

nature

BUILT FOR SPEED

Unique motion
of bristled wings
powers tiny
beetle's rapid
flight



Coronavirus

Lessons from Omicron
on the mechanics of
long-term immunity

Climate of change

Survey highlights
gender disparities
at the IPCC

Degrees of separation

Sulfidation of metals
offers practical route
to element extraction

Nature.2022.02.05

[Sat, 05 Feb 2022]

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

This Week

- **[The WHO deserves more money for its core mission — and more respect](#)** [01 February 2022]
Editorial • After decades of neglect, the World Health Organization needs to be properly funded. The United States is — wrongly — opposing a new funding plan. Other nations shouldn't wait to adopt it.
- **[UK ‘DARPA’ should let the sunshine in](#)** [02 February 2022]
Editorial • There's too much secrecy behind ARIA, the United Kingdom's ambitious planned agency for high-risk, high-reward research funding.
- **[COVID’s lesson for governments? Don’t cherry-pick advice, synthesize it](#)** [01 February 2022]
World View • Too many national leaders get good guidance yet make poor decisions.
- **[No bursting for these record-breaking bubbles](#)** [24 January 2022]
Research Highlight • By tinkering with the formula for soap bubbles, scientists create an orb that lasts more than a year.
- **[Vesuvius is off the hook for ancient Arctic ashfall](#)** [27 January 2022]
Research Highlight • Volcanic debris in a Greenland ice-core layer probably came from an Alaskan volcano instead.
- **[Hippos know strangers’ voices — and make a filthy reply](#)** [24 January 2022]
Research Highlight • The mammoth mammals can use each other's ‘wheeze-honks’ to distinguish between neighbours and strangers.
- **[Solar power’s need for a carbon-intensive metal is set to soar](#)** [27 January 2022]
Research Highlight • The shift to clean energy is expected to drive the demand for aluminium, which is used in the frames and fittings of solar panels.
- **[The foreign trees that now reign over Asia’s jungles](#)** [27 January 2022]
Research Highlight • The monarchs of southeast Asian rainforests are ancient immigrants from Africa, by way of India.

- **As lifespans grow, work time while healthy lags** [26 January 2022]

Research Highlight • The gap between average life expectancy and the number of years people are healthy and on the job is widening, with potentially grave consequences for pensioners.

- **Early Arabian capital survives vandals — and reveals a quest for the ultimate camel** [26 January 2022]

Research Highlight • Militants who occupied the centuries-old city of Hatra left behind artwork depicting hybrid camels.

- EDITORIAL
- 01 February 2022

The WHO deserves more money for its core mission — and more respect

After decades of neglect, the World Health Organization needs to be properly funded. The United States is — wrongly — opposing a new funding plan. Other nations shouldn't wait to adopt it.



The WHO's request for more core funding must not be met at the expense of work funded through voluntary contributions. Credit: Junior D. Kannah/AFP/Getty

It's a little-known fact that the World Health Organization (WHO) lives a hand-to-mouth existence. Governments contribute most of its annual US\$6-

billion budget, but the WHO is allowed to spend less than 20% of this on its core mission, which is to support public health in the poorest countries and respond to emergencies around the world. Much of the rest consists of voluntary donations earmarked for projects specified by donors, such as eliminating polio or treating HIV/AIDS.

The governments of most of the WHO's member states are now rightly saying that the agency needs more-predictable funding, as well as the flexibility to decide when and how to spend a greater proportion of its budget — during a pandemic, for example. They should consider how to move ahead with implementing this plan despite the fact that a small number of countries are opposing it.

Since the WHO was founded in 1948, its total spending power has increased nearly sixfold, when adjusted for inflation. But the majority of this money has been provided through voluntary contributions. By contrast, core funding — which comes from assessed contributions, based on a country's wealth — has decreased in real terms. As a result, the WHO is struggling to achieve a key goal [to provide universal health coverage for the most vulnerable populations](#).

For decades, the WHO's leadership has tried to persuade its donors to rebalance its funding. Last week, it tried again. At least 120 countries are backing a proposal for half of all funding for the WHO to comprise assessed contributions. This would bring its funding mechanism closer to that of many other United Nations agencies. But eight countries — including, most notably, the United States, but also Brazil, Japan and Poland — are opposing the move.

The United States wants to see how the WHO can increase efficiency before it commits more funding. In 2019 (the last year for which data are available), the United States paid \$419 million to the WHO, of which assessed contributions made up \$119 million. One way to increase efficiency would be to relieve the agency of the need to devote so many of its resources to the priorities of individual governments.

The US opposition to rebalancing the budget is a surprise, considering President Joe Biden's full-throated support of the WHO in 2020, when then-

president Donald Trump withdrew the United States and its funding from the WHO, saying the agency had become too close to China. At the time, Trump was also considering plans to create a global pandemic-response agency separate from the WHO. Biden reversed the withdrawal on taking office in January 2021, but his administration has restored only about three-quarters of US funding. And earlier this month, reports emerged that the United States is continuing to push for a pandemic-response agency — to be funded by the G20 group of the world's largest economies and administered by the World Bank in Washington DC. If brought into being, this would compete with the WHO. It would also clash with nascent plans, spearheaded by the European Union and backed by the WHO, to create a pandemic-response pact for countries to sign up to. Despite the change in US leadership, the WHO is still caught up in great-power politics. For the world's health agency to be politicized at any time is unwise; for it to happen during a pandemic is dangerous.

Countries will need to decide on the proposal to increase the WHO's assessed contributions by May, when WHO member states meet in Geneva, Switzerland, for the annual World Health Assembly. Realistically, the United States is unlikely to change its position so soon, if at all. But, while it makes up its mind, the many countries that are supporting the proposal should consider how they can press ahead and increase the agency's core funding. For example, they could increase voluntary contributions that come with no restrictions — known as 'core voluntary contributions'.

Countries are being asked to increase their assessed contributions from \$489 million in 2022–23 to \$1.1 billion in 2028–29. When shared between 194 member states, this is not a large sum of money in the context of health care. But, overall, the increase to the WHO's assessed contributions amounts to a doubling. Such a multiple will be tricky to push through national parliaments. At a time of pressure on government budgets, lawmakers in many countries will question why such an increase is needed.

Funding for the WHO tends to come out of countries' international-aid budgets, and these budgets are under strain. The pandemic, along with newer crises, such as the collapse of governance in Afghanistan, has meant that there is huge need for international humanitarian assistance, which comes out of aid budgets. At the same time, citizens in some nations are

questioning their countries' aid spending, urging governments to spend more money to tackle the health and economic consequences of the pandemic at home.

One potential solution is for countries' WHO contributions to come out of their general health spending (in effect, removing it from the aid budget). The logic is that supporting the WHO should not be seen as aid spending, but as spending to protect global public health — including anticipating and preventing or mitigating the next pandemic. Moreover, the WHO's share of a nation's health budget will be a much smaller fraction than its share of aid spending, because health budgets tend to be larger. Germany has already made the shift. In 2021, the country spent some €24 billion (US\$27 billion) on health care. By contrast, its assessed contribution to the WHO was just \$29 million.

One of the biggest risks of the WHO plan is that lawmakers might look to make savings elsewhere, by making cuts to spending on diseases such as HIV/AIDS or malaria, or to childhood immunization campaigns — all of which constitute a significant share of voluntary contributions. It's crucial that this does not happen.

More funding and a rebalancing of the WHO budget is ultimately not just about money. As important as the cash is, what matters more is the signal increased funding would send that nations value the WHO's core missions — to be led by scientific evidence and to ensure that people everywhere have the opportunity to live a healthy life.

Nature **602**, 7–8 (2022)

doi: <https://doi.org/10.1038/d41586-022-00227-y>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00227-y>.

- EDITORIAL
- 02 February 2022

UK ‘DARPA’ should let the sunshine in

There’s too much secrecy behind ARIA, the United Kingdom’s ambitious planned agency for high-risk, high-reward research funding.



No taxpayer-funded institution should be exempt from freedom of information laws. Credit: Doug Jones/Portland Press Herald/Getty

Can there be any justification for a science-funding agency that makes its decisions in secret — especially if that agency is funded by the taxpayer? The UK government seems to think so. It will shortly launch a new ‘high-risk, high-reward’ funding agency, known as ARIA (Advanced Research

and Invention Agency). ARIA, the UK equivalent of the US Defense Advanced Research Projects Agency (DARPA), will distribute £800 million (US\$1 billion) to researchers in its first 4 years — considerably more than the £110 million a year available to all of the nation’s arts and humanities researchers. But few outside the country’s government, Parliament and audit watchdog will be able to scrutinize its decisions. That’s because ARIA will be exempt from freedom of information (FoI) laws, legislation that allows the public to access information collected and held by the government.

Funds for high-risk science are a good thing, but keeping decisions on that science under wraps isn’t — and researchers and some lawmakers are rightly concerned. They’ve been campaigning for ARIA to be subject to the usual rules of public access to official data. But the government is refusing. Earlier this month, the upper house of the UK Parliament failed to get this aspect of the draft law changed.

A government spokesperson told *Nature* that information on “delivery partners” (grantees) will be published, but it will be up to ARIA’s leadership to decide what and how much. The government has [just announced who will helm the initiative](#). Peter Highnam, deputy director of the DARPA, will join as ARIA’s chief executive in May. He should, at the very least, ensure that ARIA’s grants data are integrated into the United Kingdom’s grants data portal, called [Gateway to Research](#).

ARIA was created to combat a perception that funding agencies are stuck in a ‘bog of bureaucracy’. The thinking is that many rounds of applications and peer review quash creativity: researchers think reviewers will reject applications that are outside the consensus of existing scientific or technological opinion and are therefore reluctant to take risks. By contrast, ARIA’s architects say the agency will not “shy away from failure”.

But ministers have failed to explain precisely why ARIA needs to be exempt from FoI laws. The government has said that removing the requirement to deal with FoI requests will “reduce administrative burden from ARIA’s staff” so they can “find and fund the most cutting edge research”. But this is a nonsensical argument. UK research-funding agencies of ARIA’s size receive, on average, just 50 FoI requests per year — and it doesn’t prevent

them funding world-class science. It's also nonsense to suggest that transparency is incompatible with excellence and creativity.

A defence- and security-policy paper that the government published last March, called *Global Britain in a Competitive Age*, provides another explanation for the desire to keep ARIA behind closed doors. Science and technology are referred to throughout the document. The government sees both as key to countering external threats, from countries that the United Kingdom regards as hostile, for example, or involved in terrorism or organized crime. ARIA is mentioned as one component among wider reforms to the funding and governance of science. These reforms include comparatively larger funding increases going to science that will be spent directly by government departments, and new high-level committees in which government officials will be more involved in advising scientists on research priorities.

This is a shift from the approach of previous (Conservative and Labour) governments, which saw science as much as a tool of diplomacy to address global challenges as an instrument to boost competitiveness. For example, the Global Challenges Research Fund was set up to enhance international collaborations and to meet the United Nations Sustainable Development Goals. There's a perception among researchers that this ethos has been replaced with one that aims to protect the nation from threats and to project the United Kingdom's hard power, says James Wilsdon, who researches science and government at the University of Sheffield, UK.

But a security-focused approach for UK science does not justify exemptions from FoI laws. DARPA is subject to FoI laws. Between 2011 and 2019, it handled around 50 requests each year. If ARIA is freed from scrutiny, it will join a very small and exclusive club of publicly funded institutions, alongside the British royal family and the country's various security and intelligence agencies. That makes no sense.

At [an event in London last week](#), UK chief scientific adviser Patrick Vallance rightly said more science needs to be open — including the science used in defence and security. An agency that recognizes the value of failure to eventual success needs to be open, so that researchers can build and learn from what works and what doesn't.

ARIA will give tremendous power to those who sit inside its black box, making decisions about whom and what to fund. If ARIA succeeds in creating technologies that shape society, it must not do so behind closed doors. Those who pay for it should have the ability to scrutinize its decisions.

Nature **602**, 8 (2022)

doi: <https://doi.org/10.1038/d41586-022-00226-z>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00226-z>

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

- WORLD VIEW
- 01 February 2022

COVID's lesson for governments? Don't cherry-pick advice, synthesize it



Too many national leaders get good guidance yet make poor decisions.

- [Geoff Mulgan](#) 9

Handling complex scientific issues in government is never easy — especially during a crisis, when uncertainty is high, stakes are huge and information is changing fast. But for some of the nations that have fared the worst in the COVID-19 pandemic, there's a striking imbalance between the scientific advice available and the capacity to make sense of it. Some advice is ignored because it's politically infeasible or unpragmatic. Nonetheless, much good scientific input has fallen aside because there's no means to pick it up.

Part of the problem has been a failure of synthesis — the ability to combine insights and transcend disciplinary boundaries. Creating better syntheses should be a governmental priority as the crisis moves into a new phase.

Both the theory and practice of synthesis remain inadequate. I saw this when I ran the UK Government's Strategy Unit in the early 2000s: I developed policy for everything from energy and carbon reduction to health care. I helped to set up a similar unit in Australia and have advised dozens of governments, from Canada to Bangladesh, France to Finland.

Over the past year, I've helped to run the International Public Policy Observatory (IPPO), based at University College London, with partners such as the International Network for Government Science Advice. IPPO organizes evidence syntheses on issues such as teen mental health, homelessness and the unplanned push towards online learning. Our techniques include roundtables, systematic reviews and global evidence scans.

Input from evidence synthesis is crucial for policymaking. But the capacity of governments to absorb such evidence is limited, and syntheses for decisions must go much further in terms of transparently incorporating assessments of political or practical feasibility, implementation, benefits and cost, among many other factors. The gap between input and absorption is glaring.

I've addressed teams in the UK prime minister's office, the European Commission and the German Chancellery about this issue. In responding to the pandemic, some countries (including France and the United Kingdom) have tried to look at epidemiological models alongside economic ones, but none has modelled the social or psychological effects of different policy choices, and none would claim to have achieved a truly synthetic approach.

There are dozens of good examples of holistic thinking and action: programmes to improve public health in Finland, cut UK street homelessness, reduce poverty in China. But for many governments, the capacity to see things in the round has waned over the past decade. The financial crisis of 2007 and then populism both shortened governments' time horizons for planning and policy in the United States and Europe.

Governments are now even less capable of using high-quality advice, assuming they obtain it. Ministries — such as those for agriculture and education — often have plenty of experts siloed within their own specialties. But teams around executives struggle to weave advice and evidence together. They tend to be small and consumed by ‘firefighting’. Politicians are too busy and distracted to do the job of synthesis, and civil servants are usually more comfortable with law and economics than with science or statistics, or the practicalities of implementation.

The Chinese mainland does have some central capacity to shape policy (as well as many officials with backgrounds in science and engineering), which might help it to navigate out of its zero-COVID stance. Taiwan, Singapore and South Korea — which had dramatically lower death rates than the United States and United Kingdom — also did well in leveraging data from tests, mobile phones and much more to guide policy.

The worst governments rely on intuition. But even the best resort to simple heuristics — for example, that it’s best to act fast, or that prioritizing health is also good for the economy. This was certainly true in 2020 and 2021. But that might change with higher vaccination and immunity rates.

What would it mean to transcend simple heuristics and achieve a truly synthetic approach? It would involve mapping and ranking relevant factors (from potential impacts on hospital capacity to the long-run effects of isolation); using formal and informal models to capture feedbacks, trade-offs and synergies; and more creative work to shape options.

Usually, such work is best done by teams that encompass breadth and depth, disparate disciplines, diverse perspectives and both officials and outsiders. Good examples include Singapore’s Strategy Group (and Centre for Strategic Futures), which helps the country to execute sophisticated plans on anything from cybercrime to climate resilience. But most big countries, despite having large bureaucracies, lack comparable teams.

Establishing such teams should be a priority. So should supporting a better science of synthesis. Universities’ interdisciplinary projects often stop short of true synthesis and recommending specific strategies (such as on carbon reduction) or making sharp judgements about priorities. Despite the

contributions of behavioural science, complexity theories, computer science and social sciences, understanding of how whole systems could behave better is rudimentary.

