipitation and western blot

Cells were lysed in RIPA buffer (Teknova, R3792) with HALT protease inhibitor cocktail (Thermo Scientific Scientific). As HELQ is expressed at low levels in human cell lines and the commercially available antibodies tested did not dependably detect endogenous HELQ by western blot, we validated siRNA-mediated *HELQ* knockdown using HELQ immunoprecipitation. Cells (10^6) were transfected with 2 μ g siRNA by electroporation using the Amaxa nucleofector system, and plated into 150 mm dishes. After 72 h, cells were collected and whole-cell extracts were used for immunoprecipitation. Protein (2 mg) was incubated with 1 μ g anti-HEL308 antibodies (Novus Biologicals, NBP1-91842) at 4 °C overnight with rotation. After washing 0.25 mg Pierce Protein A/G Magnetic Beads (Thermo Fisher Scientific, 88802), the antigen sample–antibody mix was added to the beads and incubated for 1 h at room temperature with rotation. The beads were washed four times and eluted in SDS–PAGE reducing sample buffer (Invitrogen) for 10 min at 96 °C. The samples were loaded onto 4–12% Bis-Tris precast gels (Invitrogen) for SDS–gel electrophoresis and transferred onto Immobilon-P PVDF membrane (Millipore). Membranes were blocked for 1 h at room temperature with Pierce clear milk blocking buffer. For western blot analysis of RAD52 and BRCA2, 50 μ g of protein was loaded onto 10% Bis-Tris or 3–8% Tris-acetate precast gels (Invitrogen) for SDS–gel electrophoresis. Proteins were transferred onto a PVDF (polyvinylidene difluoride) membrane (BioRad) for BRCA2 detection. The iBlot Gel Transfer System (Invitrogen) was used to perform dry blotting of proteins onto nitrocellulose membranes for RAD52 detection. Membranes

were blocked for 1 h at room temperature with Pierce clear milk blocking buffer. Antibodies for the western blot analysis were as follows: anti-HEL308 (Santa Cruz Biotechnology, sc-81095), anti-RAD52 (Santa Cruz Biotechnology, sc-365341), anti-BRCA2 Ab-1 (Millipore Sigma-Aldrich, OP95), anti-SMC1 (Bethyl laboratories, A300-055A).

Gene expression

Cells were collected 72 h after siRNA transfection. Cell lysis, reverse transcription and quantitative PCR were performed using the TaqMan Gene Expression Cells-to-CT kit from Invitrogen according to the manufacturer's instructions. TaqMan Gene Expression Assays for *HELQ* (Hs01127906_m1) and ACTB endogenous control (Hs99999903_m1) were run in triplicate on the QuantStudio 6 Pro real-time PCR instrument (Applied Biosystems). Relative fold gene expression was calculated using the $\Delta\Delta C_t$ method. Data represent the mean \pm s.e.m. of at least three independent experiments, and statistical analysis was performed using two-tailed paired *t*-tests.

Immunofluorescence

WT U2OS-DR and *HELQ*^{−/−} U2OS-DR cells were seeded onto four-chamber tissue culture slides (Millipore) and treated the next day with 4 µM camptothecin or 10 Gy irradiation. For siRNA knockdown, cells were transfected with non-targeting or *HELQ* siRNA, incubated overnight and seeded onto chamber slides the next day. Cells were treated with the designated damaging agent 48 h after siRNA transfection. At the designated time points, cells were fixed, blocked and permeabilized. Cells were stained with the following antibodies: anti-phosphorylated-histone H2A.X (Ser139) (05-636, Millipore Sigma-Aldrich), anti-RAD51 (PC130 Millipore Sigma-Aldrich), anti-RPA32 (2208, Cell Signaling Technology), anti-phosphorylated-RPA32 (S4/S8) (ab87277, Abcam). Secondary antibodies were as follows: AlexaFluor 488-labelled goat anti-rabbit IgG, AlexaFluor 568-labelled donkey anti-mouse IgG, AlexaFluor 568-labelled goat anti-rat (Invitrogen). Images were obtained using a Zeiss LSM 880 confocal laser scanning microscope and analysed using ImageJ. At least 100 nuclei were counted per experiment and nuclei with >5 foci were scored as positive. Data represent the mean \pm s.e.m. of at least three independent experiments, and statistical analysis was performed using two-tailed paired *t*-tests.

Software

Chemidoc MP Image Lab Touch software (Bio-Rad, v.2.2.0.08) was used for gel imaging. Quenching-based kinetic unwinding assay data were collected using Clariostar BML Labtech (v.5.20 R5). Bluelake software from LUMICKS was used for data collection of SMI (LUMICKS). Similarly, for smFRET-based assays, NimOS

software from ONI was used. BD FACSDiva software was used with the BD Biosciences LSR Fortessa Analyzer for flow cytometry data acquisition. Zen 2.3 SP1 FP3 (black) v.14.0.18.201 was used for confocal microscopy image acquisition.

The quantification of gel-based data was carried using ImageJ (NIH v.1.52k). Mars Data analysis software (BML Labtech v.3.10 R6) was used for quenching-based kinetic unwinding assays. To analyse the SMI by optical tweezer, Pylake software from Lumicks was used. Furthermore, custom scripts were used for analysing some of unwinding/translocation assays carried out by SMI (<https://github.com/singlemoleclegroup>). smFRET analysis was performed using iSMS software (open source)³⁹. For microscopy data analysis, Image J (NIH, v.1.53e) was used. QuantStudio Design and Analysis Software v.2 was used with the QuantStudio 6 Pro real-time PCR instrument for relative gene expression analysis. Flow cytometry data were analysed using BD FlowJo (v.10.6.2). FlowJo was used to gate single cells and then select for GFP⁺ and/or RFP cells depending on the assay as shown. The background signal from the untransfected control was subtracted from each experiment. Representative plots from FlowJo showing the gating strategy are provided in [Supplementary Fig. 2](#). Cas9 DSB repair assay sequencing data were analysed as in Hussain et al.⁴⁰, using PEAR software for read stitching, BLOSUM62 for alignment and code for microhomology/deletion analysis available on GitHub (<https://github.com/cjsifuen/delmh>). Statistical analysis was performed using GraphPad Prism (v.8.2.1 and v.8.4.2). All schematics except for those of the quenching-based kinetic unwinding assays were generated using Adobe illustrator v.2.3. For the schematics of quenching-based kinetic unwinding assays, Biorender.com was used.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

The datasets generated during and/or analysed during the current study are included alongside the Article. Raw datasets for experiments performed on the C-TRAP are not included and are available from S.J.B. on reasonable request. Gel source data are provided in the [Supplementary Information](#). All data are archived at the Francis Crick Institute or Sloan Kettering Institute.

Code availability

The custom script made for determining the DNA unwinding/translocation by SMI is available at GitHub (<https://github.com/singlemoleclegroup>).

References

1. 1.
Khanna, K. K. & Jackson, S. P. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat. Genet.* **27**, 247–254 (2001).
2. 2.
Marini, F. & Wood, R. D. A human DNA helicase homologous to the DNA cross-link sensitivity protein Mus308. *J. Biol. Chem.* **277**, 8716–8723 (2002).
3. 3.
Adelman, C. A. et al. HELQ promotes RAD51 parologue-dependent repair to avert germ cell loss and tumorigenesis. *Nature* **502**, 381–384 (2013).
4. 4.
Luebben, S. W. et al. *Helq* acts in parallel to *Fancc* to suppress replication-associated genome instability. *Nucleic Acids Res.* **41**, 10283–10297 (2013).
5. 5.
Takata, K., Reh, S., Tomida, J., Person, M. D. & Wood, R. D. Human DNA helicase HELQ participates in DNA interstrand crosslink tolerance with ATR and RAD51 paralogs. *Nat. Commun.* **4**, 2338 (2013).
6. 6.
Jenkins, T. et al. The HelQ human DNA repair helicase utilizes a PWI-like domain for DNA loading through interaction with RPA, triggering DNA unwinding by the HelQ helicase core. *NAR Cancer* **3**, zcaa043 (2021).
7. 7.
Tafel, A. A., Wu, L. & McHugh, P. J. Human HEL308 localizes to damaged replication forks and unwinds lagging strand structures. *J. Biol. Chem.* **286**, 15832–15840 (2011).
8. 8.

Ward, J. D. et al. Overlapping mechanisms promote postsynaptic RAD-51 filament disassembly during meiotic double-strand break repair. *Mol. Cell* **37**, 259–272 (2010).

9. 9.

Belan, O. et al. Single-molecule analysis reveals cooperative stimulation of Rad51 filament nucleation and growth by mediator proteins. *Mol. Cell* **81**, 1058–1073 (2021).

10. 10.

Sugiyama, T., New, J. H. & Kowalczykowski, S. C. DNA annealing by RAD52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. *Proc. Natl Acad. Sci. USA* **95**, 6049–6054 (1998).

11. 11.

Kagawa, W., Kurumizaka, H., Ikawa, S., Yokoyama, S. & Shibata, T. Homologous pairing promoted by the human Rad52 protein. *J. Biol. Chem.* **276**, 35201–35208 (2001).

12. 12.

Ivanov, E. L., Sugawara, N., Fishman-Lobell, J. & Haber, J. E. Genetic requirements for the single-strand annealing pathway of double-strand break repair in *Saccharomyces cerevisiae*. *Genetics* **142**, 693–704 (1996).

13. 13.

Mortensen, U. H., Bendixen, C., Sunjevaric, I. & Rothstein, R. DNA strand annealing is promoted by the yeast Rad52 protein. *Proc. Natl Acad. Sci. USA* **93**, 10729–10734 (1996).

14. 14.

Reddy, G., Golub, E. I. & Radding, C. M. Human Rad52 protein promotes single-strand DNA annealing followed by branch migration. *Mutat. Res.* **377**, 53–59 (1997).

15. 15.

Brouwer, I. et al. Human RAD52 captures and holds DNA strands, increases DNA flexibility, and prevents melting of duplex DNA: implications for DNA recombination. *Cell Rep.* **18**, 2845–2853 (2017).

16. 16.

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).

17. 17.

Newton, M. D. et al. DNA stretching induces Cas9 off-target activity. *Nat. Struct. Mol. Biol.* **26**, 185–192 (2019).

18. 18.

Barroso-González, J. et al. RAD51AP1 is an essential mediator of alternative lengthening of telomeres. *Mol. Cell* **76**, 11–26 (2019).

19. 19.

Gibb, B. et al. Protein dynamics during presynaptic-complex assembly on individual single-stranded DNA molecules. *Nat. Struct. Mol. Biol.* **21**, 893–900 (2014).

20. 20.

Hussain, S. S. et al. Measuring nonhomologous end-joining, homologous recombination and alternative end-joining simultaneously at an endogenous locus in any transfectable human cell. *Nucleic Acids Res.* **49**, e74 (2021).

21. 21.

Gallagher, D. N. et al. A Rad51-independent pathway promotes single-strand template repair in gene editing. *PLoS Genet.* **16**, e1008689 (2020).

22. 22.

Gallagher, D. N. & Haber, J. E. Single-strand template repair: key insights to increase the efficiency of gene editing. *Curr. Genet.* **67**, 747–753 (2021).

23. 23.

Moldovan, G. L. et al. DNA polymerase POLN participates in cross-link repair and homologous recombination. *Mol. Cell. Biol.* **30**, 1088–1096 (2010).

24. 24.

Takata, K. et al. Conserved overlapping gene arrangement, restricted expression, and biochemical activities of DNA polymerase v (POLN). *J. Biol. Chem.* **290**, 24278–24293 (2015).

25. 25.

Nagaraju, G., Odate, S., Xie, A. & Scully, R. Differential regulation of short- and long-tract gene conversion between sister chromatids by Rad51C. *Mol. Cell. Biol.* **26**, 8075–8086 (2006).

26. 26.

Nagaraju, G., Hartlerode, A., Kwok, A., Chandramouly, G. & Scully, R. XRCC2 and XRCC3 regulate the balance between short- and long-tract gene conversions between sister chromatids. *Mol. Cell. Biol.* **29**, 4283–4294 (2009).

27. 27.

McIlwraith, M. J. & West, S. C. DNA repair synthesis facilitates RAD52-mediated second-end capture during DSB repair. *Mol. Cell* **29**, 510–516 (2008).

28. 28.

Lao, J. P., Oh, S. D., Shinohara, M., Shinohara, A. & Hunter, N. Rad52 promotes postinvasion steps of meiotic double-strand-break repair. *Mol. Cell* **29**, 517–524 (2008).

29. 29.

Mehta, A., Beach, A. & Haber, J. E. Homology requirements and competition between gene conversion and break-induced replication during double-strand break repair. *Mol. Cell* **65**, 515–526 (2017).

30. 30.

Spirek, M. et al. Human RAD51 rapidly forms intrinsically dynamic nucleoprotein filaments modulated by nucleotide binding state. *Nucleic Acids Res.* **46**, 3967–3980 (2018).

31. 31.

Modesti, M. Fluorescent labeling of proteins. *Methods Mol. Biol.* **1665**, 115–134 (2018).

32. 32.

Modesti, M. et al. Fluorescent human RAD51 reveals multiple nucleation sites and filament segments tightly associated along a single DNA molecule. *Structure* **15**, 599–609 (2007).

33. 33.

Candelli, A. et al. Visualization and quantification of nascent RAD51 filament formation at single-monomer resolution. *Proc. Natl Acad. Sci. USA* **111**, 15090–15095 (2014).

34. 34.

King, G. A., Burla, F., Peterman, E. J. G. & Wuite, G. J. L. Supercoiling DNA optically. *Proc. Natl Acad. Sci. USA* **116**, 26534–26539 (2019).

35. 35.

Pierce, A. J., Johnson, R. D., Thompson, L. H. & Jasin, M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev.* **13**, 2633–2638 (1999).

36. 36.

Stark, J. M., Pierce, A. J., Oh, J., Pastink, A. & Jasin, M. Genetic steps of mammalian homologous repair with distinct mutagenic consequences. *Mol. Cell. Biol.* **24**, 9305–9316 (2004).

37. 37.

Bindra, R. S., Goglia, A. G., Jasin, M. & Powell, S. N. Development of an assay to measure mutagenic non-homologous end-joining repair activity in mammalian cells. *Nucleic Acids Res.* **41**, e115 (2013).

38. 38.

Chandramouly, G. et al. BRCA1 and CtIP suppress long-tract gene conversion between sister chromatids. *Nat. Commun.* **4**, 2404 (2013).

39. 39.

Preus, S., Noer, S. L., Hildebrandt, L. L., Gudnason, D. & Birkedal, V. iSMS: single-molecule FRET microscopy software. *Nat. Methods* **12**, 593–594 (2015).

40. 40.

Hussain, S. S. et al. Measuring nonhomologous end-joining, homologous recombination and alternative end-joining simultaneously at an endogenous locus in any transfectable human cell. *Nucleic Acids Res.* **49**, e74 (2021).

Acknowledgements

We thank L. Krejci (MUNI, Masaryk University, Brno) and M. Modesti (CRCM, Marseille) for providing the expression constructs for RAD51 and RPA. R.A. was supported by an EMBO long-term fellowship (ALTF975-2018) and an SNF postdoc.mobility fellowship (P400PB-186718).

Author information

Author notes

1. These authors contributed equally: Roopesh Anand, Erika Buechelmaier

Affiliations

1. DSB Repair Metabolism Laboratory, The Francis Crick Institute, London, UK

Roopesh Anand, Ondrej Belan, Matthew Newton, Aleksandra Vancevska, Tohru Takaki & Simon J. Boulton

2. Memorial Sloan Kettering Cancer Center, New York, NY, USA

Erika Buechelmaier & Simon N. Powell

3. Graduate School of Medical Sciences, Weill Cornell Medicine, New York, NY, USA

Erika Buechelmaier

4. Department of Infectious Disease, Faculty of Medicine, Imperial College London, London, UK

Artur Kaczmarczyk & David S. Rueda

5. Single Molecule Imaging Group, MRC-London Institute of Medical Sciences, London, UK

Artur Kaczmarczyk & David S. Rueda

Contributions

R.A., E.B., S.N.P. and S.J.B. developed the concept for the paper, and D.S.R. advised on SMI experiments. R.A. purified all proteins and performed all bulk biochemical assays, including DNA unwinding, strand annealing and DNA capture experiments. T.T. and A.V. expressed and purified GST–BRC4 peptides. E.B. conducted all cellular assays including HR, SSA, MMEJ and GC tract-length assays. O.B. and M.N. performed and analysed SMI and FRET-based experiments. A.K. and D.S.R. analysed part of the SMI-based DNA capture experiments. S.J.B. and R.A. wrote the paper with editorial input from all of the other authors.

Corresponding authors

Correspondence to [David S. Rueda](#), [Simon N. Powell](#) or [Simon J. Boulton](#).

Ethics declarations

Competing interests

S.J.B. is a co-founder and VP Science Strategy at Artios Pharma Ltd., Babraham Research Campus, UK.

Peer review information

Nature thanks James Haber and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 HELQ specifically unwinds substrates with 3' overhang and D-loop.

a, SDS-PAGE gel (4–12% polyacrylamide) showing purified recombinant human HELQ WT and HELQ(K365M) from insect cells. The gel was stained with Coomassie brilliant blue (CBB). We used two and single preparations of HELQ WT and HELQ(K365M) respectively in this study. **b-d**, Representative native gels (10% polyacrylamide) of DNA unwinding assay of D-loop, Y-structure and lagging strand fork with indicated concentrations of HELQ. The asterisk indicates the position of FITC labelling at 5' end of oligo. **e**, Quantification of experiments such as shown in b-d and Fig. 1a with HELQ concentration ranging from 1–270 nM. n= 4 independent experiments; mean ± S.E.M. **f-g**, Representative native gels of DNA unwinding assay of dsDNA and 5'-overhang with indicated concentrations of HELQ. **h**, Native gel showing the DNA unwinding assay of 3'-overhang by HELQ in the presence of ATP (2 mM) and ATP γ S (2 mM), a poorly hydrolysable ATP analogue (n=2). **i**, Native gel showing DNA unwinding assay of 3'-overhang with indicated concentrations of HELQ(K365M) (n =2). **j**, Electrophoretic mobility shift assay (EMSA) with ssDNA and indicated concentrations of HELQ and HELQ(K365M). The final products were resolved with native 6% polyacrylamide gels. **k**, Quantification of experiments such as shown in j. n= 3 independent experiments; mean ± S.E.M. **l**, EMSA with dsDNA and indicated concentrations of HELQ and HELQ(K365M). **m**, Quantification of experiments such as shown in l. n= 3 independent experiments; mean ± S.E.M.

Extended Data Fig. 2 RAD51 interacts directly to HELQ and promotes its helicase activity.

a, SDS-PAGE gel (4-12%) showing purified recombinant human RAD51 from *E. coli*. The gel was stained with CBB. We used two preparations of RAD51 in this study. **b**, Protein interaction analysis of MBP-HELQ-Flag and RAD51 using both amylose and Flag pull-down assay. The final eluates were run on SDS-PAGE gel (4-12%) and stained with CBB. The interaction analysis was repeated 3 times with similar results. **c-e**, Representative native gels of DNA unwinding of D-loop, Y-structure and lagging strand fork substrates with HELQ (1 nM) and indicated concentrations of either RAD51 or RecA. **f**, Representative native gel of DNA unwinding of 3'-overhang with indicated concentrations of RecA and HELQ (1 nM). **g**, Quantification of experiments such as shown in f. n = 4 independent experiments; mean ± S.E.M. **h**, SDS-PAGE gel (4-12%) showing purified recombinant GST and GST-BRC4 peptides from *E. coli*. The gel was stained with CBB. Single preparation of GST and GST-BRC4 used in this study. **i-j**, EMSA gels (6% polyacrylamide) showing RAD51 binding to 3'-overhang and Y-structure in the presence and absence of indicated concentration of GST-BRC4 and GST peptides. **k**, Representative native gel of DNA unwinding of 3'-overhang by HELQ with RAD51 and indicated concentrations of GST-BRC4 and GST peptides. **l**,

Quantification of experiments such as shown in k. n = 4 independent experiments; mean ± S.E.M. **m**, Representative native gel of DNA unwinding of Y-structure by HELQ with RAD51 and indicated concentrations of GST-BRC4 and GST peptides. **n**, Quantification of experiments such as shown in m. n = 4 independent experiments; mean ± S.E.M. **o**, Representative native gel of DNA unwinding of Y-structure by HELQ with inhibitory concentration of RAD51 and indicated concentrations of for substrates GST-BRC4 and GST peptides. **p**, Quantification of experiments such as shown in o. n = 3 independent experiments; mean ± S.E.M. **q**, Schematics representation of quenching-based kinetic DNA unwinding assay of 3'-overhang. Initially, oligo F (49-mer), labelled at 5' end with fluorescein (F), is annealed with oligo R, which is labelled with rhodamine (R) at 3' end. Due to close proximity, FRET from fluorescein is quenched by rhodamine constitutively resulting in low FRET signal. Upon DNA unwinding, DNA strands are separated and thus rhodamine no longer able to quench fluorescein, results in higher FRET signal. **r**, Relative unwinding of 3'-overhang with HELQ and indicated concentrations of RAD51 as determined by quenching-based kinetic assay. n = 3 independent experiments; shaded area represents mean ± S.E.M.; black lines represent exponential or linear fits. **s**, Relative unwinding of 3'-overhang with HELQ and indicated concentrations of RecA as determined by quenching-based kinetic assay. n = 3 independent experiments; shaded area represents mean ± S.E.M.; black lines represent exponential or linear fits.

Extended Data Fig. 3 RPA inhibits HELQ DNA unwinding activity.

a, SDS-PAGE gel (4-12%) showing purified recombinant human RPA-mRFP1 from *E. coli*. The gel was stained with CBB. Single preparation of RPA-mRFP1 was used in this study. **b**, Representative gel of DNA unwinding assay of D-loop with HELQ in the absence and presence of RPA (20 nM). **c**, Quantification of experiments such as shown in b. n = 3 independent experiments; mean ± S.E.M. **d**, Representative gel of DNA unwinding assay of 3'-overhang with HELQ (1 nM) and indicated concentrations of RPA. **e**, Quantification of experiments such as shown in d. n = 3 independent experiments; mean ± S.E.M. **f**, Representative gel of DNA unwinding assay of 3'-overhang with HELQ (3 nM) and indicated concentrations of RAD51, in the absence and presence of RPA (20 nM). **g**, Quantification of experiments such as shown in f. n = 6 independent experiments; mean ± S.E.M. **h**, Representative gel of DNA unwinding assay of D-loop with HELQ (3 nM) and indicated concentrations of RAD51, in absence and presence of RPA (20 nM). **i**, Quantification of experiments such as shown in h. n = 5 independent experiments; mean ± S.E.M.

Extended Data Fig. 4 RAD51–HELQ complex translocates along ssDNA backbone.

a–b. Bead centre displacement measured between the traps as a function of time in indicated conditions. Traces represent individual DNA molecules ($n \geq 3$). **c.** Example of two real-time bead displacement traces for indicated conditions. Unwinding bursts of linearly increasing bead distance are interspersed by pauses where no distance change is observed. **d–e.** First derivative with applied smoothening of traces shown in **c**. Height of individual peaks corresponds to the rate of individual unwinding bursts. **f–g.** Histogram showing HELQ unwinding burst rate distribution in the absence ($n = 103$) and presence ($n = 97$) of RAD51. Black line in **f** represents single Gaussian fits. **h.** Chemical labelling of RAD51. RAD51 C319S was labelled in pH 7.0 using maleimide esters of Alexa Fluor 488. After reaction termination and purification of labelled species, labelling efficiency was assessed, and free dye component was evaluated using SDS-PAGE (4–12%) and subsequent fluorescent imaging. 1:1 labelling stoichiometry was achieved as measured spectrophotometrically. Proteins were labelled typically with 80–100% labelling efficiency. Single preparation of RAD51 C319S was used in this study. **i.** Displacement of proteins bound to individual DNA tethers. The representative trajectories illustrate either unidirectional movement (red), 1D diffusion (green) and static binding with occasional diffusion (blue). The positions of molecules in time were measured by fitting a moving window of three kymograph frames with Gaussian function. **j.** Total displacement of translocating HELQ-Alx–RAD51 complexes (red) used to calculate the translocation rate. Two populations of translocating molecules can be distinguished: faster with the mean rate of 14 ± 5 nm/s and slower with the mean rate of 4 ± 1 nm/s. The slower rate might result due to “pushing” of sparsely bound Alx–RAD51 on DNA by HELQ alone species. The total displacement of a representative RAD51 filament (green) includes all frame-to-frame displacements, regardless of directionality. The stationary molecule apparent displacement (blue) of 2 ± 1 nm/s results from thermal fluctuations of the tethered DNA. **k.** Mean Square Displacement calculated from the trajectories shown in panel **i**, plotted as a function of time interval for a period up to 35 sec. Fitting the MSD plot with a power law $D\Delta t^\alpha$ resulted in the factor $\alpha > 1$ for HELQ-RAD51 complexes (red) indicating their diffusive (directed) motion, in contrast to diffusing RAD51 and HELQ(K365M)-RAD51 complexes ($\alpha \leq 1$) (green, blue).