I hope one legacy of the pandemic will be a concerted effort to improve both the theory and practice of sophisticated synthesis — to help us cope with the many crises ahead, from transitioning energy and transport to avert climate change to reducing inequality and rebuilding public trust in science.

Nature **602**, 9 (2022)

doi: <https://doi.org/10.1038/d41586-022-00212-5>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00212-5>

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

- RESEARCH HIGHLIGHT
- 24 January 2022

No bursting for these record-breaking bubbles

By tinkering with the formula for soap bubbles, scientists create an orb that lasts more than a year.

 A liquid water and glycerol bubble at 33 days old.

A 33-day-old bubble owes its longevity, in part, to a coating of plastic microparticles. Credit: Roux *et al. Phys. Rev. Fluids/APS (2022)*

Soap bubbles with a few extra ingredients can retain their structure for up to 465 days — more than 200,000 times longer than ordinary bubbles¹.

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 **602**, 10 (2022)

doi: <https://doi.org/10.1038/d41586-022-00172-w>

References

1. 1.

Roux, A., Duchesne, A. & Baudoin, M. *Phys. Rev. Fluid*
<https://doi.org/10.1103/PhysRevFluids.7.L011601> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00172-w>

- RESEARCH HIGHLIGHT
- 27 January 2022

Vesuvius is off the hook for ancient Arctic ashfall

Volcanic debris in a Greenland ice-core layer probably came from an Alaskan volcano instead.

 Satellite view of Aniakchak (center), an active volcano located on the Alaska Peninsula 670 kilometers southwest of Anchorage.

A volcano similar to Aniakchak Peak (pictured), located in Alaska, might be the true source of volcanic debris found in a Greenland ice core. Credit: Planet Observer/Universal Images Group via Getty

An ice core from Greenland that was thought to contain traces of Mount Vesuvius's famous eruption probably doesn't — a finding that could tweak the timelines of key events in the history of Earth¹.

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 **602**, 10 (2022)

doi: [*https://doi.org/10.1038/d41586-022-00174-8*](https://doi.org/10.1038/d41586-022-00174-8)

References

1. 1.

Plunkett, G. *et al.* *Clim. Past* **18**, 45–65 (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00174-8>

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

- RESEARCH HIGHLIGHT
- 24 January 2022

Hippos know strangers' voices — and make a filthy reply

The mammoth mammals can use each other's 'wheeze-honks' to distinguish between neighbours and strangers.



That familiar voice: hippos can tell members of their own groups from strangers by listening to their vocalizations. Credit: Nicolas Mathevon

Hippos can recognize each other's honking voices — and aggressively spray dung when they hear from hippo groups they dislike¹.

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 **602**, 10 (2022)

doi: <https://doi.org/10.1038/d41586-022-00173-9>

References

1. 1.

Thévenet, J., Grimault, N., Fonseca, P. & Mathevon, N. *Curr. Biol.* **32**, R70–R71 (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00173-9>

- RESEARCH HIGHLIGHT
- 27 January 2022

Solar power's need for a carbon-intensive metal is set to soar

The shift to clean energy is expected to drive the demand for aluminium, which is used in the frames and fittings of solar panels.

 A worker in protective gear removes bars of aluminium from a casting machine at Rusal's Bratsk Aluminium Smelter.

A worker at an aluminium smelter in Bratsk, Russia. Many aluminium plants consume vast amounts of coal-fired energy. Credit: Alexander Ryumin/TASS via Getty

A paradox plagues solar energy: it is key to slowing climate change, but solar installations require aluminium, the mining and processing of which releases vast amounts of greenhouse gases.

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 **602**, 10 (2022)

doi: <https://doi.org/10.1038/d41586-022-00176-6>

References

1. 1.

Lennon, A., Lunardi, M., Hallam, B. & Dias, P. R. *Nature Sustain.* <https://doi.org/10.1038/s41893-021-00838-9> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00176-6>

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

- RESEARCH HIGHLIGHT
- 27 January 2022

The foreign trees that now reign over Asia's jungles

The monarchs of southeast Asian rainforests are ancient immigrants from Africa, by way of India.

 *Shorea leprosula*, one of the most widespread dipterocarp tree species in Southeast Asia, in a nature reserve in Singapore.

A specimen of *Shorea leprosula*, one of the most widespread of the tree species known as dipterocarps in southeast Asia, towers above the ground in a nature reserve in Singapore. Credit: Jun Ying Lim

An analysis of fossilized pollen has revealed that some of the world's tallest tree species, which now dominate tropical forests in southeast Asia, originated in Africa around 100 million years ago¹.

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 **602**, 11 (2022)

doi: [*https://doi.org/10.1038/d41586-022-00182-8*](https://doi.org/10.1038/d41586-022-00182-8)

References

1. 1.

Bansal, M. *et al.* *Science* **375**, 455–460 (2022).

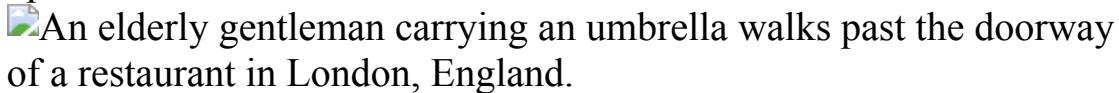
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00182-8>

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

- RESEARCH HIGHLIGHT
- 26 January 2022

As lifespans grow, work time while healthy lags

The gap between average life expectancy and the number of years people are healthy and on the job is widening, with potentially grave consequences for pensioners.



A man walks through London's financial district. Between 2015 and 2035, men in England will gain more than 3 years of life expectancy but less than half a year of healthy working life. Credit: Richard Baker/In Pictures via Getty

Several European countries plan to raise the pension age as a result of anticipated increases in life expectancy. But research in England predicts that in coming decades, people will gain years of life much faster than they will gain years of life when they're healthy and in work¹.

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 **602**, 11 (2022)

doi: <https://doi.org/10.1038/d41586-022-00136-0>

References

1. 1.

Lynch, M., Bucknall, M., Jagger, C. & Wilkie, R. *Nature Aging* **2**, 13–18 (2022).

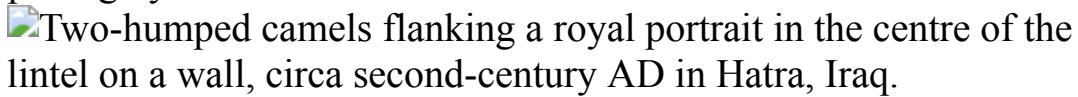
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00136-0>

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

- RESEARCH HIGHLIGHT
- 26 January 2022

Early Arabian capital survives vandals — and reveals a quest for the ultimate camel

Militants who occupied the centuries-old city of Hatra left behind artwork depicting hybrid camels.



Some of the camels carved in stone on an ancient temple have humps that suggest they are Bactrian camel-dromedary hybrids. Credit: Aliph-ISMEO project at Hatra

An ancient city damaged by the Islamist terrorist group ISIS features stone carvings of some unexpected beasts: early camel–dromedary hybrids¹.

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 **602**, 11 (2022)

doi: <https://doi.org/10.1038/d41586-022-00175-7>

References

1. 1.

Vidale, M., Berlioz, S. & Mohammed, R. *Antiquity*
<https://doi.org/10.15184/aqy.2021.169> (2022).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00175-7>

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

News in Focus

- **[Long COVID, open-access barriers and tobacco money](#)** [02 February 2022]
News Round-Up • The latest science news, in brief.
- **[China's zero-COVID strategy: what happens next?](#)** [27 January 2022]
News • With Omicron or Delta outbreaks already in multiple provinces, scientists say next week's Winter Olympics will present a major test of China's zero-tolerance approach.
- **[Three, four or more: what's the magic number for booster shots?](#)** [28 January 2022]
News • COVID vaccine boosters are proving a useful tool against Omicron, but scientists say that endless boosting might not be a practical or sustainable strategy.
- **[Big dog, little dog: mutation explains range of canine sizes](#)**
[27 January 2022]
News • The genetic variant probably comes from ancient wolves.
- **[Deltacron: the story of the variant that wasn't](#)** [21 January 2022]
News • News of a 'super variant' combining Delta and Omicron spread rapidly last week, but researchers say it never existed and the sequences might have resulted from contamination.
- **[Will Omicron end the pandemic? Here's what experts say](#)**
[31 January 2022]
News • The variant's rapid spread, different vaccine strategies and varying levels of immunity worldwide make the pandemic's future difficult to model.
- **[What the Omicron wave is revealing about human immunity](#)** [02 February 2022]
News Feature • Immunologists have raced to work out how to protect against multiple variants of SARS-CoV-2. Their research has yielded a wealth of insights and a few surprises.
- **[Where did Omicron come from? Three key theories](#)** [28 January 2022]
News Feature • The highly transmissible variant emerged with a host of unusual mutations. Now scientists are trying to work out how it evolved.

- NEWS ROUND-UP
- 02 February 2022

Long COVID, open-access barriers and tobacco money

The latest science news, in brief.



A medical worker prepares a dose of the Pfizer–BioNTech COVID-19 vaccine, which has been widely administered in Israel. Credit: Luka Dakskobler/SOPA Images/LightRocket/Getty

Long-COVID symptoms less likely in vaccinated people, Israeli data say

Data from people infected with SARS-CoV-2 early in the pandemic add to growing evidence suggesting that vaccination can help to reduce the risk of

long COVID ([P. Kuodi et al. Preprint at medRxiv https://doi.org/10.1101/2022.07.01.50029](https://doi.org/10.1101/2022.07.01.50029); 2022).

Scientists in Israel report that people who have both been infected with SARS-CoV-2 and received doses of the Pfizer–BioNTech vaccine are [much less likely to report any of a range of common long-COVID symptoms](#) than are people who were unvaccinated when infected. In fact, vaccinated people are no more likely to report symptoms than people who have never caught SARS-CoV-2. The study has not yet been peer reviewed.

“Here is another reason to get vaccinated, if you needed one,” says co-author Michael Edelstein, an epidemiologist at Bar-Ilan University in Safed, Israel.

People with long COVID have symptoms, such as fatigue, long after SARS-CoV-2 infection.

In theory, vaccination could protect against the condition by shortening the virus’s free rein in the body during breakthrough infections. But the few studies examining whether vaccines protect against long COVID have had mixed results, says Akiko Iwasaki, a viral immunologist at the Yale School of Medicine in New Haven, Connecticut.

To learn more, between July and November 2021, Edelstein and his co-authors asked some 3,000 people whether they were experiencing the most common long COVID symptoms. All had been tested for SARS-CoV-2 between March 2020 and the study period.

The researchers compared the prevalence of each symptom with self-reported vaccination status and found that fully vaccinated participants who had also had COVID-19 were 54% less likely to report headaches, 64% less likely to report fatigue and 68% less likely to report muscle pain than were their unvaccinated counterparts. Edelstein says his team’s results echo those of other research, including a UK-based study ([M. Antonelli et al. Lancet Infect. Dis. 22, 43–55; 2022](#)) from last September that found that vaccination halved the risk of long COVID.

Claire Steves, a geriatrician at King’s College London who led the UK study, agrees that the Israeli data support earlier findings. “It’s really good to

see different study designs correlating, with the same results,” she says.

High costs of open-access publishing deter authors

The fees that journals charge to publish researchers’ work open access are a barrier for authors in low- and middle-income countries, an analysis finds.

More academic journals have been making their articles free to read, driven in part by open-access requirements from research funders. Although this has made more of the scholarly literature accessible, many scientists have noted the high article-processing charges (APCs) required to publish open-access papers in some titles.

Researchers examined more than 37,000 articles published by Elsevier in 38 ‘hybrid’ journals, which contain both paywalled and freely accessible articles, and their ‘mirror’ titles — fully open-access versions of those journals ([A. C. Smith et al. *Quant. Sci. Stud.* \[https://doi.org/gn38g5; 2021\]\(https://doi.org/gn38g5\)](https://doi.org/gn38g5)). The APCs for the mirror journals cost a median of US\$2,600. Around 80% of articles in mirror journals had lead authors from high-income countries, and none had authors from low-income countries.

“When we see results like this, it just makes it crystal clear that there is a very strong financial barrier for publishing when journals charge APCs,” says Rafael Zenni, an ecologist at the Federal University of Lavras in Brazil.



Philip Morris International owns various cigarette brands, including Marlboro. Credit: Daniel Acker/Bloomberg via Getty

Publishing ban for tobacco-funded researchers

A group of international respiratory societies has [banned researchers associated with tobacco companies from publishing papers in their journals](#), after the controversial acquisition by a tobacco giant of a health-care firm that makes inhalers.

Last year, the Swiss-American tobacco company Philip Morris International, which owns some of the world's popular cigarette brands, took over UK-based pharmaceutical firm Vectura. In response, six health bodies, including the American Thoracic Society, the European Respiratory Society and the International Union Against Tuberculosis and Lung Diseases, have issued a joint statement describing the move as "highly unethical and inappropriate".

Respiratory societies have banned researchers directly funded by tobacco companies for more than a decade. Now, the organizations have extended

that ban to researchers working for firms acquired by tobacco companies.

The statement says that employees of such companies will not be allowed to publish in the societies' journals or present at meetings, and recommends that people do not use medical products newly developed by tobacco-owned firms.

Nature **602**, 13 (2022)

doi: <https://doi.org/10.1038/d41586-022-00213-4>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00213-4>

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

- NEWS
- 27 January 2022

China's zero-COVID strategy: what happens next?

With Omicron or Delta outbreaks already in multiple provinces, scientists say next week's Winter Olympics will present a major test of China's zero-tolerance approach.

- [Smriti Mallapaty](#)



Health-care workers conduct COVID-19 tests in Zhengzhou, central China.Credit: Ma Jian/VCG via Getty

China's stringent zero-COVID strategy is likely to face its toughest test yet in the next few weeks, as millions of people travel around the country for Chinese New Year, and the Winter Olympics begin in Beijing.

The approach — which was introduced by the central government early in the pandemic and has involved large-scale lockdowns, mass testing and international travel bans — has been under pressure since China's first Omicron cases were reported in mid-December. The highly infectious variant has been detected in at least 14 provinces and cities including Tianjin and Beijing, and scientists fear that fresh outbreaks might occur after next week's events.

"Omicron is going to create more of a challenge and lead to more disruptions than previous variants," says Ben Cowling, an epidemiologist at the University of Hong Kong. Outbreaks might spill into the community and will be difficult to control, he says. "The Olympics is going to be a big test."

Even though Omicron is tough to contain, its increased transmissibility and ability to evade vaccine-derived immunity have hardened support for China's unwavering strategy among some scientists.

Near-impossible to keep out

Researchers say that vaccines based on inactivated-virus technology — such as China's widely used [CoronaVac and Sinopharm vaccines](#) — offer some protection against severe disease with Omicron, but will [prevent few Omicron infections](#). "It is not the right time to reopen," says Chen Tianmu, an epidemiologist at Xiamen University.

But other researchers argue that it will be near-impossible for China to keep the variant out. "You can't stop the wind with your hand," says Rafael Araos, a physician and epidemiologist at the University for Development in Santiago. The costs of shutting borders outweigh the benefits, now that vaccines can reduce hospitalizations and deaths, he says. "It is getting harder and harder to justify the zero-COVID approach."

In the past few months, China has experienced its largest COVID-19 outbreak since April 2020. In late November, daily cases of infections

peaked at 361 — a marginal figure relative to the size of the country's population. Nevertheless, in response, China's government pursued swift and severe measures to get case numbers down.

Millions-strong cities have implemented strict lockdowns and introduced rounds of mass testing. Residents have had to make do with intermittent deliveries of food and medicines. In Xi'an in December, the government even banned all traffic and cancelled flights.

Throughout the pandemic, China's international borders have effectively been closed, preventing almost anyone from getting in or out. That has kept daily cases in the country in the hundreds or fewer, rather than the hundreds of thousands seen in the past few weeks in countries such as the United Kingdom and the United States.