Extended Data Fig. 5 ATP is important for HELQ annealing activity in presence of RPA.

a, Representative gel of DNA unwinding assay of 3'-overhang with HELQ, in the presence and absence of “cold oligo” (25 nM) i.e., unlabelled oligo with identical DNA sequence as FITC-labelled oligo. **b**, Relative unwinding of 3' overhang with indicated concentrations of HELQ as determined by quenching-based kinetic assay. $n = 3$ independent experiments; shaded area represents mean \pm S.E.M.; black lines represent exponential or linear fits. **c**, Representative gel of DNA annealing assay with HELQ (10 nM) and indicated concentrations of RPA. The black and blue colours of

substrate represent complementary DNA strands. The asterisk indicates the position of FITC labelling at 5' end. The products were resolved on 10% native polyacrylamide gel. **d**, Quantification of experiments such as shown in c. n = 5 independent experiments; mean \pm S.E.M. **e**, Representative gel of DNA annealing assay with indicated concentrations of HELQ in the absence and presence of ATP. **f**, Quantification of experiments such as shown in e. n= 3 independent experiments; mean \pm S.E.M. **g**, Native gel showing DNA annealing assay with indicated concentrations of HELQ and RPA (40 nM) in the presence of ATP γ S (n=2). **h**, Representative gel of DNA annealing assay with indicated concentrations of HELQ(K365M) in the absence and presence of RPA (40 nM). **i**, Quantification of experiments such as shown in h. n = 4 independent experiments; mean \pm S.E.M. **j**, Representative gel of DNA annealing assay with HELQ(K365M) (60 nM) and indicated concentrations of RPA. **k**, Quantification of experiments such as shown in j. n = 3 independent experiments; mean \pm S.E.M. **l**, Representative gel of DNA annealing assay with HELQ (10 nM) and various concentrations of SSB. **m**, Quantification of experiments such as shown in l. n = 4 independent experiments; mean \pm S.E.M.

Extended Data Fig. 6 HELQ strips RPA from ssDNA.

a, Native gel (longer 6% polyacrylamide) showing RPA stripping assay with HELQ (10 nM) alone and with indicated concentrations of RPA. **b**, Native gel (longer 6% polyacrylamide) showing RPA stripping assay with indicated concentrations of HELQ in the absence and presence of RPA (40 nM). **c**, Schematic of immobilized dual labelled (Cy3 and Cy5) DNA in the absence of RPA. **d–e**, Representative intensity trajectory (top) and corresponding FRET trajectory (bottom) of dual labelled DNA in the absence of RPA. **f**, Time-binned FRET histogram of DNA only, fit with gaussian. **g**, Schematic of immobilized dual labelled DNA in the presence of RPA. **h–i**, Representative intensity trajectory (top) and corresponding FRET trajectory (bottom) of dual labelled DNA in the presence of RPA. Stable high FRET is observed. **j**, Time-binned FRET histogram of DNA in the presence of RPA. Fit with gaussian. **k**, Schematic of the experimental set up of single-molecule FRET-based RPA striping assay. DNA dual-labelled with the FRET pair Cy3 and Cy5 is immobilized on the microscope slide. In the absence of RPA, a short 6 nt sequence of homology causes the DNA to fold into a high FRET state. Upon RPA binding, the DNA unfolds resulting in a low FRET state. Addition of HELQ results in cycling between the low (bound) and high (free) FRET states as RPA is bound and removed respectively. **l**, Example of single-molecule fluorescence trajectory (top, Cy3 in blue, Cy5 in red) and corresponding FRET (bottom) showing the transition from low FRET (bound) to high FRET (free). **m**, Representative FRET trajectory of DNA in the presence of 1 nM RPA and 200 nM HELQ, spikes of high FRET correspond to RPA removal events. **n**, Plot of dwell time of RPA bound (t_{off}), low FRET, state with increasing HELQ

concentration. **o**, Plot of dwell time of free state (t_{on}), high FRET, on RPA removal with increasing HELQ concentration. **p**, Representative FRET trajectory of DNA in the presence of 1 nM RPA and 200 nM HELQ KM. **q**, Representative gel of DNA annealing assay with HELQ (3 nM), RPA (40 nM) and indicated concentrations of RAD51. **r**, Quantification of experiments such as shown in q. n = 3 independent experiments; mean \pm S.E.M.

Extended Data Fig. 7 HELQ weakly displaces RAD51 from ssDNA.

a, Representative EMSA gel showing RAD51 binding to ssDNA in the absence and presence of HELQ. Reactions were incubated at 37 °C for 20 min. **b**, Quantification of experiments such as shown in a. n = 6 independent experiments; mean \pm S.E.M. **c**, Representative EMSA gel showing RAD51 binding to dsDNA in the absence and presence of HELQ. Reactions were incubated at 37 °C for 20 min. **d**, Quantification of experiments such as shown in c. n = 3 independent experiments; mean \pm S.E.M. **e**, Schematic of the experimental set up of single-molecule FRET-based RAD51 removal assay. **f**, Representative kymographs of removal of Alx–RAD51 pre-bound to gapped DNA in the presence and the absence of 50 nM HELQ or HELQ(K365M). **g**, Removal of Alx–RAD51 measured from gapped ssDNA as a function of time in indicated conditions. Traces represent individual DNA molecules (n \geq 3).

Extended Data Fig. 8 HELQ can capture non-complementary DNA strands in presence of RPA.

a, SDS-PAGE gel (4-12%) showing purified recombinant human RPA–eGFP from *E. coli*. The gel was stained with CBB. Single preparation of RPA–eGFP used in this study. **b**, A schematic of DNA pulling process. RPA–eGFP-coated ss-λ DNA tethered between the two streptavidin beads (~4.8 μm diameter) was collapsed by bringing beads at <5 μm distance. Beads were subsequently pulled apart at constant speed (step size = 0.2 μm, frequency = 500 Hz). Force-extension curves were then recorded. **c**, Force-distance curves of individual eGFP-RPA-coated ss-λ DNA molecules recorded in the absence or presence of HELQ (left and right panels, respectively). Sawtooth-like patterns in the FD curves (red) indicate disruption of ssDNA loops held together by HELQ. The average loop sizes (1.5 \pm 0.5 μm, N = 10) were estimated from the differences in the fitted contour lengths between the disruption events (dashed lines). **d**, Native gel showing the capture assay with the indicated concentrations of HELQ(K365M) in the presence of RPA (82 nM). The experiment performed two times with similar results.

Extended Data Fig. 9 HELQ is epistatic with RAD52 for SSA and additive for second-end capture.

a, HELQ protein levels in U2OS WT cells, post 72 h post siRNA transfection, confirmed by HELQ-immunoprecipitation and western blot. **b**, HELQ protein levels in U2OS WT and *HELQ*^{-/-} cells, confirmed by HELQ-immunoprecipitation and western blot. **c**, *HELQ* gene expression after treatment with siRNA post 72 h in U2OS cells. siNT n = 9, siHELQ_1 n = 9, siHELQ_2 n = 3, siHELQ_3 n = 3 independent experiments ± S.E.M, compared to siNT. **d**, Schematic representation of DR-GFP reporter assay for measuring DSB repair by HR. **e**, I-SceI-induced HR frequency in U2OS DR *HELQ*^{-/-} cells and in U2OS-DR cells treated with the indicated siRNA. siNT/WT, n = 20; siHELQ_1, n = 5; siHELQ_2, n = 13; siHELQ_3, n = 10; *HELQ*^{-/-}, n = 7 independent experiments ± S.E.M, compared to siNT or WT cells. **f**, BRCA2 and RAD52 proteins levels 72 h post siRNA transfection confirmed by western blot. **g**, I-SceI-induced SSA frequency in U2OS-SA cells treated with the indicated siRNA. siNT, n = 21; siRAD52, n = 4; siHELQ_1, n = 4; siHELQ_2, n = 10; siHELQ_3, n = 3; siHELQ_1/RAD52, n = 4, siHELQ_2/RAD52, n = 8; siHELQ_3/RAD52, n = 3 independent experiments ± S.E.M, compared to siNT. **h**, Detection of formation of γH2AX, RPA2, and RAD51 foci after 10 Gy IR treatment in U2OS WT and *HELQ*^{-/-} cells. **i**, Quantification of RPA2, RAD51 and γH2AX foci in experiments as shown in h. n = 3 independent experiments ± S.E.M, compared to siNT. **j**, Detection of formation of γH2AX, RPA2, pRPA2 S4/8 and RAD51 foci after treatment with 4 μM CPT (Camptothecin) of U2OS WT and *HELQ*^{-/-} cells. **k**, Quantification of γH2AX, pRPA S4/8 foci and RAD51 foci in experiments as shown in j. n = 3 independent experiments ± S.E.M, compared to siNT. **l**, Quantification of formation of pRPA2 S4/8 and RAD51 foci after 6 h CPT treatment in U2OS cells treated with indicated siRNA. pRPA2 S4/8 n = 5, RAD51 n = 3 independent experiments ± S.E.M, compared to siNT. siNT, siRNA not-targeted. The statistical significance was determined by using two-tailed paired t test.

Extended Data Fig. 10 A possible mechanism of HELQ function in HR and SSA.

HR, In HR, following DSB, 5' DNA ends are resected by nucleases, which generates RPA-coated 3' overhangs. RPA is displaced by RAD51 giving rise to presynaptic nucleoprotein filament formation. This RAD51-filament invades homologous DNA duplex through its 3' end. DNA invasion result in D-loop formation, which is extended by DNA synthesis. If the SDSA pathway is invoked, HELQ can disrupt the D-loop to displace and/or anneal the extended strand with the broken duplex via its helicase and strand annealing activities. Alternatively, in DSBR, the D-loop can be greatly extended and RPA-coated DNA strands in D-loop and second 3' overhang can be captured by HELQ. During this second-end capture step, HELQ strips RPA from ssDNA and anneals complementary DNA strands together to prime DNA synthesis, restoring the broken DNA and resulting in either crossover or non-crossover products. **SSA**, similar to HR, longer RPA-coated 3' overhangs can be captured by HELQ. Post capture,

HELQ strips RPA from both ssDNAs and actively anneals the DNA strands together using base-pairing. The remaining flaps are cleaved, and nicks are ligated by specific nucleases and ligases completing the DSB repair reaction.

Supplementary information

Supplementary Information

Supplementary Figs. 1 and 2.