Officials wait to process arrivals for the Olympics at Beijing airport. Credit: Carl Court/Getty

Ineffective inactivated vaccines

The heavy-handed response continues even though China has administered nearly 3 billion doses of vaccines. Some 85% of the population have been fully vaccinated, and a large fraction have received a third dose.

But a highly vaccinated population is unlikely to be a barrier against Omicron's spread. The relative inefficacy of China's vaccines at preventing infections, combined with Omicron's increased transmissibility, will make it harder for China to maintain its zero-COVID approach, says Yanzhong Huang, who studies global health in China at the Council on Foreign Relations in New York City. However, "it's precisely that lack of confidence in their vaccines" that is causing China to stick to the zero-COVID approach, he adds.

Modelling led by Chen suggests that even with 80–90% coverage with its existing vaccines, China could still experience huge numbers of hospitalizations and deaths if it relaxes its zero-tolerance strategy while a variant with similar properties to those of Omicron is spreading.

Fears about this kind of outcome have led China to double-down on its pandemic-response efforts, says Huang. The government is strongly urging people not to travel for Chinese New Year celebrations — which normally see huge numbers travelling across the country to visit family — but it has stopped short of a total ban, so there is likely to be movement anyway.

And international athletes attending the Olympics will be confined to a 'bubble', flying in on chartered flights, travelling from their hotels to sports venues in dedicated vehicles and being subjected to daily testing. Tickets will not be sold for events, and the few spectators who are allowed to attend will be instructed not to shout or cheer. In terms of precautionary measures, it's probably "the most stringent Olympics in history", says Huang.

Planning an exit

Researchers have differing views on when and how China might plan its exit from the zero-COVID strategy.

Many say ramping up booster campaigns is important. "We advise to push forward the boosters to combat the emerging variant," says Pengfei Wang, a

virologist at Fudan University in Shanghai.

Booster-vaccine coverage should be “as high as possible before we reopen the country”, reaching at least 90%, agrees Chen. “We need to build our immunity barrier high.”

Cowling argues that China should time the ramp-up of its booster campaign as close to the nation’s reopening as possible, to account for waning immunity. “Vaccines are not so critical for maintaining zero-COVID but are really critical for an exit from zero-COVID,” he says.

Furthermore, different kinds of vaccine — such as those based on messenger RNA, rather than inactivated-virus vaccines — should be considered for third, booster shots, because this might offer better immunity, says Lu Jiahai, an infectious-diseases epidemiologist at Sun Yat-sen University in Guangzhou.



Few people in Beijing have access to the Winter Olympics 'bubble', to which the athletes are confined. Credit: Kevin Frayer/Getty

Although the country's inactivated-virus vaccines have been the most widely used, China has also approved an adenovirus-vector and a protein-subunit vaccine. Furthermore, it has two mRNA vaccines in early-stage trials, says Lu. Alternatives to inactivated-virus vaccines are more widely available internationally, but China has so far shown a steadfast determination to use only Chinese jabs.

In December 2020, Shanghai-based pharmaceutical company Fosun Pharma and biotechnology company BioNTech, based in Mainz, Germany, announced that they had collaborated to produce an mRNA COVID-19 vaccine and to supply doses to China. However, the vaccine has not yet been approved by regulators. If it is, it would be the first internationally developed vaccine to be approved on the Chinese mainland.

But Shibo Jiang, a virologist at Fudan University, says it's likely that none of the world's available COVID-19 vaccines will offer sufficient protection against emerging variants. Instead, says Jiang, countries need to develop vaccines that can elicit potent neutralizing antibodies against the broader group of related coronaviruses. He adds that people in China are increasingly convinced of the need to invest in such vaccines, especially since the emergence of Omicron.

Some researchers say that outbreaks during the Olympics might start to loosen the authorities' zero-COVID approach, and contribute to a mental shift among the population towards a greater tolerance of outbreaks.

But unless COVID-19 comes under control in the rest of the world, "as long as you believe that even a small opening can be devastating for China, you're not going to expect it to abandon that approach", says Huang. "There are some tough decisions for the leaders to make."

Nature **602**, 15-16 (2022)

doi: <https://doi.org/10.1038/d41586-022-00191-7>

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

- NEWS
- 28 January 2022

Three, four or more: what's the magic number for booster shots?

COVID vaccine boosters are proving a useful tool against Omicron, but scientists say that endless boosting might not be a practical or sustainable strategy.

- [Clare Watson](#) ⁰



Health-care workers receive a fourth dose of the Pfizer–BioNTech vaccine in Santiago, Chile. Credit: Javier Torres/AFP/Getty

Late last year, studies showed that [third shots \(boosters\)](#) of COVID vaccines were effective at providing a little extra protection from infection — particularly in the face of the Omicron variant. Some countries are now offering fourth doses, but scientists say that endless boosting might not be a viable strategy, nor is it how these vaccines were meant to be used.

“We’re in totally uncharted territory for vaccinology,” says Danny Altmann, an immunologist at Imperial College London. “We’ve stumbled into a de facto programme of frequent mRNA boosters as an emergency measure, but this really doesn’t feel like the way to go.”

In early January, Israel began offering fourth doses to older and immunocompromised people and to health-care workers, hoping to shield vulnerable groups from a wave of Omicron infections, says Ran Balicer, a public-health physician at the Clalit Health Institute in Tel Aviv. This week, [preliminary data](#) from Israel revealed that a fourth dose reduces the risk of infection and severe disease.

But researchers are debating whether a third dose will be enough to confer lasting immunity against Omicron and emerging variants in most people — or whether a fourth dose, or even regular boosters, will be needed, as they are for influenza.

Some researchers say that the answer depends on the desired effect — whether boosters are intended to prevent infections and slow transmission of the virus, or whether the goal is to reduce severe disease and keep people out of hospital. Others point to evidence that extra doses could broaden the immune response enough to recognize new variants. Most agree that we need new vaccines that offer wider protection against future variants.

Limitations of boosters

Omicron changed the thinking around boosters, says Alejandro Balazs, an immunologist at the Ragon Institute in Cambridge, Massachusetts. That’s because, faced with the variant, people previously regarded as fully vaccinated now have “an antibody response that is insufficient to prevent infections”, he says.

As Omicron outbreaks have spread, boosters have been used to ramp up levels of neutralizing antibodies, curbing cases and easing strain on hospitals^{1,2}. But the concern is that boosters don't block infections for long.

Data from Israel — collected between June and November last year when Delta was dominant — and detailed online ahead of peer review, indicate that the immunity from a third (mRNA booster) shot wanes within months, mirroring the decline after two doses³.

Real-world data [from the United Kingdom](#), collected in late 2021, suggest that immunity from boosters might decrease even faster against Omicron than against Delta. However, another laboratory study, posted as a preprint which has yet to be peer reviewed, suggests that neutralizing antibodies elicited by a third dose could sustain protection against Omicron infections for up to four months⁴.

Because protection from boosters might be short-lived, rolling out endless doses — potentially at the expense of immunizing unvaccinated people in low-income nations — is not a “viable or reasonable” long-term global strategy, says Kanta Subbarao, a virologist at the Peter Doherty Institute for Infection and Immunity in Melbourne, Australia.

And, in a [statement](#) released on 11 January, the World Health Organization warned that “a vaccination strategy based on repeated booster doses of the original vaccine composition is unlikely to be appropriate or sustainable”.

Repeated booster doses of existing vaccines also probably offer only diminishing returns in terms of protection against future strains, says Miles Davenport, a computational immunologist at the University of New South Wales in Sydney, Australia. New vaccines that target specific variants are likely to be much more effective, he adds.

Whether four shots boost levels of infection-blocking antibodies any higher than a third dose remains to be seen, Davenport says, but that hasn't deterred nations including Chile, Cambodia, Denmark and Sweden from offering fourth doses to specific groups.

The [preliminary data](#) released from Israel this week, on study participants aged over 60, does, however, suggest that a fourth dose, at least four months after a third shot, revives antibody levels, doubles resistance against Omicron infection and triples protection against hospitalizations, compared with only three shots.



A medic prepares a fourth, booster dose of the Pfizer-BioNTech vaccine at a nursing home in Netanya, Israel.Credit: Jack Guez/AFP/Getty

Broad, long-lasting protection

Other studies, which looked at different parts of the body's immune response, suggest that a third shot might already provide long-lasting immunity in most cases. Protection against severe illness seems [more durable](#) and is probably due to memory B cells and T cells, which remain capable of battling Omicron even as antibody defences decline^{5,6}.

Real-world data from the United States, the United Kingdom and Israel show that a third (booster) shot of an mRNA vaccine protects most people

against hospitalization for up to five months against Delta — and for three months or more [against Omicron](#)^{7,8,9}. This more durable immunity “also wanes, but to a lesser extent”, says Balicer, meaning that a third shot might be enough to prevent people getting critically ill.

Work led by Balazs further suggests that a third dose of an mRNA vaccine (which are used largely in the West) not only restores antibody levels but also potentially broadens responses to variants¹⁰. After that booster, “the antibodies actually see Omicron now, where they effectively didn’t see it before”, he says.

“Hopefully, this third shot is enough” for most people to prevent severe disease and offer some protection against infection, Balazs adds. But [some studies suggest](#) that people who are immunized with inactivated-virus vaccines — such as China’s CoronaVac and Sinopharm jabs, which are widely used in middle- and low-income nations — might need two additional doses of an mRNA vaccine to combat Omicron.

Altmann says that, with differing levels of natural immunity from past infections in communities worldwide, and with people having had many combinations of vaccines, “we may need to take a deep breath and re-evaluate which approaches really give the most enduring immunity when overlaid on what we have so far”.

Better solution than endless boosters

Rather than administering endless booster shots, says Balicer, a better way to slow the pandemic would be to develop new vaccines that “have a longer, enduring effect, and that allow adequate protection against multiple existing and emerging strains”.

The first data on Omicron-specific vaccines are expected within months — although even that might be too late given how quickly the variant spreads. A pan-coronavirus vaccine that covers all strains as well as related viruses would be preferable, but “whether this will be possible isn’t yet clear,” says disease ecologist Marm Kilpatrick at the University of California Santa

Cruz. “There is always substantial uncertainty when dealing with viral evolution.”

Peter McIntyre, an infectious-disease specialist at the University of Otago in Dunedin, New Zealand, argues that, until we have new vaccines, strategies should prioritize protecting individuals against severe illness, boosting to shield vulnerable groups and using antivirals to keep people out of hospital.

“We need to keep our focus very firmly on protection against severe disease,” he says. “That is the yardstick we should be judging ourselves by.”

Nature **602**, 17-18 (2022)

doi: <https://doi.org/10.1038/d41586-022-00200-9>

References

1. 1.

Gardner, B. J. & Kilpatrick, M. Preprint at medRxiv
<https://doi.org/10.1101/2021.12.10.21267594> (2021).

2. 2.

Belik, M. *et al.* Preprint at Research Square
<https://doi.org/10.21203/rs.3.rs-1199296/v1> (2022).

3. 3.

Levine-Tiefenbrun, M. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.27.21268424> (2021).

4. 4.

Xia, H. *et al.* Preprint at bioRxiv
<https://doi.org/10.1101/2022.01.21.476344> (2022).

5. 5.

Goel, R. G. *et al.* *Science* <https://doi.org/10.1126/science.abm0829> (2021).

6. 6.

Naranbhai, V. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2022.01.04.21268586> (2022).

7. 7.

Barda, N. *et al.* *Lancet* [https://doi.org/10.1016/S0140-6736\(21\)02249-2](https://doi.org/10.1016/S0140-6736(21)02249-2) (2021).

8. 8.

Andrews, N. *et al.* *Nature Med.* <https://doi.org/10.1038/s41591-022-01699-1> (2022).

9. 9.

Arbel, R. *et al.* *N. Engl. J. Med.*
<https://doi.org/10.1056/NEJMoa2115624> (2021).

10. 10.

Garcia-Beltran, W. F. *et al.* *Cell*
<https://doi.org/10.1016/j.cell.2021.12.033> (2021).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00200-9>

- NEWS
- 27 January 2022

Big dog, little dog: mutation explains range of canine sizes

The genetic variant probably comes from ancient wolves.

- [Ewen Callaway](#)



Dogs differ in size more than any other mammal. Credit: Brand-X Pictures/Getty

From chihuahuas to great Danes, dogs differ more in size than any other mammal species on the planet. A mutation behind such variation has been

traced to an unexpected source: ancient wolves¹.

The mutation lies near a gene called *IGF1*, which researchers flagged 15 years ago as having a major role in the size variation of domestic dogs. It was the first of around two dozen such genes identified. But efforts to pinpoint the gene variant responsible had come up empty.

“*IGF1* has been a thorn in our side,” says Elaine Ostrander, a geneticist at the US National Human Genome Research Institute in Bethesda, Maryland, who led the 2007 study that first identified *IGF1*’s role in dog size², as well as the 27 January study in *Current Biology* that now fulfils the quest.

Ancient dogs, domesticated from wolves in the past 30,000 years, differed in size to some extent. But the current extreme size differences — the largest breeds are up to 40 times bigger than the smallest — emerged in the past 200 years, as humans established modern breeds.

Ostrander and colleagues including geneticist Jocelyn Plassais at INSERM-University of Rennes, France, analysed the genomes of more than 1,400 canids, including ancient dogs, wolves, coyotes and 230 modern dog breeds.

Growth control

When they compared variation in the region around the *IGF1* gene with body size in dogs and wild canids, one variant stood out. It lies in a stretch of DNA that encodes a molecule called a long non-coding RNA, which is involved in controlling levels of the IGF1 protein, a potent growth hormone.

The researchers identified two versions, or alleles, of the variant. Across all breeds, dogs with two copies of one allele tended to weigh less than 15 kilograms, whereas two copies of the other version were more common in dogs weighing more than 25 kilograms. Dogs with one copy of each allele tended to be intermediate in size, says Ostrander. Canines with two copies of the large-bodied allele also had higher levels of the IGF1 protein in their blood, compared with those with two copies of the ‘small’ allele.



The mutation that causes dogs' size variation has been traced to ancient wolves. Credit: Alexander Sviridov/Shutterstock

When the researchers looked at the genomes of other canids, they found a similar relationship. "This wasn't just a dog story. This was a wolf story and a fox story and a coyote story and everything story. It was canine-wide," says Ostrander.

Diminutive ancestors

The researchers think that the allele linked to small bodies is, evolutionarily, much older than the large-bodied version. Coyotes, jackals, foxes and most other canids they analysed had two copies of the 'small' version, suggesting that this version was present in a common ancestor of these animals.

It's not clear when the large-bodied allele evolved. The researchers found that an ancient wolf that lived in Siberia around 53,000 years ago carried one copy of this version. Other ancient wolves and modern grey wolves tend to

have two, suggesting that the large-bodied allele might have been beneficial to wolves.

The prevailing view among scientists used to be that small body size was probably linked to relatively new genetic changes, potentially unique to domestic dogs, says Robert Wayne, an evolutionary biologist at the University of California, Los Angeles. “This turns the whole story on its head. That’s what’s marvellous about the whole thing.”

The study could be a sign that dogs were domesticated from smaller-bodied wolves, dissimilar to present-day grey wolf populations, says Elinor Karlsson, a geneticist at the University of Massachusetts Chan Medical School in Worcester. “We don’t know what the wolves that led to dogs even looked like,” she says.

Researchers also caution that the story of dog size is far from complete. Plassais wants to work out how the variants influence levels of the *IGF1* protein. And the variant isn’t the only determinant of size in dogs: the *IGF1* gene itself accounts for about 15% of variation between breeds.

“We’re not talking about a mutation that makes a wolf chihuahua-sized,” says Karlsson. “We’re talking about one of many mutations that tends to make you a bit smaller.”

Nature **602**, 18 (2022)

doi: <https://doi.org/10.1038/d41586-022-00209-0>

References

1. 1.

Plassais, J. *et al.* *Curr. Biol.* <https://doi.org/10.1016/j.cub.2021.12.036> (2021).