Reporting Summary

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Reprints and Permissions

About this article

Cite this article

Anand, R., Buechelmaier, E., Belan, O. *et al.* HELQ is a dual-function DSB repair enzyme modulated by RPA and RAD51. *Nature* **601**, 268–273 (2022).
<https://doi.org/10.1038/s41586-021-04261-0>

- Received: 02 June 2021
- Accepted: 17 November 2021
- Published: 22 December 2021

- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04261-0>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04261-0>

| [Section menu](#) | [Main menu](#) |

- Article
- [Published: 08 December 2021](#)

Structure and mechanism of the SGLT family of glucose transporters

- [Lei Han¹](#) na1,
- [Qianhui Qu^{1,2}](#) na1 nAff5,
- [Deniz Aydin](#) [ORCID: orcid.org/0000-0002-2758-2480](#)^{1,2,3,4} na1,
- [Ouliana Panova](#)^{1,2},
- [Michael J. Robertson](#) [ORCID: orcid.org/0000-0003-2610-680X](#)^{1,2},
- [Yan Xu](#)¹,
- [Ron O. Dror](#) [ORCID: orcid.org/0000-0002-6418-2793](#)^{1,2,3,4},
- [Georgios Skiniotis](#) [ORCID: orcid.org/0000-0003-0238-7846](#)^{1,2} &
- [Liang Feng](#) [ORCID: orcid.org/0000-0001-6171-2050](#)^{1,2}

[Nature](#) volume **601**, pages 274–279 (2022)

- 7238 Accesses
- 1 Citations
- 43 Altmetric
- [Metrics details](#)

Subjects

- [Cryoelectron microscopy](#)
- [Permeation and transport](#)

- [Transporters](#)

Abstract

Glucose is a primary energy source in living cells. The discovery in 1960s that a sodium gradient powers the active uptake of glucose in the intestine¹ heralded the concept of a secondary active transporter that can catalyse the movement of a substrate against an electrochemical gradient by harnessing energy from another coupled substrate. Subsequently, coupled Na⁺/glucose transport was found to be mediated by sodium–glucose cotransporters^{2,3} (SGLTs). SGLTs are responsible for active glucose and galactose absorption in the intestine and for glucose reabsorption in the kidney⁴, and are targeted by multiple drugs to treat diabetes⁵. Several members within the SGLT family transport key metabolites other than glucose². Here we report cryo-electron microscopy structures of the prototypic human SGLT1 and a related monocarboxylate transporter SMCT1 from the same family. The structures, together with molecular dynamics simulations and functional studies, define the architecture of SGLTs, uncover the mechanism of substrate binding and selectivity, and shed light on water permeability of SGLT1. These results provide insights into the multifaceted functions of SGLTs.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

[Subscribe to Journal](#)

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

[Buy article](#)

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Fig. 1: Overall structure of SGLT1.

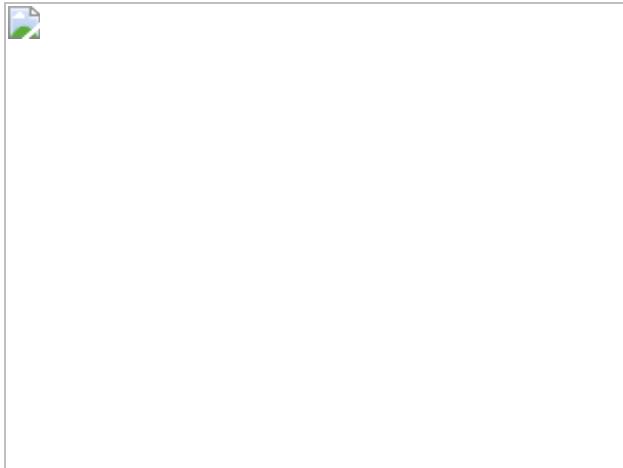


Fig. 2: SGLT1 substrate binding and water permeation.

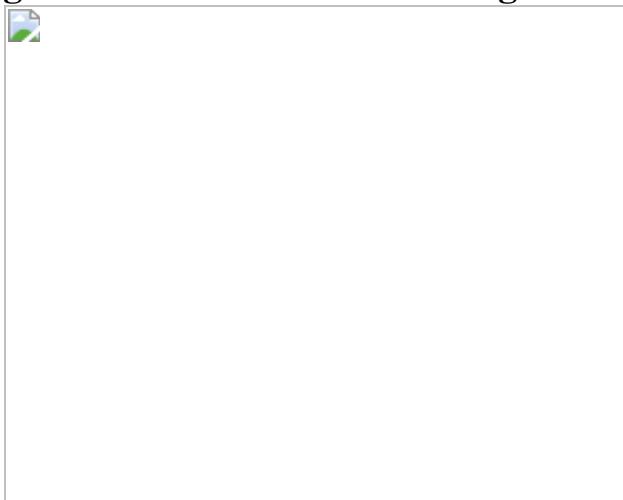


Fig. 3: Overall structure and substrate-binding pocket of SMCT1.

Data availability

The cryo-EM maps have been deposited into the Electron Microscopy Data Bank under accession numbers [EMD-25194](#), [EMD-25195](#) and [EMD-25196](#). The coordinates have been deposited into the Protein Data Bank under accession numbers [7SL8](#), [7SL9](#) and [7SLA](#).

References

1. 1.
Kleinzeller, A. & Kotyk, A. *Membrane Transport and Metabolism* (Publishing House of the Czechoslovak Academy of Sciences, 1961).
2. 2.
Wright, E. M., Loo, D. D. F. & Hirayama, B. A. Biology of human sodium glucose transporters. *Physiol. Rev.* **91**, 733–794 (2011).
3. 3.
Hopfer, U., Nelson, K. & Isselbacher, K. J. Specific glucose transport in isolated brush border membranes from rat small-intestine. *J. Biol. Chem.* **248**, 25–32 (1973).
4. 4.
Wright, E. M. Renal Na^+ -glucose cotransporters. *Am. J. Physiol. Renal Physiol.* **280**, F10–F18 (2001).
5. 5.
Hsia, D. S., Grove, O. & Cefalu, W. T. An update on sodium-glucose co-transporter-2 inhibitors for the treatment of diabetes mellitus. *Curr. Opin. Endocrinol. Diabetes Obes.* **24**, 73–79 (2017).
6. 6.
Hummel, C. S. et al. Glucose transport by human renal Na^+ /d-glucose cotransporters SGLT1 and SGLT2. *Am. J. Physiol. Cell Physiol.* **300**,

C14–C21 (2011).

7. 7.

Turk, E., Zabel, B., Mundlos, S., Dyer, J. & Wright, E. M. Glucose galactose malabsorption caused by a defect in the Na^+ /glucose cotransporter. *Nature* **350**, 354–356 (1991).

8. 8.

Hirschhorn, N. et al. Decrease in net stool output in cholera during intestinal perfusion with glucose-containing solutions. *N. Engl. J. Med.* **279**, 176–181 (1968).

9. 9.

Zeuthen, T., Gorraitz, E., Her, K., Wright, E. M. & Loo, D. D. F. Structural and functional significance of water permeation through cotransporters. *Proc. Natl Acad. Sci. USA* **113**, E6887–E6894 (2016).

10. 10.

Loo, D. D. et al. Passive water and ion transport by cotransporters. *J. Physiol.* **518**, 195–202 (1999).

11. 11.

Faham, S. et al. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+ /sugar symport. *Science* **321**, 810–814 (2008).

12. 12.

Wahlgren, W. Y. et al. Substrate-bound outward-open structure of a Na^+ -coupled sialic acid symporter reveals a new Na^+ site. *Nat. Commun.* **9**, 1753 (2018).

13. 13.

Leung, D. W., Turk, E., Kim, O. & Wright, E. M. Functional expression of the *Vibrio parahaemolyticus* Na^+ /galactose (vSGLT) cotransporter in *Xenopus laevis* oocytes. *J. Membr. Biol.* **187**, 65–70 (2002).

14. 14.

Canul-Tec, J. C. et al. Structure and allosteric inhibition of excitatory amino acid transporter 1. *Nature* **544**, 446–451 (2017).

15. 15.

Suzuki, T. et al. Apical localization of sodium-dependent glucose transporter SGLT1 is maintained by cholesterol and microtubules. *Acta Histochem. Cytochem.* **39**, 155–161 (2006).

16. 16.

Ghezzi, C., Calmettes, G., Morand, P., Ribalet, B. & John, S. Real-time imaging of sodium glucose transporter (SGLT1) trafficking and activity in single cells. *Physiol. Rep.* **5**, e13062 (2017).

17. 17.

Penmatsa, A., Wang, K. H. & Gouaux, E. X-ray structure of dopamine transporter elucidates antidepressant mechanism. *Nature* **503**, 85–90 (2013).

18. 18.

Gagnon, D. G., Bissonnette, P. & Lapointe, J. Y. Identification of a disulfide bridge linking the fourth and the seventh extracellular loops of the Na^+ /glucose cotransporter. *J. Gen. Physiol.* **127**, 145–158 (2006).

19. 19.

Wright, E. M. in *Genetic Diseases of the Kidney* (eds Lifton, R. P. et al.) 131–140 (2009).

20. 20.

Sala-Rabanal, M. et al. Bridging the gap between structure and kinetics of human SGLT1. *Am. J. Physiol. Cell Physiol.* **302**, C1293–C1305 (2012).

21. 21.

Bisignano, P. et al. Inhibitor binding mode and allosteric regulation of Na^+ -glucose symporters. *Nat. Commun.* **9**, 5245 (2018).

22. 22.

Gopal, E. et al. Expression of slc5a8 in kidney and its role in Na^+ -coupled transport of lactate. *J. Biol. Chem.* **279**, 44522–44532 (2004).

23. 23.

Miyauchi, S., Gopal, E., Fei, Y. J. & Ganapathy, V. Functional identification of SLC5A8, a tumor suppressor down-regulated in colon cancer, as a Na^+ -coupled transporter for short-chain fatty acids. *J. Biol. Chem.* **279**, 13293–13296 (2004).

24. 24.

Cui, D. & Morris, M. E. The drug of abuse γ -hydroxybutyrate is a substrate for sodium-coupled monocarboxylate transporter (SMCT) 1 (SLC5A8): characterization of SMCT-mediated uptake and inhibition. *Drug Metab. Dispos.* **37**, 1404–1410 (2009).

25. 25.

Gopal, E. et al. Transport of nicotinate and structurally related compounds by human SMCT1 (SLC5A8) and its relevance to drug

transport in the mammalian intestinal tract. *Pharm. Res.* **24**, 575–584 (2007).

26. 26.

Li, H. et al. SLC5A8, a sodium transporter, is a tumor suppressor gene silenced by methylation in human colon aberrant crypt foci and cancers. *Proc. Natl Acad. Sci. USA* **100**, 8412–8417 (2003).

27. 27.

Ganapathy, V. et al. Sodium-coupled monocarboxylate transporters in normal tissues and in cancer. *AAPS J.* **10**, 193–199 (2008).

28. 28.

Diez-Sampedro, A., Lostao, M. P., Wright, E. M. & Hirayama, B. A. Glycoside binding and translocation in Na^+ -dependent glucose cotransporters: comparison of SGLT1 and SGLT3. *J. Membr. Biol.* **176**, 111–117 (2000).

29. 29.

Choe, S., Rosenberg, J. M., Abramson, J., Wright, E. M. & Grabe, M. Water permeation through the sodium-dependent galactose cotransporter vSGLT. *Biophys. J.* **99**, L56–L58 (2010).

30. 30.

Vandenberg, R. J., Handford, C. A., Campbell, E. M., Ryan, R. M. & Yool, A. J. Water and urea permeation pathways of the human excitatory amino acid transporter EAAT1. *Biochem. J.* **439**, 333–340 (2011).

31. 31.

Turk, E. et al. Molecular characterization of *Vibrio parahaemolyticus* vSGLT: a model for sodium-coupled sugar cotransporters. *J. Biol. Chem.* **275**, 25711–25716 (2000).

32. 32.

Goehring, A. et al. Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat. Protoc.* **9**, 2574–2585 (2014).

33. 33.

Pardon, E. et al. A general protocol for the generation of nanobodies for structural biology. *Nat. Protoc.* **9**, 674–693 (2014).

34. 34.

McMahon, C. et al. Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat. Struct. Mol. Biol.* **25**, 289–296 (2018).

35. 35.

Hediger, M. A., Ikeda, T., Coady, M., Gundersen, C. B. & Wright, E. M. Expression of size-selected mRNA encoding the intestinal Na⁺/glucose cotransporter in *Xenopus laevis* oocytes. *Proc. Natl Acad. Sci. USA* **84**, 2634–2637 (1987).

36. 36.

Leung, D. W., Loo, D. D., Hirayama, B. A., Zeuthen, T. & Wright, E. M. Urea transport by cotransporters. *J. Physiol.* **528** 251–257 (2000).

37. 37.

Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

38. 38.

Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*

14, 331–332 (2017).

39. 39.

Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

40. 40.

Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

41. 41.

Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3–18 (2007).

42. 42.

Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

43. 43.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

44. 44.

Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

45. 45.

Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

46. 46.

The PyMOL Molecular Graphics System v.2.0. (Schrödinger, 2017).

47. 47.

Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

48. 48.

Goddard, T. D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

49. 49.

Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786 (2011).

50. 50.

Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. OPM: orientations of proteins in membranes database. *Bioinformatics* **22**, 623–625 (2006).

51. 51.

Jacobson, M. P., Friesner, R. A., Xiang, Z. & Honig, B. On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **320**, 597–608 (2002).

52. 52.

Betz, R. Dabble (v2.6.3). *Zenodo*
<https://zenodo.org/record/836914#.YZz6AWDP2M8> (2017).

53. 53.

Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).

54. 54.

Klauda, J. B. et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).

55. 55.

Guvench, O., Hatcher, E. R., Venable, R. M., Pastor, R. W. & Mackerell, A. D. CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Chem. Theory Comput.* **5**, 2353–2370 (2009).

56. 56.

Case, D. A., et al. AMBER 2018 (Univ. California, San Francisco, 2018).

57. 57.

Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015).

58. 58.

Jean-Paul Ryckaert, G. C., Herman, J. C. B. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).

59. 59.

Roe, D. R. & Cheatham, T. E. III. PTTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).

60. 60.

William Humphrey, A. D., Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

61. 61.

Nguyen, C. N., Young, T. K. & Gilson, M. K. Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J. Chem. Phys.* **137**, 044101 (2012).

62. 62.

Ramsey, S. et al. Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *J. Comput. Chem.* **37**, 2029–2037 (2016).

Acknowledgements

This work was made possible by support from Stanford University and the Harold and Leila Y. Mathers Charitable Foundation to L.F. and G.S., a Dean’s Fellowship to L.H., a Stanford Bio-X seed grant to R.O.D. and G.S., the SNSF Early Postdoctoral Mobility fellowship P2ELP3_187989 and the EMBO Long-Term Fellowship ALTF 544-2019 to D.A. L.F. was an NIH Director’s New Innovator awardee. Cryo-EM data for this work was collected at the Stanford-SLAC cryo-EM facility. We thank members of the Feng and Skiniotis laboratories for helpful discussions, Dror laboratory members J. Paggi, S. Eismann, M. Dämgen and M. Vögele for assistance with molecular dynamics simulations and analysis, and E. Montabana for support with data collection.

Author information

Author notes

1. Qianhui Qu

Present address: Shanghai Stomatological Hospital, Institutes of Biomedical Science, Department of Systems Biology for Medicine, Fudan University, Shanghai, China

2. These authors contributed equally: Lei Han, Qianhui Qu, Deniz Aydin

Affiliations

1. Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, USA

Lei Han, Qianhui Qu, Deniz Aydin, Ouliana Panova, Michael J. Robertson, Yan Xu, Ron O. Dror, Georgios Skiniotis & Liang Feng

2. Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA

Qianhui Qu, Deniz Aydin, Ouliana Panova, Michael J. Robertson, Ron O. Dror, Georgios Skiniotis & Liang Feng

3. Department of Computer Science, Stanford University, Stanford, CA, USA

Deniz Aydin & Ron O. Dror

4. Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA, USA

Deniz Aydin & Ron O. Dror

Contributions

L.H. performed molecular biology, protein expression and purification, biochemistry and functional studies. Q.Q. prepared cryo-EM grids, collected and processed data, and reconstructed the maps of SGLT1_{con}, SGLT1_{conHA} and SMCT1. L.H. built the model with the input from Q.Q. D.A. performed and analysed molecular dynamics simulations and contributed to preparation of the figures and manuscript. M.J.R. assisted

with model building. Y.X. performed biochemistry and protein engineering. O.P. prepared cryo-EM grids, collected and processed data, and reconstructed the map of SGLT1_{conHA}. R.O.D. oversaw the molecular dynamics simulations and contributed to the manuscript preparation. L.H., Q.Q., G.S. and L.F. wrote the manuscript with the input from all co-authors. G.S. and L.F. supervised the project.

Corresponding authors

Correspondence to [Georgios Skiniotis](#) or [Liang Feng](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature* thanks David Drew, Bernard Thorens and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

[Extended Data Fig. 1 Sequence alignments of selected SSS transporters.](#)

The sequences of selected SSS transporters were aligned using Clustal Omega (<http://www.uniprot.org/>) and adjusted manually. The secondary structural elements of SGLT1 are indicated above the sequence alignment. The mutations in SGLT1_{con} are highlighted.

Extended Data Fig. 2 Sequence alignments of SMCT homologs.

The sequences of selected SMCT homologs were aligned using Clustal Omega (<http://www.uniprot.org/>) and adjusted manually. The secondary structural elements of SMCT1 are indicated above the sequence alignment.

Extended Data Fig. 3 Glucose uptake of SGLT1 mutants.

a, Uptake activities of SGLT1_{con} and SGLT1_{conHA}. SGLT1_{conHA} is the same as SGLT1_{con}, except that W660 and G661 of SGLT1_{con} are reversed to H660 and A661. The uptake buffer contains different combinations of sodium (Na), choline (Ch), or phlorizin (Pz) as indicated. Uptake of α MDG is shown (mean \pm SEM; n=4 biological replicates). **b**, Uptake activities of SGLT1 and SGLT1-WG mutant, in which H660 and A661 are substituted with W660 and G661 (mean \pm SEM; n=4 biological replicates). **c**, Uptake activities of SGLT1 with mutations in the cholesterol-binding site (mean \pm SEM; n=4 biological replicates). **d**, Thermostability of SGLT1_{con} (left) or SGLT1 (right) and their variants under conditions with or without cholestryl hemisuccinate (CHS) (mean \pm SEM; n=3 biological replicates). **e**, The transport of α MDG by SGLT1 WT (left) and SGLT1_{conHA} (right) in the presence of various concentrations of α MDG. Data were plotted according to the equation, $U=U_{\max} \times [S] / (K_{0.5}+[S])$ (mean \pm SEM; n=3 biological replicates).

Extended Data Fig. 4 Cryo-EM sample preparation and data processing of SGLT1.

a, The elution profile of SGLT1_{con}-Nb1 on a size-exclusion column. The insert shows SDS-PAGE analysis of the purified sample. Data are representative of five independent experiments with similar results. **b**, Representative cryo-EM micrograph of SGLT1_{con}-Nb1 complex particles (from 15,039 micrographs with similar results). **c**, Selected 2D class averages of SGLT1_{con}-Nb1 complex (from 100 classes with similar results). **d, e**, The workflow of classification and refinement. The overall nominal resolutions of the SGLT1_{con}-Nb1 complex and the SGLT1_{conHA}-Nb1

complex were determined by the ‘gold standard’ FSC curve using the FSC=0.143 criterion. **f**, Overlay of the SGLT1_{con} model with the density map of SGLT1_{conHA}.

Extended Data Fig. 5 Cryo-EM densities and refined models.

a, Local resolution of the cryo-EM map of the SGLT1_{conHA}-Nb1 complex. **b**, Cryo-EM densities and model of SGLT1_{conHA} transmembrane helices and extracellular loops. **c**, Local resolution of the cryo-EM map of the SGLT1_{con}-Nb1 complex. **d**, Cryo-EM densities and model of SGLT1_{con} transmembrane helices. **e**, Local resolution of the cryo-EM map of the SMCT1-Nb2 complex. **f**, Cryo-EM densities and model of SMCT1 transmembrane helices.

Extended Data Fig. 6 Functional characterization, cryo-EM sample preparation, and data processing of SMCT1.

a, Time course of pyruvate uptake by SMCT1 expressing oocytes (mean ± SEM; n=3 biological replicates). **b**, The transport of pyruvate by SMCT1 in the presence of various concentrations of pyruvate (mean ± SEM; n=3 biological replicates). **c**, The size-exclusion chromatography of SMCT1-Nb2 complex and SDS-PAGE analysis of the purified sample. Data are representative of five independent experiments with similar results. **d**, Representative cryo-EM micrograph of the SMCT1-Nb2 complex (from 8,823 micrographs with similar results). **e**, Selected 2D class averages of SMCT1-Nb2 complex particles (from 100 classes with similar results). **f**, The workflow of data processing on the SMCT1-Nb2 complex. The overall nominal resolutions of the SMCT1-Nb2 complex were determined by the ‘gold standard’ FSC curve using the FSC=0.143 criterion.

Extended Data Fig. 7 Structural features of SGLT1 and substrate selectivity.

a, The density map of the cholesteryl hemisuccinate binding site of SGLT1_{conHA}. The cholesteryl hemisuccinate density map is indicated by

dashed red oval. **b**, Structural overlay of SGLT1_{conHA} (orange) and a dopamine transporter DAT (cyan, PDB: 4M48). The helices near the cholesterol binding site are shown as ribbons. **c**, The overall organization of the lid domain of SGLT1_{con}. Disulfide bonds are indicated by dashed blue circles. **d**, Close view of the four disulfide bonds. **e**, For five simulations of cholesterol-bound SGLT1_{con}, the heavy atom RMSD of cholesterol from its initial position is plotted over time. **f**, Conservation surface mapping of SGLT1. The conservation scores are calculated from 200 SGLT1 sequences using ConSurf. The cytosolic vestibule (middle) and extracellular cavity (right) are indicated by dashed yellow oval. **g**, Uptake activities of SGLT1 T287A variant (mean \pm SEM; n=3 biological replicates). WT and control are the same as in Fig. 2c). **h**, α MDG (left) and galactose (right) uptake activities of SGLT1 variants with point mutations in the substrate binding pocket on residues that are different between SGLT1 and vSGLT (mean \pm SEM; n=3 biological replicates).

Extended Data Fig. 8 Sodium-binding sites and SMCT1's substrate binding site.

a, Structural overlay of SGLT1_{conHA} (orange), SMCT1 (cyan), and SiaT (gray, PDB: 5NV9). The helices involved in sodium binding are shown as ribbons. The shift of TM helices from SiaT to SGLT1_{conHA} or SMCT1-butyrate is depicted by red arrows. **b**, Sodium-binding sites compared between SGLT1_{conHA} (yellow) or SMCT1 (yellow) and SiaT (gray). Residues in Na2 or Na3 sites are shown as sticks. **c**, Uptake activities of SGLT1 with mutations in the Na3-binding site. Uptake activities are normalized to WT (mean \pm SEM; n=4 biological replicates). **d**, SMCT1-butyrat interaction shown as Ligplot⁺ diagram (yellow dashed lines, hydrogen bonds; spokes, hydrophobic interactions). **e**, Pyruvate uptake activities of SMCT1 mutants compared to WT (mean \pm SEM; n=3 biological replicates).

Extended Data Fig. 9 Structural comparison among members of SSS family.

a, Superposition of the binding pockets of SMCT1 and apo SGLT1_{conHA}. Transmembrane helices involved in forming the central pockets are labeled. The relative positions of substrates are indicated by 3D shapes: pink hexagon for glucose and orange oval for butyrate. SMCT1 and SGLT1_{conHA} are colored in cyan and gray, respectively. **b**, Surface representation of the central cavity of SMCT1 (left) and SGLT1_{conHA} (right). Substrates are placed in the binding sites, the extension of which is indicated by dashed yellow ovals. **c**, The sequence comparison of the substrate binding site residues. For the sugar-transporting branch, the positions equivalent to SGLT1's sugar binding site residues are highlighted in orange. For the metabolite-transporting branch, the positions equivalent to SMCT1's substrate binding site residues are highlighted in blue. **d**, Sliced view of SGLT1_{conHA} (left panel) and SMCT1-butyrat (right panel). **e**, The superimposed SGLT1 (inward-facing) and SiaT (outward-facing). Zoomed-in view of regions that undergo considerable conformational changes are shown in blue boxes on the right (unrelated helices or loops are removed for clarity). The shift of TM helices between SiaT (gray) and SGLT1 (orange) is indicated by black arrows. From outward-open to inward-facing conformation, the N-terminal half of TM10 undergoes significant inward movement around a Gly-Pro-Pro motif at the center of the helix. Concomitantly, the short loop connecting TM9-TM10 and the C-terminal part of TM9 also moves inward. As a result, the N-terminal end of TM10 and the TM9-TM10 loop come into contact with TM2 and EL4 of the extracellular domain, which collapses the extracellular vestibule and stabilizes the closed conformation of the extracellular gate. Phe453^{TM10}, at the end of TM10, thus moves into a position to contact other extracellular gate residues to shield the substrate-binding pocket from the extracellular solution. In association, TM11 and TM12 tilt away to accommodate the movement of TM10. On the intracellular side, TM5, together with TM4, tilts outward while TM8 and TM9 tilt away from TM1, TM5 and TM6. These movements open the intracellular entrance of the vestibule and widen the permeation pathway to enable substrate release. The increased distance between TM8 and TM1/TM5 is linked with the disruption of both Na₂ and Na₃ sites. Thus, Na⁺ binding is coupled to the conformational changes during state transitions and glucose transport. During state transitions, the extracellular lid domain is also expected to undergo significant

conformational changes, which might help stabilize conformational states or give rise to distinct surface features to modulate Na⁺ transport. **f**, Structural comparison of SGLT1_{conHA} (orange), SMCT1 (cyan) and vSGLT (gray, PDB: 3DH4). **g**, Structural comparison of SGLT1_{conHA} (orange), SMCT1 (cyan). The orientation difference of the extracellular domain is zoomed-in in red box.