2. 2.

Sutter, N. B. *et al.* *Science* **316**, 112–115 (2007).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00209-0>

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

- NEWS
- 21 January 2022

Deltacron: the story of the variant that wasn't

News of a ‘super variant’ combining Delta and Omicron spread rapidly last week, but researchers say it never existed and the sequences might have resulted from contamination.

- [Freda Kreier](#)



Researchers say that the ‘Deltacron’ sequences might be the result of lab errors. Credit: T. Narayan/Bloomberg/Getty

On 7 January, virologist Leondios Kostrikis announced on local television that his research group at the University of Cyprus in Nicosia had identified several SARS-CoV-2 genomes that featured elements of both the Delta and Omicron variants.

Named by them as ‘Deltacron,’ Kostrikis and his team uploaded 25 of the sequences to the popular public repository GISAID that evening, and another 27 a few days later. On 8 January, financial news outlet Bloomberg [picked up the story](#), and Deltacron became international news.

The response from the scientific community was swift. Many specialists declared both on social media and to the press that the 52 sequences did not point to a new variant, and were not the result of recombination — the genetic sharing of information — between viruses, but instead probably resulted from contamination in the laboratory.

“There is no such thing as #Deltacron,” [tweeted Krutika Kuppalli](#), a member of the World Health Organization’s COVID-19 technical team based at the Medical University of South Carolina in Charleston, on 9 January.
“#Omicron and #Delta did NOT form a super variant.”

Spread of misinformation

The story behind how a small crop of SARS-CoV-2 sequences became the focus of a brief and intense scientific controversy is complicated. And although some researchers applaud the system for quickly catching a possible sequencing error, others warn that the events of last week might offer a cautionary tale on the spread of misinformation during the pandemic.

Kostrikis says that aspects of his original hypothesis have been misconstrued, and that — despite the confusing name that some of the media took to mean that the sequences were those of a Delta–Omicron recombinant virus — he never said that the sequences represented a hybrid of the two.

Nevertheless, 72 hours after the researchers uploaded the sequences, Kostrikis removed them from public view on the database, pending further investigation.

Cheryl Bennett, an official at the GISAID Foundation's Washington DC office says, given that more than seven million SARS-CoV-2 genomes have been uploaded to the [GISAID](#) database since January 2020, some sequencing mistakes should not come as a surprise.

"However, rushing to conclusions on data that have just been made available by labs that find themselves under significant time pressure to generate data in a timely manner is not helpful in any outbreak," she says.

An error in the sequence?

The 'Deltacron' sequences were generated from virus samples obtained by Kostrikis and his team in December as part of an effort to track the spread of SARS-CoV-2 variants in Cyprus. While examining some of their sequences, the researchers noticed an Omicron-like genetic signature in the gene for the spike protein, which helps the virus to enter cells.

In an e-mail to *Nature*, Kostrikis explains that his initial hypothesis was that some Delta virus particles had independently evolved mutations in the spike gene similar to those common in Omicron. But after the wide news coverage, other scientists working on genetic sequencing and COVID-19 pointed out another possibility: a lab error.

Sequencing any genome depends on primers — short bits of manufactured DNA that serve as the starting point for sequencing by binding to the target sequence.

Delta, however, has a mutation in the spike gene that reduces some primers' ability to bind to it, making it harder to sequence this region of the genome. Omicron doesn't share this mutation, so if any Omicron particles were mixed into the sample owing to contamination, it might make the sequenced spike gene seem to be similar to that in Omicron, says Jeremy Kamil, a virologist at Louisiana State University Health Shreveport.

This type of contamination, says Kamil, is "so, so common".

Kostrikis counters that if Deltacron was a product of contamination, sequencing should have turned up Omicron sequences with Delta-like

mutations, because Omicron has its own primer-hindering mutation. He adds that the Deltacron lab contamination argument was “spearheaded by social media without considering our complete data, and without providing any real solid evidence that it is not real”.

Debunking debacle

However, other researchers have also pointed out that even if the sequences aren’t the result of contamination, the mutations identified by Kostrakis are not exclusive to Omicron and are found in other variants, making ‘Deltacron’ something of a misnomer.

In fact, GISAID is littered with sequences that have elements of sequences seen in other variants, says Thomas Peacock, a virologist at Imperial College London. Such sequences “get uploaded all the time”, he says. “But, generally, people don’t have to debunk them because there isn’t a load of international press all over them.”

“Scientists need to be very careful about what they are saying,” one virologist, who wanted to remain anonymous to avoid becoming embroiled in the controversy, told *Nature*. “When we say something, borders can be closed.”

Kostrakis now says he is “in the process of investigating all the crucial views expressed by prominent scientists around the world about my recent announcement”. He says he plans to submit the research for peer review.

In the interim, Kamil and other researchers fear that such incidents could make researchers more hesitant to share time-sensitive data. “You have to allow for the scientific community to self-correct,” he says. “And, in a pandemic, you have to facilitate the rapid sharing of viral genome data, because that’s how we find variants.”

Nature 602, 19 (2022)

doi: <https://doi.org/10.1038/d41586-022-00149-9>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00149-9>

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

- NEWS
- 31 January 2022

Will Omicron end the pandemic? Here's what experts say

The variant's rapid spread, different vaccine strategies and varying levels of immunity worldwide make the pandemic's future difficult to model.

- [David Adam](#)



Testing for Omicron: the variant has spread rapidly in countries worldwide. Credit: Fabrice Coffrini/AFP/Getty

On 11 January, just seven weeks after the Omicron variant was first reported, the World Health Organization (WHO) warned of a “tidal wave” of infection washing from west to east across the world. Fifty of the 53 countries in Europe and central Asia had reported cases of Omicron, said Hans Henri Kluge, the WHO’s regional director for Europe.

Countries would have to cope as best they could, he said, guided by their individual epidemiological situation, available resources, vaccination-uptake status and socio-economic context. In recent weeks, countries in Europe and the United States have felt the full force of the Omicron wave; in the United Kingdom, which has reported most infections, daily COVID-19 cases peaked at more than 160,000 earlier this month. Scientists there say all nations are facing the same major problem: the sheer speed at which the variant spreads.

And although the WHO and others have suggested that huge numbers of Omicron infections could signal the end of the pandemic, because of the short-term surge in immunity that will follow, researchers warn that the situation remains volatile and difficult to model.

“It moves so fast that it gives very little time to prepare any kind of response. So, decisions have to be made under huge uncertainty,” says Graham Medley, an infectious-diseases modeller at the London School of Hygiene & Tropical Medicine, who advises the UK government.

Rapid spread

Numbers of Omicron infections can double in less than two days, which is significantly faster than previous SARS-CoV-2 variants and closer to what public-health officials would expect from the milder influenza virus.

“Omicron is flu on acid,” one scientist says.

“We haven’t seen that speed before, and it meant you couldn’t vax your way out of it,” adds Christina Pagel, a health-care data analyst at University College London. “Even if you could vax everybody, it still takes two weeks for the vaccine to kick in, and by then you’re in the middle of it.”

That places policymakers and the researchers who advise them in an unenviable position. “It was a situation where you either put in restrictions very, very early, or you do nothing,” Pagel says. “But if you wait to see what happens, then it’s too late.”

Along with other countries, Britain tightened regulations in December. But it was a controversial move, particularly because reports from South Africa, which was hit by Omicron the previous month, suggested that the variant seemed to cause less severe illness and hospitalization — a conclusion now supported by the experience of the United Kingdom and other places.



The Philippines has experienced one of the worst surges in Omicron cases in southeast Asia. Credit: Ezra Acayan/Getty

Difficult to model

UK modellers were initially torn about how to use information from South Africa. It’s relatively straightforward to update a computer model to account for changes in the biological properties of a new variant. However, as the

pandemic has progressed, it has become much harder to simulate the baseline immune response of a country's population, and so to judge how that will limit spread.

In the early days of the pandemic, researchers could assume that most people worldwide were equally susceptible to infection, because COVID-19 was a new disease and no vaccines were available. But 12 months of different vaccine strategies, types and take-up rates from country to country, as well as fluctuating rates of infection and recovery, have left a diverse immunological landscape.

"The probability that infection will put someone in hospital is absolutely a key parameter. But we are now estimating that in an obviously not naive population," says Mark Woolhouse, an infectious-disease epidemiologist at the University of Edinburgh, UK, who also advises the government. "When you are making those sorts of estimates, formally you really should remake them for every population you're interested in. And that applies everywhere."

Vaccination differences

Modellers were confounded by the lack of specifics in South African data about reduced severity. "There was no quantitative analysis," Woolhouse says. "So, what numbers do you plug in? Are you saying 10% less pathogenic, or 50% less, or 90% less?"

Still, speaking in a personal capacity, Woolhouse says that some influential modellers in the United Kingdom were wrong not to allow for any reduced severity, instead working with assumed hospitalization rates for Omicron that were identical to those of previous variants. "That's clearly a pessimistic assumption," he says. "I do think it could have been much clearer from the beginning that there was this possibility it was less pathogenic and, you know, being crystal clear on what the policy implications of that difference might be."

The heterogeneity in immunological baselines and other important factors, such as population dynamics from country to country, make it difficult to

predict international spread of Omicron with any precision, or to assess — for example — how the variant might take hold in countries with lower vaccination rates. “It’s very hard to answer that question,” says Julian Tang, a consultant virologist at the Leicester Royal Infirmary, UK. “And it’s not very useful, because if you say it’s spreading in pattern XYZ across western Europe and then ABC across North America and MNO in Africa, that doesn’t really help anybody.”

Waning protection

The waning protection against infection that vaccines offer against Omicron also complicates the picture. Laboratory studies have indicated that inactivated-virus vaccines, which make up almost half of the ten billion doses distributed worldwide, [elicit few antibodies against the variant](#). Does that mean that Omicron will rip through places that rely on these shots even faster?

Not necessarily, says Woolhouse. “The inactivated-virus vaccines might induce a broader immunity that would react to a wider range of strains because it will elicit immune responses against viral proteins other than spike, which is particularly variable,” he says. “It’s a very interesting question but I simply haven’t seen a formal analysis of it yet.”

That’s because there are few real-world data to go on. “It’s only just hitting countries that have used them,” says Pagel.



Credit: Daniel Munoz/AFP/Getty

Among the countries that rely on inactivated-virus vaccines, Omicron seems to be making the most headway in the Philippines, which saw an exponential rise in COVID-19 cases this month, particularly in Manila. The number of new infections in the capital does seem to be dropping, but the virus is spreading farther afield. “Definitely cases are slowing down in the [National Capital Region], but in other regions it is now increasing,” said Maria Rosario Vergeire, health spokesperson for the Philippine government.

Vaccination rates are still relatively low in the Philippines, with just 53% of the population fully vaccinated. Officials there say they want to vaccinate all the country’s 77 million adults by May.

Although vaccines are likely to continue protecting against severe symptoms, Pagel says, infection will continue to spread. “I think the assumption is that none of the vaccines are going to give you long-lasting protection against infection,” she says.

Tang agrees: “I don’t think vaccines are the way this pandemic is going to end.”

When will it end?

So, how will it end? Not with Omicron, researchers predict. “This will not be the last variant, and so the next variant will have its own characteristics,” Medley says.

Given that the virus is unlikely to disappear completely, [COVID-19 will inevitably become an endemic disease](#), scientists say. But that’s a slippery concept, and one that means different things to different people. “I think it’s the expectation that the general behaviour is somehow towards the situation where we have so much immunity in the population that we would no longer see very deadly epidemics,” says Sebastian Funk, an epidemiologist at the London School of Hygiene & Tropical Medicine.

The transition to endemicity, or “living with the virus” without restrictions and safeguards, is difficult to model with any accuracy, he adds. That’s partly because even the best disease models struggle to make sensible forecasts beyond a few weeks ahead. It’s also because endemicity reflects a judgement call on how many deaths societies are willing to tolerate while the global population steadily builds up immunity.

For Woolhouse, COVID-19 will truly become endemic only when most adults are protected against severe infection because they have been exposed multiple times to the virus as children, and so have developed natural immunity. That will take decades, and it means many older people today (who were not exposed as children) will remain vulnerable and might need continued vaccinations.

That strategy has its flaws. Some of those exposed as children will develop long COVID. And it relies on children continuing to show much lower rates of severe illness as variants evolve.

There are no guarantees that the next variant will be milder, but Tang says that seems to be the pattern so far. “This virus is getting milder and milder with each iteration,” he says.

doi: <https://doi.org/10.1038/d41586-022-00210-7>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00210-7>

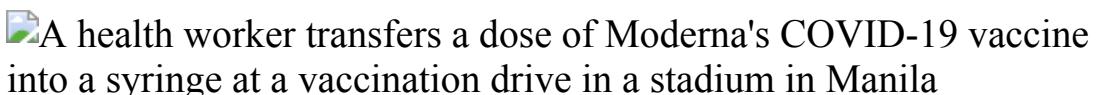
| [Section menu](#) | [Main menu](#) |

- NEWS FEATURE
- 02 February 2022

What the Omicron wave is revealing about human immunity

Immunologists have raced to work out how to protect against multiple variants of SARS-CoV-2. Their research has yielded a wealth of insights and a few surprises.

- [Cassandra Willyard](#) ⁰



A health worker in the Philippines prepares a dose of COVID-19 vaccine during the January surge of Omicron cases. Credit: Basilio Sepe/ZUMA Wire/Shutterstock

No one anticipated how quickly Omicron would sweep the globe. Although the surge from the variant is starting to decline in many countries, worldwide case numbers are still on the rise. The last full week of January saw about 23 million confirmed new cases; previous peaks topped out at about 5 million per week. Beleaguered public-health officials are still scrambling to curtail the virus's spread so that people with COVID-19 don't overwhelm the hospitals.

Omicron also presented immunologists with a new and urgent puzzle. Initial data suggest that existing vaccines, designed around the original SARS-CoV-2, do not provide much protection from becoming infected with the variant, even if they do seem to reduce the risk of hospitalization or death. The protection provided by two doses of a messenger RNA vaccine drops to less than 40% just a few months after the second dose^{1,2}. But a third,

‘booster’ dose seems to help. One report found about 60–70% protection from infection at two weeks after a third shot¹, and protection from severe illness seems strong².

“This is very exciting,” says Mark Slifka, an immunologist at Oregon Health & Science University in Portland. It’s also a little surprising. Why would a third encounter with a vaccine targeted to the original virus’s spike protein — which it uses to enter cells — work against this variant, which has more than 30 mutations in the spike?

The human immune system’s ability to remember past infections is one of its hallmarks, but a durable response is not guaranteed. Some infections and immunizations elicit lifelong protection, but for others, the response is modest and requires regular reminders in the form of booster shots or new, reformulated vaccines. COVID-19 has forced on the world a chance to explore the intricacies of this complex and crucial biological phenomenon. “It’s an amazing natural experiment,” says Donna Farber, an immunologist at Columbia University in New York City. “It’s just this unbelievable opportunity to look at human immune responses in real time.”

With around ten billion shots of a dozen COVID-19 vaccines already in people’s arms, and five worrying variants pulsing around the globe, scientists are scrambling to answer key questions. How long will vaccination protect people for? What will that protection look like? And, of course, how will a vaccine developed against the original SARS-CoV-2 fare against other variants, such as Omicron?

“We are just at the beginning of a wave of discovery,” says John Wherry, an immunologist at the University of Pennsylvania’s Perelman School of Medicine in Philadelphia. What emerges will be crucial not only for fighting COVID-19, but for understanding some of the most fundamental features of immune memory.

Making memories last

The immune system kicks into action soon after a pathogen enters the body. But it can take several days for the specialized cells that target viruses and

bacteria to join the battle. These B cells and T cells work to eradicate the infection; after the fight is over, they remember the intruder.