Extended Data Fig. 10 Substrate binding and water permeation of SGLT1.

a, For five simulations of SGLT1_{con} with glucose and two sodium ions initially placed in the glucose and sodium binding sites, the backbone RMSD of SGLT1_{con} from the cryo-EM structure is plotted over time. **b**, For five simulations of SGLT1_{con} with glucose and two sodium ions initially placed in the glucose and sodium binding sites, the heavy atom RMSD of glucose from its initial position is plotted over time. **c**, A frame from every 200 ns of the 2 μ s molecular dynamics simulation (simulation no. 1 in panels (**a**, **b**)) shows the position of glucose in the binding pocket (red at t = 0, transitioning to yellow at t = 2 μ s). **d**, Hydrogen bonds between glucose and SGLT1_{con} binding pocket residues are shown in black lines for simulation no. 1 in panels (**a**, **b**). **e**, Hydrogen bonds between glucose and the SGLT1_{con} binding pocket residues are shown as the percentage of total simulation time for all simulations in panels (**a**, **b**). **f**, Water occupancy averaged over time for glucose-bound SGLT1_{con} simulation no. 1 in panels (**a**, **b**) (left), and for a 4 μ s apo SGLT1_{con} simulation (right). Water density is shown in dark blue mesh (contoured at 0.0334 water molecules/ \AA^3 , approximately the bulk density), and glucose is shown in yellow sticks. **g**, The backbone RMSD of SGLT1_{con} from the cryo-EM structure (left), and the heavy atom RMSD of glucose from its initial position (right) for five simulations of SGLT1_{con} with glucose initially placed in the glucose binding site and with no sodium initially placed in the sodium binding sites. Sodium did not enter the sodium binding sites during the timescale of these simulations. **h**, Uptake activities of SGLT1 mutants. Oocyte-based urea uptake activities with or without inhibitor (200 μ M phlorizin, Pz) are compared to the wild-type (WT) transporter (mean \pm SEM; n=3 biological

replicates). **i**, Time course of urea uptake by SGLT1 WT and mutants with or without inhibitor phlorizin (Pz). Uptake data were plotted using linear regression (mean \pm SEM; n=3 biological replicates).

Extended Data Table 1 Cryo-EM data collection, refinement, and validation statistics

Supplementary information

Reporting Summary

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Han, L., Qu, Q., Aydin, D. *et al.* Structure and mechanism of the SGLT family of glucose transporters. *Nature* **601**, 274–279 (2022).
<https://doi.org/10.1038/s41586-021-04211-w>

- Received: 16 October 2019
- Accepted: 04 November 2021
- Published: 08 December 2021
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04211-w>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [Transporter-protein structures show how salt gets a sweet ride into cells](#)

- David Drew

Nature (2022)

[Transporter-protein structures show how salt gets a sweet ride into cells](#)

- David Drew

News & Views 08 Dec 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04211-w>

- Article
- [Published: 08 December 2021](#)

Structural basis of inhibition of the human SGLT2–MAP17 glucose transporter

- [Yange Niu¹](#),
- [Rui Liu ORCID: orcid.org/0000-0002-3758-6493¹](#),
- [Chengcheng Guan¹](#),
- [Yuan Zhang¹](#),
- [Zhixing Chen ORCID: orcid.org/0000-0001-5962-7359^{1,2}](#),
- [Stefan Hoerer³](#),
- [Herbert Nar ORCID: orcid.org/0000-0002-3878-6964³](#) &
- [Lei Chen ORCID: orcid.org/0000-0002-7619-8311^{1,2}](#)

Nature volume 601, pages 280–284 (2022)

- 5901 Accesses
- 1 Citations
- 36 Altmetric
- [Metrics details](#)

Subjects

- [Cryoelectron microscopy](#)
- [Kidney](#)
- [Metabolic disorders](#)

- [Receptor pharmacology](#)
- [Type 2 diabetes](#)

Abstract

Human sodium–glucose cotransporter 2 (hSGLT2) mediates the reabsorption of the majority of filtrated glucose in the kidney¹. Pharmacological inhibition of hSGLT2 by oral small-molecule inhibitors, such as empagliflozin, leads to enhanced excretion of glucose and is widely used in the clinic to manage blood glucose levels for the treatment of type 2 diabetes¹. Here we determined the cryogenic electron microscopy structure of the hSGLT2–MAP17 complex in the empagliflozin-bound state to an overall resolution of 2.95 Å. Our structure shows eukaryotic SGLT-specific structural features. MAP17 interacts with transmembrane helix 13 of hSGLT2. Empagliflozin occupies both the sugar-substrate-binding site and the external vestibule to lock hSGLT2 in an outward-open conformation, thus inhibiting the transport cycle. Our work provides a framework for understanding the mechanism of SLC5A family glucose transporters and also develops a foundation for the future rational design and optimization of new inhibitors targeting these transporters.

This is a preview of subscription content

Access options

Subscribe to nature+

Get immediate online access to the entire Nature family of 50+ journals

\$29.99

monthly

[Subscribe](#)

Subscribe to Journal

Get full journal access for 1 year

\$199.00

only \$3.90 per issue

[Subscribe](#)

All prices are NET prices.

VAT will be added later in the checkout.

Tax calculation will be finalised during checkout.

Buy article

Get time limited or full article access on ReadCube.

\$32.00

[Buy](#)

All prices are NET prices.

Additional access options:

- [Log in](#)
- [Learn about institutional subscriptions](#)

Fig. 1: Architecture of the hSGLT2–MAP17 complex.

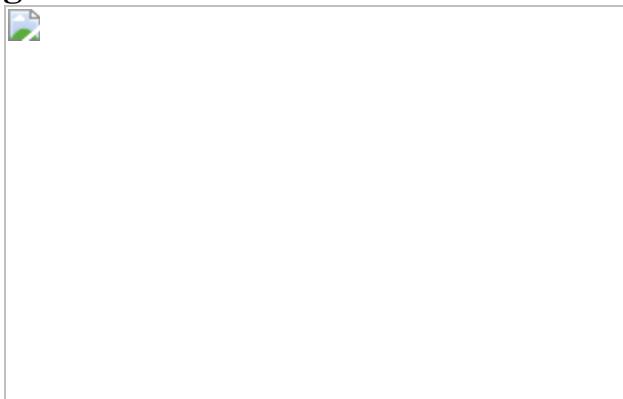


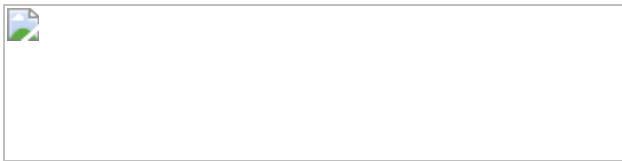
Fig. 2: The eukaryotic-specific features of hSGLT2.



Fig. 3: The empagliflozin-binding site of hSGLT2.



Fig. 4: Empagliflozin locks hSGLT2 in the outward-facing conformation.



Data availability

Cryo-EM maps and atomic coordinates of the hSGLT2–MAP17 complex have been deposited in the Electron Microscopy Data Bank and PDB under the ID codes [EMD-31558](#) and [7FEN](#), respectively. [Source data](#) are provided with this paper.

References

1. 1.

Ferrannini, E. Sodium-glucose co-transporters and their inhibition: clinical physiology. *Cell Metab.* **26**, 27–38 (2017).

2. 2.

Wright, E. M., Loo, D. D. & Hirayama, B. A. Biology of human sodium glucose transporters. *Physiol. Rev.* **91**, 733–794 (2011).

3. 3.

Coady, M. J. et al. MAP17 is a necessary activator of renal Na^+ /glucose cotransporter SGLT2. *J. Am. Soc. Nephrol.* **28**, 85–93 (2017).

4. 4.

Santer, R. & Calado, J. Familial renal glucosuria and SGLT2: from a mendelian trait to a therapeutic target. *Clin. J. Am. Soc. Nephrol.* **5**, 133–141 (2010).

5. 5.

Faillie, J. L. Pharmacological aspects of the safety of gliflozins. *Pharmacol. Res.* **118**, 71–81 (2017).

6. 6.

Choi, C. I. Sodium-glucose cotransporter 2 (SGLT2) inhibitors from natural products: discovery of next-generation antihyperglycemic agents. *Molecules* **21**, 1136 (2016).

7. 7.

Frampton, J. E. Empagliflozin: a review in type 2 diabetes. *Drugs* **78**, 1037–1048 (2018).

8. 8.

Ghezzi, C. et al. SGLT2 inhibitors act from the extracellular surface of the cell membrane. *Physiol. Rep.* **2**, e12058 (2014).

9. 9.

Faham, S. et al. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+ /sugar symport. *Science* **321**, 810–814 (2008).

10. 10.

Watanabe, A. et al. The mechanism of sodium and substrate release from the binding pocket of vSGLT. *Nature* **468**, 988–991 (2010).

11. 11.

Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).

12. 12.

Kirchhofer, A. et al. Modulation of protein properties in living cells using nanobodies. *Nat. Struct. Mol. Biol.* **17**, 133–138 (2010).

13. 13.

Chang, H. C. et al. Development of a novel non-radioactive cell-based method for the screening of SGLT1 and SGLT2 inhibitors using 1-NBDG. *Mol. Biosyst.* **9**, 2010–2020 (2013).

14. 14.

Patel, A., Toso, D., Litvak, A. & Nogales, E. Efficient graphene oxide coating improves cryo-EM sample preparation and data collection from tilted grids. Preprint at <https://doi.org/10.1101/2021.03.08.434344> (2021).

15. 15.

Wang, N. et al. Structural basis of human monocarboxylate transporter 1 inhibition by anti-cancer drug candidates. *Cell* **184**, 370–383.e13 (2021).

16. 16.

Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na^+/Cl^- -dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).

17. 17.

Jungnickel, K. E. J., Parker, J. L. & Newstead, S. Structural basis for amino acid transport by the CAT family of SLC7 transporters. *Nat. Commun.* **9**, 550 (2018).

18. 18.

Yan, R., Zhao, X., Lei, J. & Zhou, Q. Structure of the human LAT1-4F2hc heteromeric amino acid transporter complex. *Nature* **568**, 127–130 (2019).

19. 19.

Lee, Y. et al. Cryo-EM structure of the human l-type amino acid transporter 1 in complex with glycoprotein CD98hc. *Nat. Struct. Mol. Biol.* **26**, 510–517 (2019).

20. 20.

Gagnon, D. G., Bissonnette, P. & Lapointe, J. Y. Identification of a disulfide bridge linking the fourth and the seventh extracellular loops of the Na⁺/glucose cotransporter. *J. Gen. Physiol.* **127**, 145–158 (2006).

21. 21.

Wright, E. M., Turk, E. & Martin, M. G. Molecular basis for glucose-galactose malabsorption. *Cell Biochem. Biophys.* **36**, 115–121 (2002).

22. 22.

Santer, R. et al. Molecular analysis of the SGLT2 gene in patients with renal glucosuria. *J. Am. Soc. Nephrol.* **14**, 2873–2882 (2003).

23. 23.