B cells “are the first responders”, says Ali Ellebedy, an immunologist at the Washington University School of Medicine in St. Louis, Missouri. During a first exposure to a pathogen, B cells that get activated divide rapidly and differentiate into plasma cells that churn out proteins called antibodies. Antibodies can flag suspicious intruders for destruction, and some might bind to a part of a pathogen that prevents it from infecting cells altogether. These are the ‘neutralizing’ antibodies. “They’re the only thing that can truly give you sterilizing immunity,” says Shane Crotty, an immunologist at the La Jolla Institute for Immunology in California. That’s why researchers typically use the presence of these antibodies as a proxy for immune protection.

By September 2020, a handful of studies^{3,4} reported that neutralizing-antibody levels were dropping in people who had recovered from COVID-19. Some experts expressed alarm that immunity to SARS-CoV-2 might therefore be fleeting.

Immunologists, however, weren’t surprised. Antibodies are supposed to wane after an infection. The short-lived B cells that churn out antibodies right away die off quickly. “This is something we’ve known forever,” says Rafi Ahmed, an immunologist and director of the Emory Vaccine Center at Emory University in Atlanta, Georgia.

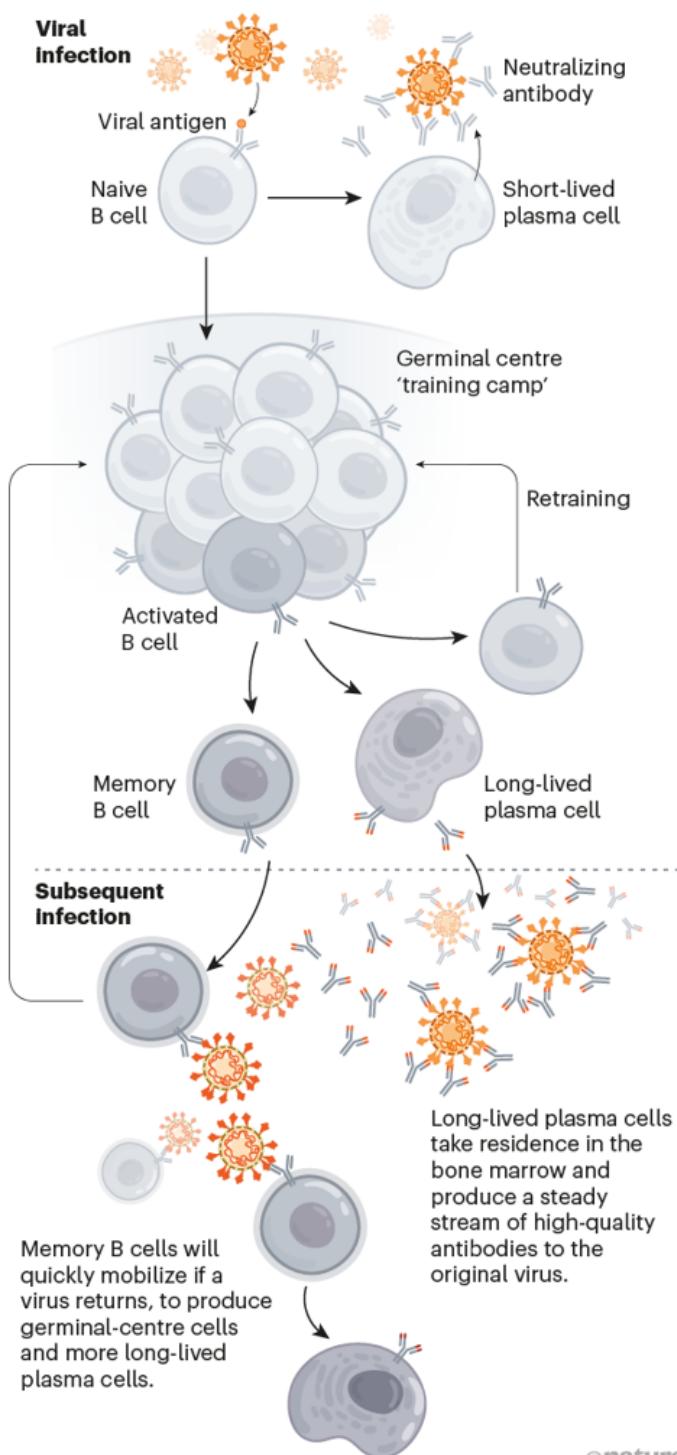
What matters is whether the body makes long-lived B cells that can target the pathogen if it reappears. These cells typically develop inside structures called germinal centres, which arise in the lymph nodes during an infection and serve as a sort of B-cell training camp. There, the cells multiply and acquire mutations. Only those that produce the best antibodies, the ones that latch most securely on to the surface of the virus, survive. It’s “almost a winnowing process”, Ellebedy says.

Within a month or so, some of the cells that produce these super-binders become memory B cells that circulate in the blood (see ‘B-cell memory’). They don’t produce antibodies, but if they encounter the virus or its proteins, they can rapidly divide and become plasma cells that do. The rest become

long-lived plasma cells that reside mainly in the bone marrow and secrete a small-but-steady stream of high-quality antibodies. “Those cells basically live with us for the rest of our lives,” Ellebedy says.

B-CELL MEMORY

When activated by an antigen molecule from a virus or vaccine, B cells multiply and go on to different fates. Some become short-lived plasma cells that produce antibodies. Other B cells enter germinal centres in the lymph nodes, where they multiply and mutate to enhance immunological memory with better antibodies and specialized cells.



Credit: Nik Spencer/*Nature*

A drop in antibody levels after infection is normal. What immunologists really want to know is where — or whether — the decline will stop. In April 2020, Ahmed and his team began studying people who had recovered from COVID-19. The scientists found that those people's antibody levels dropped quickly for the first two or three months after infection. But then, after about four months, the researchers saw the curve start to flatten. They have published results on the first eight months⁵, but now have data up to 450 days, and Ahmed is encouraged by what they see. So far, “looking at the shape of the curve, it looks pretty damn good”, he says. “It is really quite stable.”

The immune response after vaccination more or less mimics what happens after infection, with one major difference. In a SARS-CoV-2 infection, the immune system sees the whole virus. The most effective vaccines, however, are using just one viral protein to elicit a response: spike. And whether antibody levels will also plateau after vaccination isn't yet clear. Wherry and his colleagues analysed immune responses in 61 people for 6 months after their first shot, finding that antibody levels peaked about a week after the second shot and then fell quickly for a couple of months. After that, they declined more slowly⁶.

With that decline came a drop in protection. The shots, which became widely available in some countries as early as December 2020, showed impressive effectiveness initially. But by July 2021, reports began to surface of breakthrough infections. Data from Israel, which had launched an aggressive vaccination campaign using the Pfizer–BioNTech mRNA vaccine, suggested that this vaccine's protection against infection dropped from 95% to just 39% over the course of 5 months (see go.nature.com/3hjdxtn; in Hebrew and English).

Those numbers make it sound as though the vaccine is faltering. And researchers have seen that, over time, it does lose its ability to keep infection at bay. But vaccines have retained their ability to prevent serious illness. Protection from infection might be waning, but protection against hospitalization seems to be holding up. “You're probably going to have protective immunity for years,” Crotty says.

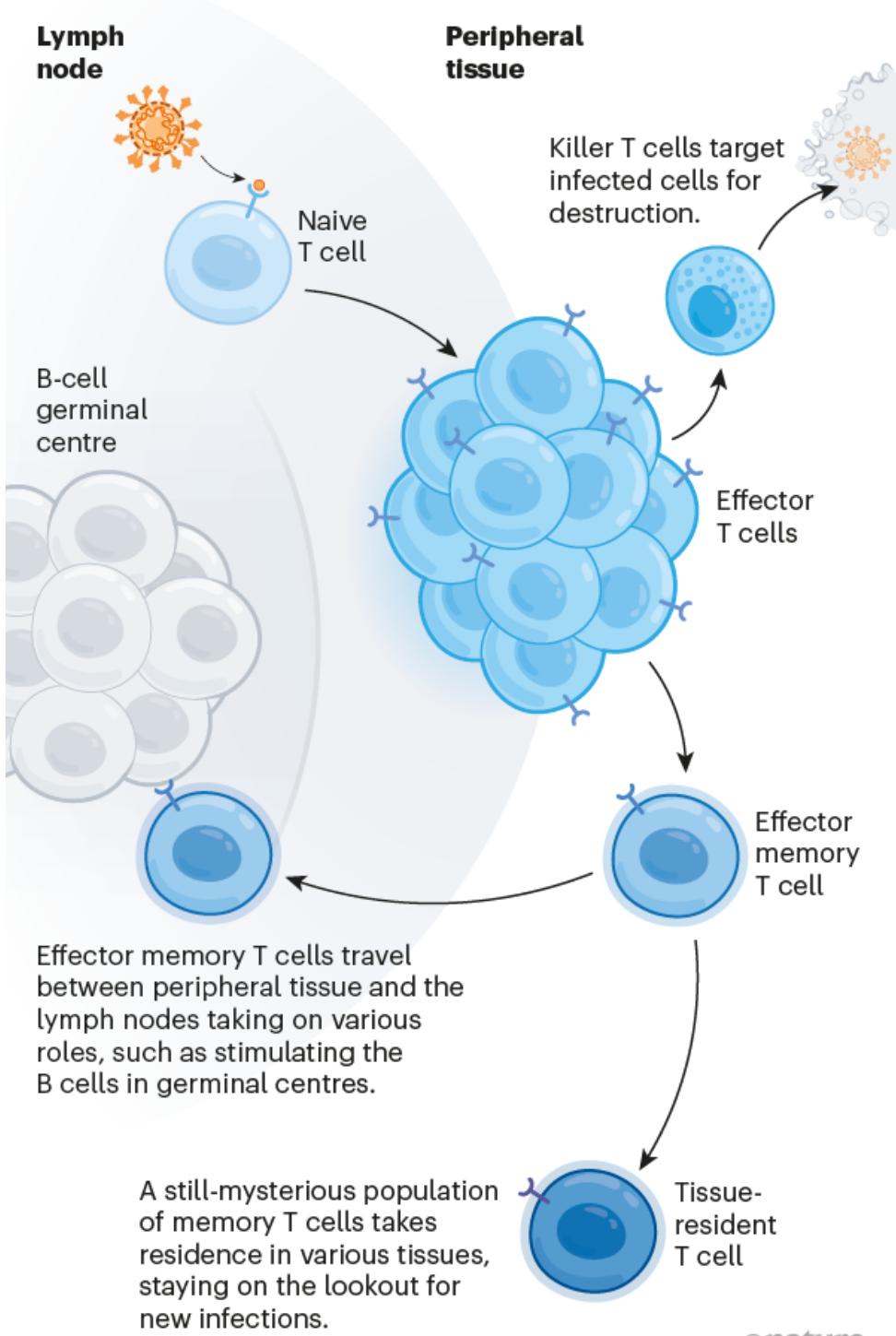
The cells will save us

Immune memory depends on more than just antibodies. Even when antibody levels drop, memory B cells can recognize a return invader, divide, and quickly start churning out antibodies to fight it. And the memory B-cell response improves over time, at least in the short term. Six months after vaccination, the individuals in Wherry's study⁶ had elevated numbers of memory B cells that responded not only to the original SARS-CoV-2, but also to three other variants of concern.

And then there are T cells, the third pillar of immune memory. On coming into contact with an antigen, these multiply into a pool of effector cells that act to wipe out the infection. Killer T cells quickly divide to assassinate infected cells, and various types of helper T cell secrete chemical signals that stimulate other parts of the immune system, including B cells. After the threat has passed, some of these cells persist as memory T cells (see 'T-cell memory').

T-CELL MEMORY

After a naive T cell encounters a viral antigen, it multiplies into several different kinds of effector cell that help to regulate the immune response. As the infection subsides, some will become memory T cells.



Credit: Nik Spencer/*Nature*

Some people might carry memory T cells from past coronavirus infections — such as those that cause common colds — that can recognize SARS-CoV-2. These cells could help to fight the infection, or even stop it completely. One study⁷ found that health-care workers who were exposed to SARS-CoV-2 but never tested positive had subtle signs of a response to infection. The researchers hypothesize that cross-reactive T cells shut the infection down before it could take hold. “These people did have an infection in a sort of loosest sense of the word,” says Mala Maini, an immunologist at University College London who led the study. But “there’s probably not much virus around because it’s being shut down very quickly”.

This idea is still controversial, and the phenomenon might be rare. Memory cells typically can’t block infection in the way that neutralizing antibodies can, but they don’t necessarily need to. With COVID-19, infection happens quickly, but it takes a little while to cause serious illness. That gives memory T cells some time to do their jobs. When re-exposed to a virus or booster, these cells will kick into overdrive, “proliferating like crazy”, Crotty says. “In a 24-hour period, you can get a tenfold increase in the number of your memory T cells.” That’s probably not fast enough to have much of an effect on getting sick, he adds. But it could be fast enough to prevent hospitalization.

And it’s much harder for the virus to find a way around the T-cell response. That’s because T cells in one individual recognize different parts of the virus than do T cells in another individual. So a virus could mutate to escape one person’s T-cell response, but not another’s. “Escape is meaningless at the population level,” Crotty says. Also, T cells can see parts of the virus (or the spike protein) that antibodies can’t, including pieces that are less likely to mutate.

Several studies have found that people who had been vaccinated or had been infected with SARS-CoV-2 had about the same T-cell response to Omicron as they did to the Delta variant, despite the large number of mutations^{8,9}. Observations of Omicron’s spread also suggest that this is so. A T-cell response is possibly also helping to drive the phenomenon known as ‘decoupling’. In areas with higher immunity because of past infections or

vaccination, the number of cases of Omicron has risen quickly, but the number of hospitalizations and deaths has increased much more slowly.

Evolution of immunity

A perfect vaccine would induce an immune response that is not only durable, but also broad enough to protect against the virus as it mutates and evolves. With Omicron raging, it seems the vaccines have lost some ground. But the immune system still has a number of tricks to deal with viruses that keep changing.

One of those tricks happens inside the germinal centres. There, the B-cell training not only improves how well antibodies bind to their original target; it can also boost the number of binding sites they recognize, increasing the odds that they can identify a variant.

“Indirectly, the whole success of vaccination depends on how robust the germinal centre is,” Ellebedy says. Dogma suggests that without the germinal centre, “we don’t have memory”.

But that might not be entirely true. The immune system has “a grab bag of other pathways” that are more nuanced and less well studied, says Stephanie Eisenbarth, director of the Center for Human Immunobiology at Northwestern University Feinberg School of Medicine in Chicago, Illinois. Research by Eisenbarth and her colleagues shows that even mice that lack the ability to make germinal centres can generate long-lived plasma cells¹⁰. How these cells arise isn’t entirely clear, but just like the plasma cells that come through the germinal centre, these seem to bind tightly to their targets.

Emerging data suggest that Omicron is, nevertheless, able to largely circumvent the antibodies generated by past infection or vaccination. Pfizer reported a 25-fold drop in the neutralization of Omicron (compared with the original SARS-CoV-2) in people who had received two vaccine doses. Why a third-dose booster might bring back protection isn’t entirely clear.

It’s possible that a third shot simply boosts all antibody levels equally, including the small proportion that can recognize pieces of Omicron’s spike

protein that haven't changed. "We know already from some of the data released by the companies that antibodies get boosted very, very efficiently," says Wherry. But it's looking likely that a third shot actually increases the breadth of the response.



Israeli officials authorized a fourth dose of a COVID-19 vaccine in January in hopes of boosting people's immunological memory.Credit: Amir Levy/Getty

In one study¹¹, researchers assessed blood from people who had received vaccines from Moderna, Pfizer–BioNTech or Johnson & Johnson to assess how well their antibodies neutralized a virus containing spike protein from SARS-CoV-2 variants. Blood from individuals who received one or two doses had little ability to neutralize Omicron. But blood from people who had received a booster dose of an mRNA vaccine fought the variant effectively. Their neutralization capacity against Omicron was only four- to sixfold lower than against the original strain.

People who have received two doses of vaccine have memory B cells that can bind to Omicron¹². It's possible that a third shot prompts these memory

cells to become antibody-producing cells. “One of the major jobs of memory B cells is to be a library of guesses by the immune system about what a variant may look like,” Crotty says.