Calado, J. et al. Twenty-one additional cases of familial renal glucosuria: absence of genetic heterogeneity, high prevalence of

private mutations and further evidence of volume depletion. *Nephrol. Dial. Transplant.* **23**, 3874–3879 (2008).

24. 24.

Bhattacharya, S. et al. An exhaustive perspective on structural insights of SGLT2 inhibitors: a novel class of antidiabetic agent. *Eur. J. Med. Chem.* **204**, 112523 (2020).

25. 25.

Turner, R. J. & Silverman, M. Sugar uptake into brush border vesicles from normal human kidney. *Proc. Natl Acad. Sci. USA* **74**, 2825–2829 (1977).

26. 26.

Coady, M. J., Wallendorff, B. & Lapointe, J. Y. Characterization of the transport activity of SGLT2/MAP17, the renal low-affinity Na^+ -glucose cotransporter. *Am. J. Physiol. Renal Physiol.* **313**, F467–F474 (2017).

27. 27.

Abramson, J. & Wright, E. M. Structure and function of Na^+ -symporters with inverted repeats. *Curr. Opin. Struct. Biol.* **19**, 425–432 (2009).

28. 28.

Penmatsa, A. & Gouaux, E. How LeuT shapes our understanding of the mechanisms of sodium-coupled neurotransmitter transporters. *J. Physiol.* **592**, 863–869 (2014).

29. 29.

Forrest, L. R. & Rudnick, G. The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters. *Physiology* **24**,

377–386 (2009).

30. 30.

Penmatsa, A., Wang, K. H. & Gouaux, E. X-ray structure of dopamine transporter elucidates antidepressant mechanism. *Nature* **503**, 85–90 (2013).

31. 31.

Coleman, J. A., Green, E. M. & Gouaux, E. X-ray structures and mechanism of the human serotonin transporter. *Nature* **532**, 334–339 (2016).

32. 32.

Li, N. et al. Structure of a pancreatic ATP-sensitive potassium channel. *Cell* **168**, 101–110.e10 (2017).

33. 33.

Scheich, C., Kummel, D., Soumailakakis, D., Heinemann, U. & Bussow, K. Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res.* **35**, e43 (2007).

34. 34.

Goehring, A. et al. Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat. Protoc.* **9**, 2574–2585 (2014).

35. 35.

Nasr, M. L. et al. Covalently circularized nanodiscs for studying membrane proteins and viral entry. *Nat. Methods* **14**, 49–52 (2017).

36. 36.

Hummel, C. S. et al. Glucose transport by human renal Na⁺/d-glucose cotransporters SGLT1 and SGLT2. *Am. J. Physiol. Cell Physiol.* **300**, C14–C21 (2011).

37. 37.

Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

38. 38.

Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

39. 39.

Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

40. 40.

Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

41. 41.

Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).

42. 42.

Biasini, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).

43. 43.

Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

44. 44.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

45. 45.

Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

Acknowledgements

We thank J. Ye and W. Zhang from Boehringer Ingelheim International GmbH for assistance; L.-C. Hsu from the School of Pharmacy, National Taiwan University for providing the 1-NBD-glucose sample for our initial pilot experiment; and S. Huang for assistance with the radioactive experiments. Cryo-EM data collection was supported by the Electron Microscopy Laboratory and Cryo-EM Platform of Peking University with the assistance of X. Li, Z. Guo, B. Shao, X. Pei and G. Wang. Part of the structural computation was also performed on the Computing Platform of the Center for Life Science and High-performance Computing Platform of Peking University. We thank the National Center for Protein Sciences at Peking University in Beijing, China, for assistance with negative-stain EM. The work is supported by grants from the National Natural Science Foundation of China (91957201 and 31870833 to L.C., 31821091 to L.C. and Z.C., and 31971375 to Z.C.) and Beijing Municipal Science & Technology Commission Project (Z201100005320017 to Z.C.). C.G. was partially supported by a postdoctoral fellowship sponsored by Boehringer Ingelheim International GmbH.

Author information

Affiliations

1. State Key Laboratory of Membrane Biology, College of Future Technology, Institute of Molecular Medicine, Peking University, Beijing Key Laboratory of Cardiometabolic Molecular Medicine, Beijing, China

Yange Niu, Rui Liu, Chengcheng Guan, Yuan Zhang, Zhixing Chen & Lei Chen

2. Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China

Zhixing Chen & Lei Chen

3. Boehringer-Ingelheim Pharma, GmbH & Co KG, Biberach, Germany

Stefan Hoerer & Herbert Nar

Contributions

L.C. initiated the project and wrote the draft of the manuscript. C.G. screened the SGLT homologues. Y.N. purified protein, prepared the cryo-EM sample, processed data and carried out functional assays. Y.N. and R.L. collected the cryo-EM data. L.C. built and refined the model. Y.Z. and Z.C. synthesized 1-NBD-glucose. S.H. and H.N. provided empagliflozin and [³H]-empagliflozin. All authors contributed to the preparation of the manuscript.

Corresponding author

Correspondence to [Lei Chen](#).

Ethics declarations

Competing interests

S.H. and H.N. are employees of Boehringer Ingelheim Pharma, GmbH & Co KG.

Additional information

Peer review information *Nature* thanks David Drew and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

[Extended Data Fig. 1 Biochemical characterization of the human SGLT2-MAP17 heterodimer.](#)

a, FSEC traces of GFP-tagged hSGLT2 (red) alone, in complex with non-tagged MAP17 (green) and in complex with C-terminal MBP-mScarlet tagged MAP17 (blue). The fluorescence signals were monitored in the GFP channel. **b**, HPLC chromatograph of synthesized 1-NBD-glucose, showing high purity. ESI-MS: *m/z* calculated for $C_{12}H_{14}N_4O_8 [M+H]^+$ 343.09, found 343.18. **c**, [3 H]-empagliflozin binding to the purified hSGLT2-MAP17 protein. Unlabeled empagliflozin was used for competition. CPM, counts per minute. (Data are shown as means \pm standard deviations, $n = 3$ biologically independent samples). **d**, Size-exclusion chromatography profile of the hSGLT2-MAP17 complex in nanodisc in the presence of empagliflozin. Fractions between dashes were used for cryo-EM sample preparation. **e**, SDS-PAGE gel of purified hSGLT2-MAP17 protein in the nanodisc. Asterisk indicates the fraction for cryo-EM sample preparation. The purifications were repeated 3 times with similar results. (For gel source data, see Supplementary Fig. 1.).

[Source data](#)

Extended Data Fig. 2 Cryo-EM image analysis of the hSGLT2-MAP17 complex.

a, Representative raw micrograph of hSGLT2-MAP17 complex in nanodisc. The other 7899 micrographs are similar to this one. **b**, 2D class averages of hSGLT2-MAP17 complex output from cryoSPARC. **c**, Data processing flow chart. **d**, The locations of GFP (green) and nanobody (brown) shown in 4 Å low-pass filtered cryo-EM map. **e**, Gold-standard Fourier shell correlation curves of the final reconstruction. **f**, Angular distributions of the final reconstruction. **g**, Local resolution distribution of the hSGLT2-MAP17 complex. **h**, The cut-open view of local resolution distribution of the hSGLT2-MAP17 complex.

[Source data](#)

Extended Data Fig. 3 Representative cryo-EM density maps of the hSGLT2-MAP17 complex.

The densities of transmembrane helices, empagliflozin, and putative palmitoylated Cys residues are contoured at 1.3σ , 1.0σ and 1.0σ , respectively.

Extended Data Fig. 4 Sequence alignment of SGLT.

a, The sequences of the *Homo sapiens* SGLT2 (hSGLT2), *Mus Musculus* SGLT2 (mSGLT2), *Homo sapiens* SGLT1 (hSGLT1), and *Vibrio parahaemolyticus* SGLT (vSGLT) are aligned using MEGA-X. α helices are shown as cylinders. Unmodeled residues are shown as dashed lines. Conserved residues are colored from cyan to blue. Residues that form disulfide bonds are highlighted in yellow. FRG mutations on extracellular protrusion are indicated by asterisks. The residues that are mutated in Fig. 3g are boxed in red. The colors of cylinders are the same as in Fig. 1d. **b**, The sequences of transmembrane helices of *Homo sapiens* MAP17, *Homo sapiens* SMIM24 are aligned using MEGA-X. Conserved residues are colored from cyan to blue. Residues that interact with hSGLT2 are boxed in red. **c**, The superposition of hSGLT2 (colored) with vSGLT (gray). The

scaffold domain of hSGLT2 is colored in green. The bundle helices of TM1, TM2, TM6 and TM7 of hSGLT2 are colored in brown, blue, pink and red respectively. **d**, A 90° rotated view of **c**. **e**, Interactions between hydroxyl groups of empagliflozin and hSGLT2. The helices of hSGLT2 are colored the same as in **c**. Empagliflozin and its interacting residues are shown as sticks. Putative hydrogen bonds are depicted as black dashed lines. **f**, Interactions of D-galactose with vSGLT. The helices of vSGLT are numbered according to LeuT nomenclature for comparison. D-galactose and its interacting residues are shown as sticks. Hydrogen bonds are depicted as black dashed lines. **g**, Top view of the cross-section of the transmembrane domain of the hSGLT2-MAP17 complex at the approximate level indicated by the dashed lines in Fig. [1c](#), colored the same as in Fig. [1c](#). The numbers of transmembrane helices are labeled above. **h**, Top view of the cross-section of the transmembrane domain of the GkApcT-MgtS complex. GkApcT is colored in grey and MgtS is colored in yellow. **i**, Top view of the cross-section of the transmembrane domain of the LAT1-4F2hc complex. LAT1 is colored in grey and the transmembrane helix of 4F2hc is colored in yellow.

[Extended Data Fig. 5 Distributions of human disease mutations.](#)

a, hSGLT1 mutations found in glucose galactose malabsorption human patients are mapped onto the homology structural model of hSGLT1. C α positions of mutations are shown as purple spheres. hSGLT1 is shown in blue ribbons. **b**, A 180° rotated view of **a**. **c**, hSGLT2 mutations found in familial renal glycosuria patients are mapped onto the structure of hSGLT2. C α positions of mutations are shown as purple spheres. hSGLT2 is shown in brown ribbons. **d**, A 180° rotated view of **c**.

[Extended Data Fig. 6 Chemical structures of related compounds.](#)

a, Chemical structures of glucose and its various analogues. Carbons of D-glucose are numbered in blue. The C2-OH and C3-OH groups of glucose

that are important for hSGLT2 binding are shown in red dashed circles. **b**, Chemical structures of representative SGLT2i.

Extended Data Fig. 7 Electron density map of hSGLT2 at the putative sodium-binding site.

a, The electron density map of hSGLT2 is contoured at 1.8σ . **b**, a 180° rotated view of **a**.

Extended Data Fig. 8 The model of hSGLT2 inhibition by SGLT2i.

a, hSGLT2 is shown in blue. MAP17 is shown in red. SGLT2i is shown in green. Glucose is shown in yellow. Sodium ions are shown as purple spheres. **b**, The cut-open view of hSGLT2 shown with the surface colored in pink. Empagliflozin is shown as green spheres. **c**, The cut-open view of SERT (PDB ID: 5I73) with the surface colored in brown. (S)-citalopram is shown as green spheres. **d**, The cut-open view of DAT (PDB ID: 4M48), with the surface colored in red. Nortriptyline is shown as green spheres.

Extended Data Table 1 Cryo-EM data collection, refinement and validation statistics

Extended Data Table 2 The potencies of empagliflozin on various hSGLT2 constructs

Supplementary information

Supplementary Fig. 1

The uncropped gel of Extended Data Fig. 1e with size marker indicated.