Wherry offers another possibility. The booster might be triggering the formation of germinal centres, setting off another cascade of mutation among B cells. “That’s one of the things that we’re going to be watching carefully,” he says.

Slifka posits that the first dose of the vaccine generates antibodies that bind well to the features of the spike protein that are readily accessible. When subsequent doses arrive, existing antibodies quickly coat those accessible features, leaving less-accessible targets available for B cells to latch on to.

The good news about boosters, however, comes with a caveat. It’s not clear how long booster protection will last. Data from the United Kingdom suggest it could wane quickly¹³. Three doses of the Pfizer–BioNTech vaccine provided 70% protection initially. But by 10 weeks, protection against infection had dropped to 45%. And reports emerging from Israel suggest that [a fourth-dose booster doesn’t seem to elevate protection effectively](#). This suggests that the best next move might be to develop Omicron-specific booster shots.

Pfizer and Moderna are already working on mRNA versions of such jabs. In January, Pfizer chief executive Albert Bourla said that an Omicron-specific vaccine should be ready to launch by March. By then, however, many will already have been infected with the variant and gained some immunity that way. Pfizer is also working on a shot that would include both the original spike and one from Omicron. The ultimate goal, of course, is to develop a jab that would provide long-lasting immunity without multiple boosters.

The magic ingredient

SARS-CoV-2 could provide other opportunities for learning how to improve vaccination. In 2019, Slifka and his colleague Ian Amanna published a review¹⁴ looking at different types of vaccine and hunting for patterns that might help to predict why some induce durable immunity and others don’t.

Of the vaccine types they looked at, the longest-lasting protection tended to come from live-virus vaccines. These consist of pathogens that have been altered so that they can't cause disease. Because they mimic the actual infection so well, they tend to elicit a durable response. But those that contained whole inactivated virus or pieces of viral protein elicited good memory, too. What seems to matter, Slifka says, is the amount of time the antigen sticks around. "You don't have to be chronically infected," he says, "but it has to maintain stimulation of the immune system for a certain amount of time."

Slifka and Amanna didn't include mRNA vaccines in the paper — the technology wasn't in common use — but these do seem to fit the trend. For mRNA vaccines, the antigen gets produced by cells in the body (from an mRNA template). It sticks around for just a few weeks. And the evidence so far suggests that immunity might also be transient. But [RNA vaccines that have the ability to replicate in the body](#) might bring about longer-lasting immunity.

SARS-CoV-2 has given scientists a plethora of vaccines to observe and compare against the backdrop of an active pandemic, including those using whole, inactivated virus; protein; or mRNA, or those based on an adenovirus, such as Oxford–AstraZeneca's or Johnson & Johnson's offerings. There have been surprises. The response after a shot of the Johnson & Johnson vaccine, for example, elicits a weaker immune response than the mRNA vaccines initially, "and then it actually starts to get better over time", says Deepta Bhattacharya, an immunologist at the University of Arizona in Tucson. "Something interesting is happening there."

Scientists are also eager to understand what happens when people mix and match vaccines. A UK study known as Com-CoV has been investigating this phenomenon since early in the pandemic. Its most recent data¹⁵ show that people who received a first dose of either Oxford–AstraZeneca or Pfizer–BioNTech followed by Moderna had a higher antibody response than those who received a second dose of the same vaccine.

"You can think about it like cross training," Wherry says. Mixing and matching different kinds of vaccine might create a more flexible, diverse immune memory.

Adding more targets might also trigger better protection. The most effective current vaccines target the spike protein, but T cells can see the whole virus, says Bali Pulendran, an immunologist at Stanford University in California. He thinks of immunological memory as an enormous chandelier suspended by three thin wires: one represents the antibody response, one is memory B cells and the third is memory T cells. Each is important and should be considered in vaccine design. If one or two of the strands were severed, “would we be confident standing under it?” Pulendran asks.

A shot with broad, durable neutralizing activity against SARS-CoV-2 was always going to be a tall order. Much of that comes down to the nature of the virus itself. “If you look at respiratory infections, these historically have been very hard to prevent,” Ahmed says. That applies to influenza, [respiratory syncytial virus](#) and “we definitely see it with the common cold”. With a systemic infection, such as measles, it takes time for the virus to spread through the body and cause illness. With respiratory infections, it’s happening right at the point of entry. For such pathogens, protecting against serious illness might be the best anyone can hope for.

Many are still optimistic, however. “Everyone and their mother is studying SARS-CoV-2 right now,” says Scott Hensley, an immunologist at the Perelman School of Medicine. That surge of interest has led to remarkable advances in immunologists’ ability to dissect the immune response. The insights might finally help them to unlock the recipe for a vaccine that offers long-lasting, broad protection.

“What’s the magic sauce?” Pulendran asks. “Therein lies a deep, deep mystery, a fundamental challenge, which if it is solved will have a transformative effect on vaccinology.”

Nature **602**, 22-25 (2022)

doi: <https://doi.org/10.1038/d41586-022-00214-3>

References

1. 1.

Andrews, N. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.14.21267615> (2021).

2. 2.

Tseng, H. F. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2022.01.07.22268919> (2021).

3. 3.

Long, Q.-X. *et al.* *Nature Med.* **26**, 1200–1204 (2020).

4. 4.

Ibarrondo, F. J. *et al.* *N. Engl. J. Med.* **383**, 1085–1087 (2020).

5. 5.

Cohen, K. W. *et al.* *Cell Rep. Med.* **2**, 100354 (2021).

6. 6.

Goel, R. R. *et al.* *Science* **374**, eabm0829 (2021).

7. 7.

Swadling, L. *et al.* *Nature* **601**, 110–117 (2022).

8. 8.

Keeton, R. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.26.21268380> (2021).

9. 9.

Gao, Y. *et al.* *Nature Med.* <https://doi.org/10.1038/s41591-022-01700-x> (2022).

10. 10.

Chen, J. S. *et al.* *Sci. Immunol.*
<https://doi.org/10.1126/sciimmunol.abl5652> (2021).

11. 11.

Garcia-Beltran, W. F. *et al.* *Cell*
<https://doi.org/10.1016/j.cell.2021.12.033> (2021).

12. 12.

Tarke, A. *et al.* *Cell* <https://doi.org/10.1016/j.cell.2022.01.015> (2022).

13. 13.

UK Health Security Agency. *SARS-CoV-2 Variants of Concern and Variants Under Investigation in England: Technical Briefing 33* (2021).

14. 14.

Slifka, M. A. & Amanna, I. J. *Front. Immunol.*
<https://doi.org/10.3389/fimmu.2019.00956> (2019).

15. 15.

Stuart, A. S. V. *et al.* *Lancet* **399**, 36–49 (2022).

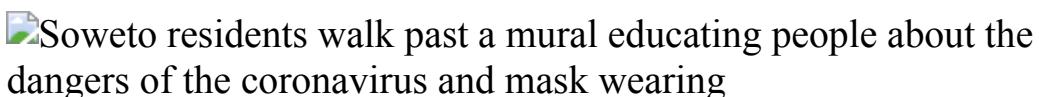
This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00214-3>

- NEWS FEATURE
- 28 January 2022

Where did Omicron come from? Three key theories

The highly transmissible variant emerged with a host of unusual mutations. Now scientists are trying to work out how it evolved.

- [Smriti Mallapaty](#)



People in Johannesburg, South Africa, near where cases of Omicron were first identified. Credit: Kim Ludbrook/EPA-EFE/Shutterstock

Little more than two months after it was first spotted in South Africa, the Omicron variant of the coronavirus SARS-CoV-2 has spread around the world faster than any previous versions. Scientists have tracked it in more than 120 countries, but remain puzzled by a key question: where did Omicron come from?

There's no transparent path of transmission linking Omicron to its predecessors. Instead, the variant has an unusual array of mutations, which it evolved entirely outside the view of researchers. Omicron is so different from earlier variants, such as Alpha and Delta, that evolutionary virologists estimate its closest-known genetic ancestor probably dates back to more than a year ago, some time after mid-2020 (ref. [1](#)). "It just came out of nowhere," says Darren Martin, a computational biologist at the University of Cape Town, South Africa.

The question of Omicron's origins is of more than academic importance. Working out under what conditions this highly transmissible variant arose

might help scientists to understand the risk of new variants emerging, and suggest steps to minimize it, says Angela Rasmussen, a virologist at the University of Saskatchewan Vaccine and Infectious Disease Organization in Saskatoon, Canada. “It’s very difficult to try to mitigate a risk that you can’t even remotely wrap your head around,” she says.

The World Health Organization’s recently formed Scientific Advisory Group for the Origins of Novel Pathogens (SAGO) met in January to discuss Omicron’s origins. The group is expected to release a report in early February, according to Marietjie Venter, a medical virologist at the University of Pretoria in South Africa, who chairs SAGO.

Ahead of that report, scientists are investigating three theories. Although researchers have sequenced millions of SARS-CoV-2 genomes, they might simply have missed a series of mutations that eventually led to Omicron. Alternatively, the variant might have evolved mutations in one person, as part of a long-term infection. Or it could have emerged unseen in other animal hosts, such as mice or rats.

For now, whichever idea a researcher favours “often comes down to gut feeling rather than any sort of principled argument”, says Richard Neher, a computational biologist at the University of Basel in Switzerland. “They are all fair game,” says Jinal Bhiman, a medical scientist at the National Institute for Communicable Diseases in Johannesburg, South Africa. “Everyone has their favourite hypothesis.”

Craziest genome

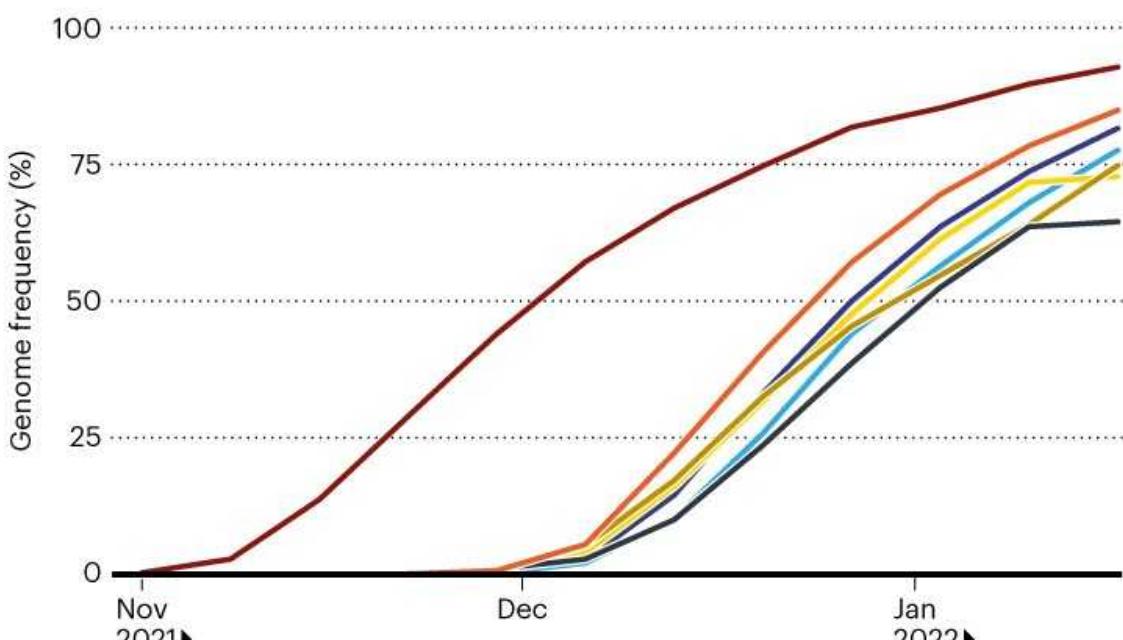
Researchers agree that Omicron is a recent arrival. It was first detected in South Africa and Botswana in early November 2021 (see ‘Omicron takeover’); retrospective testing has since found earlier samples from individuals in England on 1 and 3 November, and in South Africa, Nigeria and the United States on 2 November. An analysis of the mutation rate in hundreds of sequenced genomes, and of how quickly the virus had spread through populations by December, dates its emergence to not long before that — around the end of September or early October last year². In southern Africa, Omicron probably spread from the dense urban province of Gauteng,

between Johannesburg and Pretoria, to other provinces and to neighbouring Botswana.

OMICRON TAKEOVER

Omicron has quickly spread to become the world's dominant variant of the SARS-CoV-2 coronavirus — as shown by its prevalence in viral genomes uploaded to the GISAID database from various regions.

— Africa — Europe* — North America — South America
— Oceania — Global — Asia



* Including United Kingdom.

©nature

Source: [GISAID](#)

But because Johannesburg is home to the largest airport on the African continent, the variant could have emerged anywhere in the world — merely being picked up in South Africa because of the country's sophisticated genetic surveillance, says Tului de Oliveira, a bioinformatician at the University of KwaZulu-Natal in Durban and at Stellenbosch University's Centre for Epidemic Response and Innovation, [who has led South Africa's efforts to track viral variants](#), including Omicron.



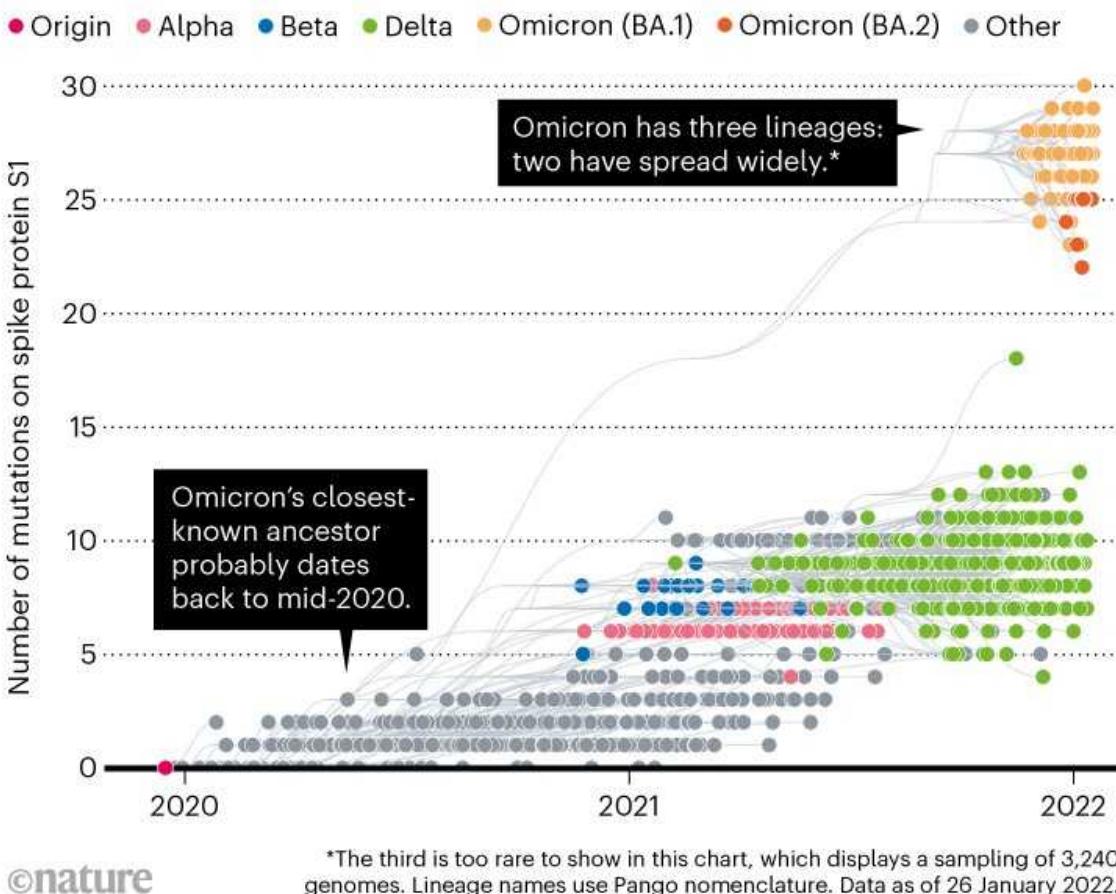
PhD student Upasana Ramphal in the laboratory of Tulio de Oliveira at the University of KwaZulu-Natal in Durban, whose group has led efforts to track Omicron and other variants in southern Africa. Credit: Joao Silva/NYT/Redux/eyevine

What stands out about Omicron is its remarkable number of mutations. Martin heard about it when he took a phone call from de Oliveira, who asked him to look at the craziest SARS-CoV-2 genome he had ever seen.