Reporting Summary

Source data

[**Source Data Fig. 1**](#)

[**Source Data Fig. 3**](#)

[**Source Data Extended Data Fig. 1**](#)

[**Source Data Extended Data Fig. 2**](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Niu, Y., Liu, R., Guan, C. *et al.* Structural basis of inhibition of the human SGLT2–MAP17 glucose transporter. *Nature* **601**, 280–284 (2022).
<https://doi.org/10.1038/s41586-021-04212-9>

- Received: 13 September 2021
- Accepted: 03 November 2021
- Published: 08 December 2021
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04212-9>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

Further reading

- [Transporter-protein structures show how salt gets a sweet ride into cells](#)
 - David Drew

Nature (2022)

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04212-9>

| [Section menu](#) | [Main menu](#) |

Amendments & Corrections

- [**Author Correction: Schizophrenia risk from complex variation of complement component 4**](#) [15 December 2021]
Author Correction •
- [**Publisher Correction: Burden and characteristics of COVID-19 in the United States during 2020**](#) [17 December 2021]
Publisher Correction •
- [**Publisher Correction: Disease variant prediction with deep generative models of evolutionary data**](#) [17 December 2021]
Publisher Correction •

- Author Correction
- [Published: 15 December 2021](#)

Author Correction: Schizophrenia risk from complex variation of complement component 4

- [Aswin Sekar^{1,2,3},](#)
- [Allison R. Bialas^{4,5},](#)
- [Heather de Rivera^{1,2},](#)
- [Avery Davis^{1,2},](#)
- [Timothy R. Hammond⁴,](#)
- [Nolan Kamitaki^{1,2},](#)
- [Katherine Tooley^{1,2},](#)
- [Jessy Presumey⁵,](#)
- [Matthew Baum^{1,2,3,4},](#)
- [Vanessa Van Doren¹,](#)
- [Giulio Genovese^{1,2},](#)
- [Samuel A. Rose²,](#)
- [Robert E. Handsaker^{1,2},](#)
- [Schizophrenia Working Group of the Psychiatric Genomics Consortium](#),
- [Mark J. Daly^{2,6},](#)
- [Michael C. Carroll⁵,](#)
- [Beth Stevens^{2,4} &](#)
- [Steven A. McCarroll^{1,2}](#)

[Nature](#) volume 601, pages E4–E5 (2022)

- 1078 Accesses
- 1 Altmetric
- [Metrics details](#)

The [Original Article](#) was published on 27 January 2016

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/nature16549> Published online 27 January 2016

Figure 7d and Supplementary Fig. 10 of this Article reported experiments on synaptic refinement in the mouse lateral geniculate nucleus (LGN). We identified an error in the description of this experiment in the figure legends and text; the error does not affect the results or conclusions. The text and figure legends described comparisons of *C4*-mutant mice to wild-type littermate controls; the data analysis in fact drew upon two litters from het × het crosses and three additional knockout (KO) and wild-type animals (total $n = 17$ animals; wild-type $n = 5$, het $n = 7$, KO $n = 5$), as in the following Table 1.

Table 1 Corrected listing of animals analyzed by type

Animals were age-matched, and images from animals were analysed blind to genotype. The data and quantification/barplots in the original paper correspond to these animals and are thus unchanged.

We correct the descriptions of these experiments as follows:

In the main text, fifth paragraph of subsection “*C4* in the central nervous system,” the third sentence originally reading “We found that in mice deficient in *C4* (ref. 36), C3 immunostaining in the dLGN was greatly reduced compared to wild-type littermates (Fig. 7a, b)...,” the text “wild-type littermates” should instead have read “age-matched wild-type controls.” In the sixth paragraph of the same subsection, the second sentence originally reading “Mice lacking functional *C4* exhibited greater overlap between RGC inputs from the two eyes ($P < 0.001$) than wild-type littermate controls,” the text “wild-type littermate controls” should have read “wild-type controls.”

In the Fig. 7a caption originally reading “**a**, Representative confocal images of immunohistochemistry for C3 in the P5 dLGN showed reduced C3 deposition in the dLGN of *C4*^{−/−} mice compared to wild-type (WT) littermates,” the text “wild-type (WT) littermates” should instead have read “age-matched wild-type (WT) mice.” In the Fig. 7c caption originally reading “**c**, Co-localization analysis revealed a reduction in the fraction of VGLUT2⁺ puncta that were C3⁺ in *C4*-deficient mice relative to their WT littermates,” the text “their WT littermates” should have read “age-matched WT mice.” In the third sentence of the Fig. 7d caption originally reading “The overlapping area is significantly increased in *C4*^{−/−} mice ($n = 6$ mice per group...” the text “ $n = 6$ mice per group” should instead have read “ $n = 17$ animals total; wild-type $n = 5$, het

$n = 7$, KO $n = 5$.” In the sixth sentence of the Fig. 7d caption, in the text originally reading “Compared to their wild-type littermates, *C4*-deficient mice exhibited lower R value variance, indicating defects in synaptic refinement ($n = 6$ mice per group,” the text “their wild-type littermates” should have read “age-matched WT mice,” while “ $n = 6$ mice per group” should have read “ $n = 17$ animals total; wild-type $n = 5$, het $n = 7$, KO $n = 5$.”

In Extended Data Fig. 10 with the title originally reading “Mouse *C4* genes and additional analyses of the dLGN eye segregation phenotype in *C4* mutant mice and wild-type and heterozygous littermate controls” the latter portion should have read “... *C4* mutant, wild-type and heterozygous mice.” In the Extended Data Fig. 10e caption appearing originally as “e, Quantification of the percentage of total dLGN area receiving both contralateral and ipsilateral projections shows a significant increase in *C4*^{−/−} compared to wild-type littermates (ANOVA, $n = 5$ mice per group,” the text “wild-type littermates” should have read “age-matched wild-type mice,” while “ $n = 5$ mice per group” should have read “ $n = 17$ mice; wild-type $n = 5$, het $n = 7$, KO $n = 5$.” Similarly, in the Extended Data Fig. 7g caption originally reading “g, Quantification of dLGN area receiving ipsilateral innervation showed a significant increase in ipsilateral territory in the *C4*^{−/−} mice compared to wild-type littermates (AVOVA, $n = 5$ mice per group,” the text “wild-type littermates” should have read “wild-type mice” and “ $n = 5$ mice per group” should have read “ $n = 17$ mice total; wild-type $n = 5$, het $n = 7$, KO $n = 5$.”

Also, in the original Fig. 7d, which showed example images from individual animals, an image of a wild-type animal from a different experiment was utilized in the lower left panel. That image is consistent with the retinogeniculate refinement pattern for wild-type mice in our experiments, but it had been prepared for and used in a previous publication¹ by team members and was not attributed to that publication or experiment here. We have replaced it with an image from the experiment in this study as Fig. 1.

Fig. 1: Replacement panel of original Fig. 7d



None of these corrections affect the results and conclusions of the original paper.

References

1. 1.

Bialas, A. R. & Stevens, B. TGF- β signaling regulates neuronal C1q expression and developmental synaptic refinement. *Nat. Neurosci.* **16**, 1773–1782 (2013); retracted <https://doi.org/10.1038/s41593-021-00877-7> (2021).

Author information

Affiliations

1. Department of Genetics, Harvard Medical School, Boston, Massachusetts, 02115, USA

Aswin Sekar, Heather de Rivera, Avery Davis, Nolan Kamitaki, Katherine Tooley, Matthew Baum, Vanessa Van Doren, Giulio Genovese, Robert E. Handsaker & Steven A. McCarroll

2. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, 02142, USA

Aswin Sekar, Heather de Rivera, Avery Davis, Nolan Kamitaki, Katherine Tooley, Matthew Baum, Giulio Genovese, Samuel A. Rose, Robert E. Handsaker, Mark J. Daly, Beth Stevens & Steven A. McCarroll

3. MD-PhD Program, Harvard Medical School, Boston, Massachusetts, 02115, USA

Aswin Sekar & Matthew Baum

4. Department of Neurology, F.M. Kirby Neurobiology Center, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, 02115, USA

Allison R. Bialas, Timothy R. Hammond, Matthew Baum & Beth Stevens

5. Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, Massachusetts, 02115, USA

Allison R. Bialas, Jessy Presumey & Michael C. Carroll

6. Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, 02114, USA

Mark J. Daly

Consortia

Schizophrenia Working Group of the Psychiatric Genomics Consortium

Corresponding author

Correspondence to [Steven A. McCarroll](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Sekar, A., Bialas, A.R., de Rivera, H. *et al.* Author Correction: Schizophrenia risk from complex variation of complement component 4. *Nature* **601**, E4–E5 (2022). <https://doi.org/10.1038/s41586-021-04202-x>

- Published: 15 December 2021
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04202-x>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04202-x>

| [Section menu](#) | [Main menu](#) |

- Publisher Correction
- [Published: 17 December 2021](#)

Publisher Correction: Burden and characteristics of COVID-19 in the United States during 2020

- [Sen Pei](#) [ORCID: orcid.org/0000-0002-7072-2995¹](#),
- [Teresa K. Yamana¹](#),
- [Sasikiran Kandula¹](#),
- [Marta Galanti](#) [ORCID: orcid.org/0000-0002-9060-1250¹](#) &
- [Jeffrey Shaman](#) [ORCID: orcid.org/0000-0002-7216-7809¹](#)

[Nature](#) volume 601, page E6 (2022)

- 1134 Accesses
- 1 Altmetric
- [Metrics details](#)

The [Original Article](#) was published on 26 August 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03914-4>
Published online 26 August 2021

The correct name of the first author is Sen Pei, as it now appears. In the first online version of the article, the author's name was shown in the correct order. At the proof stage, however, the production team reversed the order.

Author information

Affiliations

1. Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA

Sen Pei, Teresa K. Yamana, Sasikiran Kandula, Marta Galanti & Jeffrey Shaman

Corresponding authors

Correspondence to [Sen Pei](#) or [Jeffrey Shaman](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Pei, S., Yamana, T.K., Kandula, S. *et al.* Publisher Correction: Burden and characteristics of COVID-19 in the United States during 2020. *Nature* **601**, E6 (2022). <https://doi.org/10.1038/s41586-021-04172-0>

- Published: 17 December 2021
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04172-0>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04172-0>

| [Section menu](#) | [Main menu](#) |

- Publisher Correction
- [Published: 17 December 2021](#)

Publisher Correction: Disease variant prediction with deep generative models of evolutionary data

- [Jonathan Frazer](#)¹ na1,
- [Pascal Notin](#)² na1,
- [Mafalda Dias](#)¹ na1,
- [Aidan Gomez](#)²,
- [Joseph K. Min](#) [ORCID: orcid.org/0000-0002-2781-7390](#)¹,
- [Kelly Brock](#)¹,
- [Yarin Gal](#) [ORCID: orcid.org/0000-0002-2733-2078](#)² &
- [Debora S. Marks](#) [ORCID: orcid.org/0000-0001-9388-2281](#)^{1,3}

Nature volume **601**, page E7 (2022)

- 1331 Accesses
- 1 Altmetric
- [Metrics details](#)

Subjects

- [Computational models](#)
- [Disease genetics](#)
- [Genetic variation](#)

- [Genetics research](#)
- [Machine learning](#)

The [Original Article](#) was published on 27 October 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-04043-8> published online 27 October 2021

In the version of this article initially published online, there was an error in Fig. 1. Specifically, in the “Evolutionary index” box, the approximate to symbol was missing in the formula now reading:

$$\$ \$ \{E\}_{v} \sim -\log \frac{P(x)_v}{P(x)_{\text{WT}}}$$

This error has been corrected in the online version of the Article.

Author information

Author notes

1. These authors contributed equally: Jonathan Frazer, Pascal Notin, Mafalda Dias

Affiliations

1. Marks Group, Department of Systems Biology, Harvard Medical School, Boston, MA, USA

Jonathan Frazer, Mafalda Dias, Joseph K. Min, Kelly Brock & Debora S. Marks

2. OATML Group, Department of Computer Science, University of Oxford, Oxford, UK

Pascal Notin, Aidan Gomez & Yarin Gal

3. Broad Institute of Harvard and MIT, Cambridge, MA, USA

Debora S. Marks

Corresponding authors

Correspondence to [Yarin Gal](#) or [Debora S. Marks](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Frazer, J., Notin, P., Dias, M. *et al.* Publisher Correction: Disease variant prediction with deep generative models of evolutionary data. *Nature* **601**, E7 (2022). <https://doi.org/10.1038/s41586-021-04207-6>

- Published: 17 December 2021
- Issue Date: 13 January 2022
- DOI: <https://doi.org/10.1038/s41586-021-04207-6>

Share this article

Anyone you share the following link with will be able to read this content:

[Get shareable link](#)

Sorry, a shareable link is not currently available for this article.

[Copy to clipboard](#)

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04207-6>

| [Section menu](#) | [Main menu](#) |