The variant has more than 50 mutations when compared with the original SARS-CoV-2 virus isolated in Wuhan, China (see go.nature.com/32utxva). Some 30 of these contribute to changes in amino acids in the spike protein¹, which the coronavirus uses to attach to and fuse with cells. Previous variants of concern have had no more than ten such spike mutations. “That is a hell of a lot of changes,” says Neher (see ‘Most mutated’).

MOST MUTATED

The Omicron variant of the SARS-CoV-2 coronavirus has more mutations than any known predecessor. This chart shows mutations in the S1 subunit of the spike protein, which attaches to host cells.



Source: [Nextstrain](#)

Researchers have seen many of these mutations before. Some were previously known to give the virus an increased ability to bind to the ACE2 receptor protein — which adorns host cells and is the docking point for SARS-CoV-2 — or to help it evade the body’s immune system. Omicron forms a stronger grip on ACE2 than do previously seen variants³. It is also better at evading the virus-blocking ‘neutralizing’ antibodies⁴ produced by people who have been vaccinated, or who have been infected with earlier variants. Other changes in the spike protein seem to have modified how Omicron enters cells: it appears to be less adept at fusing directly with the

cell's membrane, and instead tends to gain entry after being engulfed in an endosome (a lipid-surrounded bubble)³.

But more than a dozen of Omicron's mutations are extremely rare: some have not been seen at all before, and others have popped up but disappeared again quickly, presumably because they gave the virus a disadvantage¹.

Another curious feature of Omicron is that, from a genomic viewpoint, it consists of three distinct sublineages (called BA.1, BA.2 and BA.3) that all seem to have emerged at around the same time — two of which have taken off globally. That means Omicron had time to diversify before scientists noticed it. Any theory about its origins has to take this feature into account, as well as the number of mutations, notes Joel Wertheim, a molecular epidemiologist at the University of California, San Diego.

Silent spread

Researchers have explained the emergence of previous variants of concern through a simple process of gradual evolution. As SARS-CoV-2 replicates and transmits from person to person, random changes crop up in its RNA sequence, some of which persist. Scientists have observed that, in a given lineage, about one or two single-letter mutations a month make it into the general viral circulation — a mutation rate about half that of influenza. It is also possible for chunks of coronavirus genomes to shuffle and recombine wholesale, adds Kristian Andersen, an infectious-disease researcher at Scripps Research in La Jolla, California. And viruses can evolve faster when there is selection pressure, he says, because mutations are more likely to stick around if they give the virus an increased ability to propagate under certain environmental conditions.

Some scientists think that person-to-person spread would not be conducive to accumulating as many changes as Omicron has since mid-2020. "It does seem like a year and a half is a really short period of time for that many mutations to emerge and to apparently be selected for," says Rasmussen.

But Bhiman argues that enough time has elapsed. She thinks the mutation process could have occurred unseen, in a region of the world that has limited

genomic sequencing and among people who don't typically get tested, perhaps because they didn't have symptoms. At some point in the past few months, she says, something happened to help Omicron explode, maybe because the progress of other variants — such as Delta — was gradually impeded by the immunity built up from vaccination and previous infection, whereas Omicron was able to evade this barrier.

Although researchers have submitted almost 7.5 million SARS-CoV-2 sequences to the GISAID genome database, hundreds of millions of viral genomes from people with COVID-19 worldwide have not been sequenced. South Africa, with some 28,000 genomes, has sequenced less than 1% of its known COVID-19 cases, and many nearby countries, from Tanzania to Zimbabwe and Mozambique, have submitted fewer than 1,000 sequences to GISAID (see 'Missing genomes').

MISSING GENOMES

The GISAID database contains sequenced SARS-CoV-2 genomes representing less than 1% of the reported COVID-19 cases in each of Africa, Asia and South America.



©nature

Data as of 27 January 2022.

Source: [GISAID](#)

Martin says that researchers need to sequence SARS-CoV-2 genomes from these countries to get a better sense of the likelihood of unobserved evolution. It is possible that the three sublineages of Omicron each separately arrived in South Africa from a region with limited sequencing capacity, he says.

But de Oliveira says the scenario that Omicron evolved unseen through person-to-person transmission is “extremely implausible”. Intermediate steps in Omicron’s evolution should have been picked up in viral genomes from people travelling from countries that do little sequencing to those that do a lot.

“This is not the nineteenth century, where you take six months to go from point to point by sailboat,” says Sergei Pond, a computational evolutionary biologist at Temple University in Philadelphia, Pennsylvania.

And Andersen adds that, because some of Omicron’s mutations haven’t been seen before, the variant might have evolved in an environment not involving person-to-person chains of transmission. Some of the changes in Omicron don’t match any seen even in the broader viral group of sarbecoviruses, which includes the virus that causes severe acute respiratory syndrome (SARS). For example, one particular site on the genomes of all known sarbecoviruses encodes a serine amino acid, but a mutation in Omicron means the variant has a lysine at that position¹, which changes the biochemistry of that region, Andersen says.

However, says Jesse Bloom, a viral evolutionary geneticist at the Fred Hutchinson Cancer Research Center in Seattle, Washington, SARS-CoV-2 has not yet explored all of its possibilities in people. “The virus is still expanding in the evolutionary space.”

Chronic infection

An alternative incubator for fast-paced evolution is a person with a chronic infection. There, the virus can multiply for weeks or months, and different types of mutation can emerge to dodge the body’s immune system. Chronic infections give the virus “the opportunity to play cat and mouse with the immune system”, says Pond, who thinks it is a plausible hypothesis for Omicron’s emergence.

Such chronic infections have been observed in people with compromised immune systems who cannot easily get rid of SARS-CoV-2. For example, a December 2020 case report described a 45-year-old man with a persistent

infection⁵. During almost five months in its host, SARS-CoV-2 accumulated close to a dozen amino-acid changes in its spike protein. Some researchers suggest Alpha emerged in someone with a chronic infection, because, like Omicron, it seems to have accumulated changes at an accelerated rate (see go.nature.com/3yj6kmh).

“The virus has to change to stick around,” says Ben Murrell, an interdisciplinary virologist at the Karolinska Institute in Stockholm. The receptor-binding domain, where many of Omicron’s mutations are concentrated, is an easy target for antibodies, and probably comes under pressure to change in a long-term infection.



Health workers stand outside a building under lockdown in Hong Kong, amid a rise in Omicron coronavirus cases. Credit: Louise Delmotte/AFP/Getty

But none of the viruses from individuals with chronic infections studied so far has had the scale of mutations observed in Omicron. Achieving that would require high rates of viral replication for a long time, which would

presumably make that person very unwell, says Rasmussen. “It seems like a lot of mutations for just one person.”

Further complicating the picture, Omicron’s properties could stem from combinations of mutations working together. For example, two mutations found in Omicron — N501Y together with Q498R — increase a variant’s ability to bind to the ACE2 protein by almost 20 times, according to cell studies⁶. Preliminary research by Martin and his colleagues suggests that the dozen or so rare mutations in Omicron form three separate clusters, in which they seem to work together to compensate for the negative effects of any single one¹.

If this is the case, it means that the virus would have to replicate sufficiently in a person’s body to explore the effects of combinations of mutations — which would take longer to achieve than if it were sampling the space of possible mutations one by one.

One possibility is that multiple individuals with chronic infections were involved, or that Omicron’s ancestor came from someone with a long-term infection and then spent some time in the general population before being detected. “There are a lot of open questions,” says Rasmussen.

Proving this theory is close to impossible, because researchers would need to be lucky enough to find the particular person or group that could have sparked Omicron’s emergence. Still, more comprehensive studies of SARS-CoV-2’s evolution in chronic infections would help to map out the range of possibilities, says Neher.

Mouse or rat

Omicron might not have emerged in a person at all. SARS-CoV-2 is a promiscuous virus: it has spread to a wild leopard, to hyenas and hippopotamuses at zoos, and into pet ferrets and hamsters. It has caused havoc in mink farms across Europe, and has infiltrated populations of white-tailed deer throughout North America. And Omicron might be able to enter a broader selection of animals. Cell-based studies have found that, unlike

earlier variants, Omicron's spike protein can bind to the ACE2 protein of turkeys, chickens and mice^{[3,7](#)}.

One study found that the N501Y–Q498R combination of mutations allows variants to bind tightly to rat ACE2 (ref. ^{[6](#)}). And Robert Garry, a virologist at Tulane University in New Orleans, Louisiana, notes that several other mutations in Omicron have been seen in SARS-CoV-2 viruses adapting to rodents in laboratory experiments.

The types of single-nucleotide substitution observed in Omicron's genome also seem to reflect those typically observed when coronaviruses evolve in mice, and do not match as well with the switches that are observed in coronaviruses adapting to people, according to a study of 45 mutations in Omicron^{[8](#)}. The study noted that, in human hosts, G to U substitutions tend to occur in RNA viruses at a higher rate than C to A switches do, but that Omicron does not show this pattern.

It is possible, then, that SARS-CoV-2 could have acquired mutations that gave it access to rats — jumping from an ill person to a rat, possibly through contaminated sewage — and then spread and evolved into Omicron in that animal population. An infected rat could later have come into contact with a person, sparking the emergence of Omicron. The three sublineages of Omicron are sufficiently distinct that, according to this theory, each would represent a separate jump from animal to human.

A large population of animals with infections lasting longer than in humans could give SARS-CoV-2 room to explore a wide diversity of mutations and “build up a large ghost population of viruses that no one knows about”, says Martin, who says he finds this ‘reverse zoonosis’ theory convincing. Changes that make the virus better at spreading in its animal host won’t necessarily affect its ability to infect people, he says.

An animal reservoir could also explain why some of the mutations in Omicron have been rarely seen before in people, says Andersen.

In the dark

But others say that even a single viral jump from an animal to a person is a rare event — let alone three. Meanwhile, the virus has had plenty of opportunities to slip between people. And although some of Omicron's mutations have been seen in rodents, that doesn't mean they can't happen or haven't occurred in people, too, and have simply been missed.

Murrell also points out that SARS-CoV-2 didn't immediately go through a period of accelerated evolution after jumping to people for the first time. When it spread to mink and deer, it did pick up changes, but not as many mutations as Omicron has accumulated, says Spyros Lytras, an evolutionary virologist at the University of Glasgow, UK. This means that the evidence isn't sufficient to suggest Omicron's predecessor would have undergone rapid selection after finding a new home in the wild.

To confirm this theory, researchers would need to find close relatives of Omicron in another animal, but they haven't been looking — "something that has been horribly neglected", says Martin. Since the pandemic began, researchers have sequenced fewer than 2,000 SARS-CoV-2 genomes isolated from other animals, mostly from mink, cats and deer.

Now that Omicron has taken off, how it evolves in people could offer more clues about its origins. It might, for instance, shed mutations that, in retrospect, are found to have helped it adapt to a different animal host, or in a person with a chronic infection. But it could also not change by much, leaving researchers in the dark.

The answer to Omicron's emergence will probably be one or a combination of the three scenarios, says Bloom. But, he adds, researchers are far from explaining the processes that brought Omicron here, let alone predicting what the next variant will look like.

And many scientists say they might never find out where Omicron came from. "Omicron really shows us the need for humility in thinking about our ability to understand the processes that are shaping the evolution of viruses like SARS-CoV-2," says Bloom.

doi: <https://doi.org/10.1038/d41586-022-00215-2>

References

1. 1.

Martin, D. P. *et al.* Preprint at bioRxiv
<https://doi.org/10.1101/2022.01.14.476382> (2022).

2. 2.

Viana, R. *et al.* *Nature* <https://doi.org/10.1038/s41586-022-04411-y> (2022).

3. 3.

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

4. 4.

Cele, S. *et al.* Preprint at medRxiv
<https://doi.org/10.1101/2021.12.08.21267417> (2021).

5. 5.

Choi, B. *et al.* *N. Engl. J. Med.* **383**, 2291–2293 (2020).

6. 6.

Bate, N. *et al.* Preprint at bioRxiv
<https://doi.org/10.1101/2021.12.23.473975> (2021).

7. 7.

Cameroni, E. *et al.* Preprint at bioRxiv
<https://doi.org/10.1101/2021.12.12.472269> (2021).

8. 8.

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00215-2>

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

Opinion

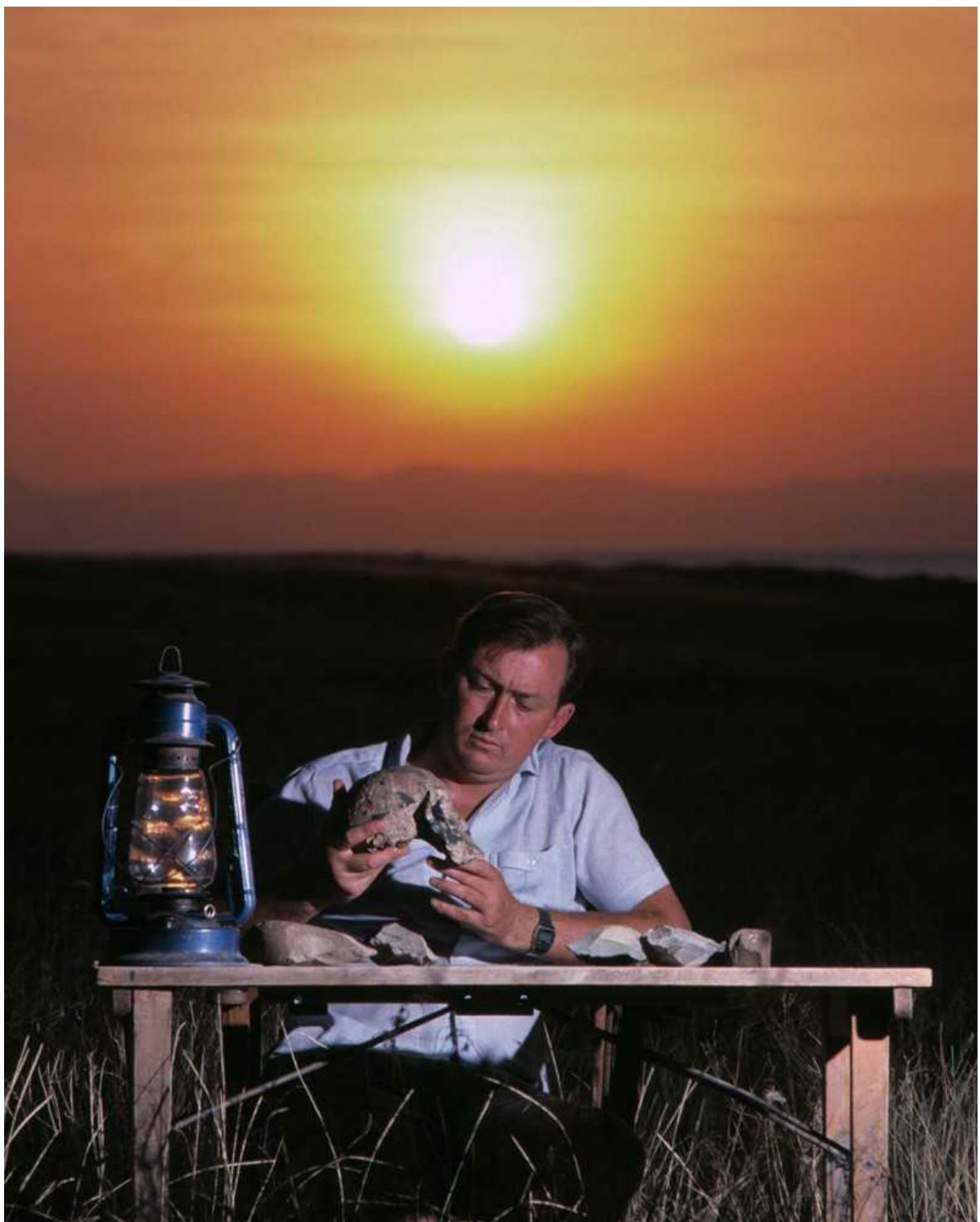
- **[Richard Leakey \(1944–2022\)](#)** [28 January 2022]
Obituary • Palaeontologist of human origins, conservationist and politician.
- **[Survey of gender bias in the IPCC](#)** [01 February 2022]
Comment • The Intergovernmental Panel on Climate Change needs to do more to include the expertise and voices of women, even as numbers and policies improve.
- **[COVID-19: LMICs need antivirals as well as vaccines](#)** [01 February 2022]
Correspondence •
- **[Emissions cuts take political and social innovation too](#)** [01 February 2022]
Correspondence •
- **[FAO should focus on real not nominal food prices](#)** [01 February 2022]
Correspondence •
- **[Early-career satisfaction: industry beats academia](#)** [01 February 2022]
Correspondence •

- OBITUARY
- 28 January 2022

Richard Leakey (1944–2022)

Palaeontologist of human origins, conservationist and politician.

- [Marta Mirazón Lahr](#) 0



Credit: William Campbell/Sygma/Getty

Richard Leakey made palaeontological discoveries of lasting significance, and brought animal poaching to the world's attention. His fossil finds at

Koobi Fora on the shores of Lake Turkana, Kenya, transformed our understanding of the diversity of human ancestors. He directed Kenya's national museum, reorganized the country's wildlife services and headed Kenya's civil service. He died aged 77, at home in the Ngong Hills, Kenya.

In science, he liked exploration, big-picture problems and building institutions. He made huge strides in conservation, empowering organizations and deploying shock tactics. He entered politics, creating an opposition party, then worked in government, finally becoming its corruption watchdog. He mentored young Kenyan scholars, conservationists and artists who are now leaders in their field.

Born in Nairobi, Richard was the middle child of pioneers in African palaeontology and archaeology Louis and Mary Leakey. He abandoned school at 16 to open an animal-trapping and safari business, earning enough to pay for flying lessons and his own small plane. In 1963, a mix of interest in his parents' world and a wish to prove himself to them lured him into the study of the past, and he found his first important hominin fossil — a 1.5 million-year-old mandible of *Paranthropus boisei* — in 1964.

In 1967, Leakey's father asked him to direct an expedition to the Omo Valley of southern Ethiopia. There, Leakey found two *Homo sapiens* fossils now known to be 230,000 years old ([C. M. Vidal et al. *Nature* <https://doi.org/gn3794; 2022>](https://doi.org/gn3794; 2022)), key evidence of our species' African origins. Flying over the eastern shore of Lake Turkana, he recognized the potential of sediments at Koobi Fora, which proved to be a trove of hominin fossils. The discovery of different hominin species living at the same time between 2 million and 1.5 million years ago (*P. boisei*, *Homo habilis*, *Homo rudolfensis* and *Homo erectus*) changed views of how humans evolved.

In 1968, Leakey became director of the National Museums of Kenya, which became a hub of thriving research. Soon afterwards, he met the young British zoologist Meave Epps. They married after his first marriage ended, and became life-long personal and scientific partners. Their work with researchers dubbed the Hominid Gang, led by Kamoya Kimeu, resulted in the discovery of dozens of hominin fossils, including a new genus and four new species (*Paranthropus aethiopicus*, *Australopithecus anamensis* and *Kenyanthropus platyops*, as well as *H. rudolfensis*). A 1.6-million-year-old

skeleton of a juvenile *H. erectus* proved to have grown more slowly than apes and faster than humans, giving insights into the evolution of human life-history.

Leakey became involved in acrimonious scientific arguments — sometimes he was right, sometimes not — which, during the 1970s, gave an antagonistic tone to human-origins research. His health deteriorated, and he had his first kidney transplant (donated by his brother Philip) in 1980. In 1989, Kenya's president, Daniel arap Moi, asked him to run the Kenya Wildlife Service (KWS). Leakey declared war on poachers, burnt the stockpile of Kenyan ivory and massively reduced elephant deaths. His controversial tactics had an impact on a web of corrupt practices and created serious enemies. In 1993, the plane he was piloting crashed; both his legs had to be amputated below the knee. Sabotage was rumoured.

The relationship with Moi became increasingly hostile. In 1995, Leakey left KWS to create an opposition party, Safina, becoming a member of the Kenyan parliament in 1998. His time in opposition was tense. Leakey was beaten and received death threats. But Kenya needed large investments, and funders demanded assurances. Capitalizing on Leakey's reputation for integrity, in 1998 Moi asked him to direct KWS again, and in 1999 to head the civil service. Over three years, Leakey raised hundreds of millions of dollars for Kenya and fought corruption.

In 2002, he accepted a position at Stony Brook University, New York, that allowed him to live in Kenya and create the Turkana Basin Institute (TBI), which he chaired from 2005 until his death. TBI fostered a burst of discoveries: Miocene primates, hominins, the oldest stone tools in the world at 3.3 million years, evidence of prehistoric warfare, and the earliest monumental architecture in sub-Saharan Africa. In 2004, Leakey founded WildlifeDirect, a non-governmental conservation body, serving on its board for 10 years. In 2007, he became chair of Transparency International Kenya, continuing his battle against corruption.

By this time, Leakey had skin cancer and progressively worse health. He underwent a second kidney transplant in 2006, with Meave as the donor, and a liver transplant in 2013. Yet, in 2015, he accepted President Uhuru Kenyatta's request to return to KWS as chair until 2018. For the past six

years, he worked to create a new Kenyan museum, called Ngaren — of which I am a board member — to celebrate science, evolution and humanity's African origins.

Richard was special — fun, insightful, generous, with a sharp sense of humour, and a fabulous cook and sommelier. He embraced life, good and bad, and imbued those around him with the sheer excitement of what could be done, discovered, resolved and enjoyed.

Nature **602**, 29 (2022)

doi: <https://doi.org/10.1038/d41586-022-00211-6>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00211-6>

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

- COMMENT
- 01 February 2022

Survey of gender bias in the IPCC

The Intergovernmental Panel on Climate Change needs to do more to include the expertise and voices of women, even as numbers and policies improve.

- [Diana Liverman](#)⁰,
- [Nicolena vonHedemann](#)¹,
- [Patricia Nying'uro](#)²,
- [Markku Rummukainen](#)³,
- [Kerstin Stendahl](#)⁴,
- [Miriam Gay-Antaki](#)⁵,
- [Marlies Craig](#)⁶,
- [Lorena Aguilar](#)⁷,
- [Paulette Bynoe](#)⁸,
- [Friedemann Call](#)⁹,
- [Sarah Connors](#)¹⁰,
- [Laura David](#)¹¹,
- [Andrew Ferrone](#)¹²,
- [Bronwyn Hayward](#)¹³,
- [Shiromani Jayawardena](#)¹⁴,
- [Lamin Mai Touray](#)¹⁵,
- [Jyoti Parikh](#)¹⁶,
- [Minal Pathak](#)¹⁷,
- [Rosa Perez](#)¹⁸,
- [Anna Pirani](#)¹⁹,
- [Anjal Prakash](#)²⁰,
- [Christiane Textor](#)²¹,
- [Lourdes Tibig](#)²²,

- [Melinda Tignor](#)²³,
- [Çiğdem Tuğac](#)²⁴,
- [Carolina Vera](#)²⁵ &
- [Radha Wagle](#)²⁶



IPCC researchers huddle in Switzerland to address the impacts of climate change on land for a special report in August 2019.Credit: Mike Muzurakis/IISD/ENB

Women are increasingly prominent in climate negotiations. Familiar figures include United Nations climate chiefs Patricia Espinosa and Christiana Figueres, Barbados Prime Minister Mia Mottley and youth activist Greta Thunberg. Yet gender equity is far from being realized across the climate research community, including in the Intergovernmental Panel on Climate Change (IPCC). Although the numbers of women involved in writing IPCC reports have increased steadily since the 1990s, a gender imbalance and barriers to women's participation persist.

In 2018, the IPCC established the Task Group on Gender to compile a report and make recommendations. The report, presented at a plenary session in May 2019, included the results of a survey of IPCC participants that showed ongoing gender biases and barriers. As members of that task group — including report authors, staff members and government representatives — we distil our findings here. We also describe subsequent actions and set out what still needs to be done as the IPCC wraps up its sixth assessment cycle in 2023.

Gendered experiences

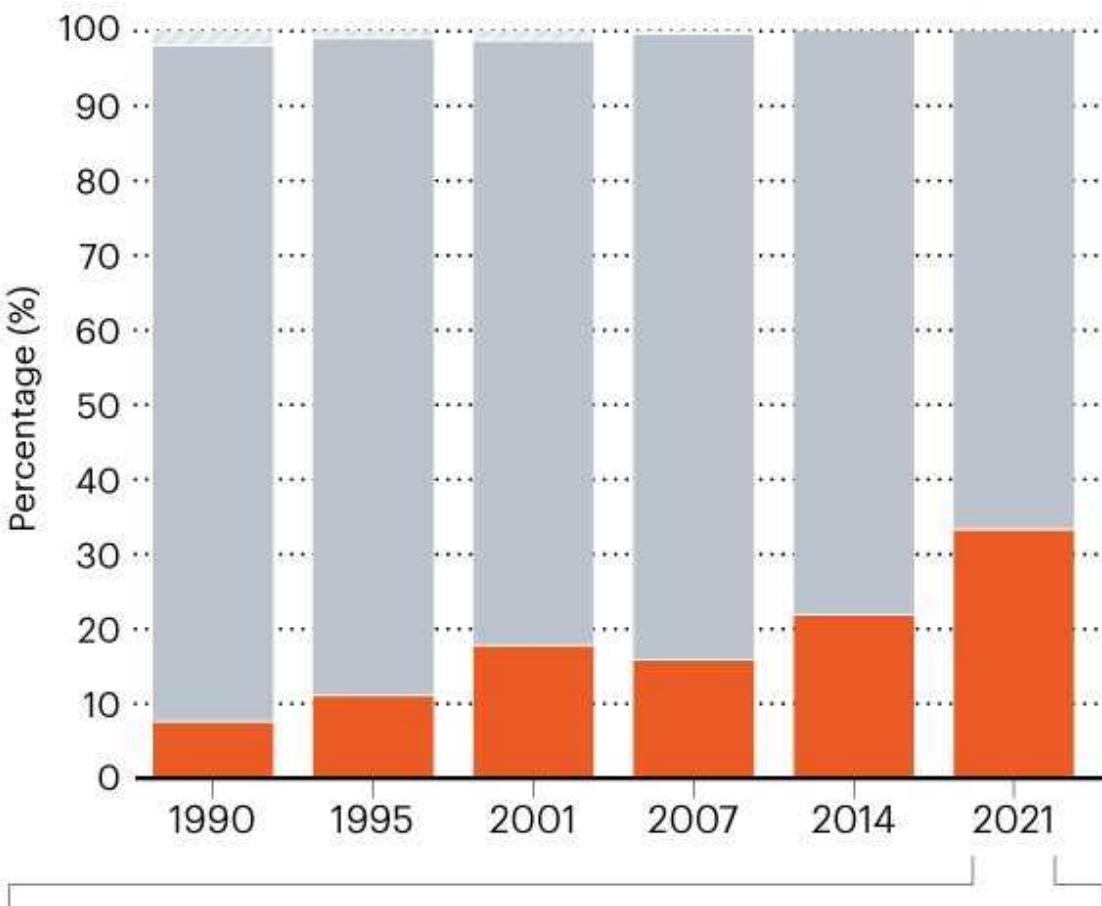
The IPCC task group sent a survey to 1,520 contributors to understand how they perceived and experienced gender bias and barriers during their work with the IPCC. We received 533 replies. The response rate was higher for women (39%) than for men (28%; see Supplementary information; one person chose ‘other’ gender).

We found that women’s representation has grown — from just 8% of authors in the first assessment report in 1990, to one-third in the Sixth Assessment Report (AR6; see ‘IPCC gender breakdown’). However, the survey revealed ongoing challenges (see ‘IPCC gender survey: select results’).

IPCC GENDER BREAKDOWN

The percentage of female core authors on teams writing assessment reports has grown, from 8% in 1990 to 33% in 2021.

■ Female ■ Male ■ Unknown



Female participation varies across working groups (WG).



Source: D. Liverman *et al.*

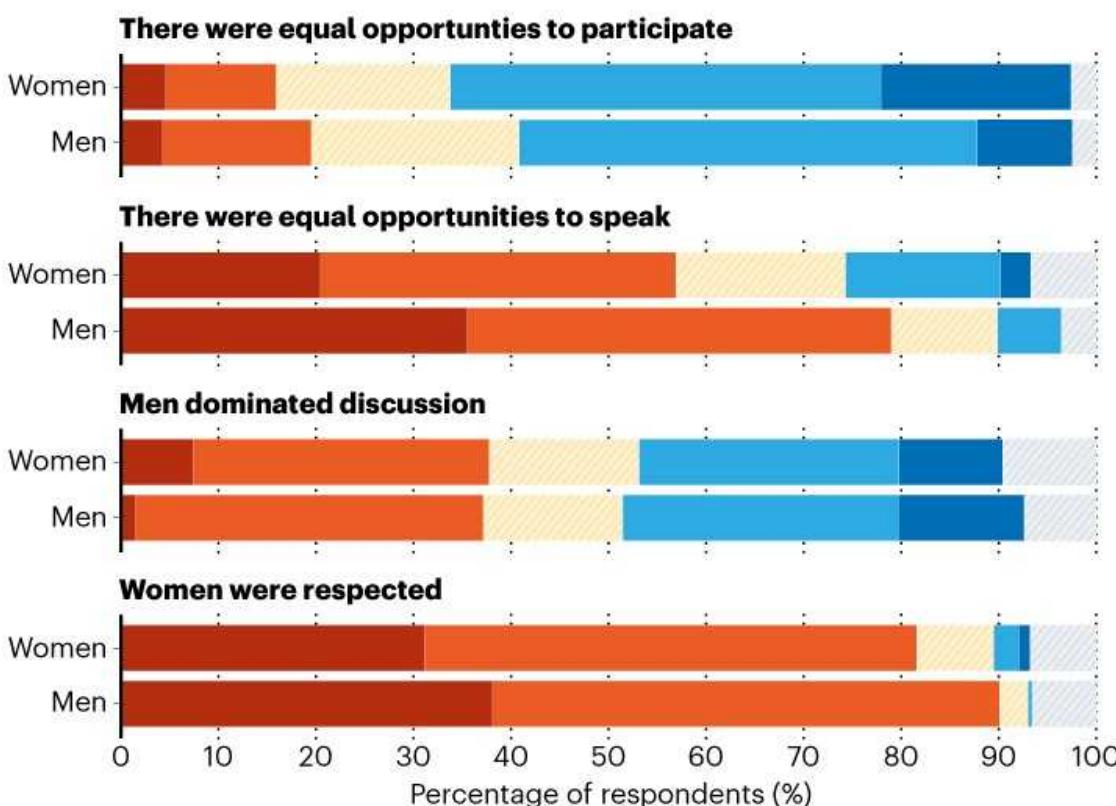
Why does this matter? Fair representation and broad expertise are essential when considering an issue as global, urgent and cross-cutting as climate change. Research suggests that, overall, women are more likely than men to be affected by climate change. Events such as drought and storms undermine people's ability to provide food, water and child care, and those roles tend to be taken on by women in many societies. Women are more likely to lack access to land, insurance and disaster relief^{1,2}. And as caregivers, farmers and leaders of communities and organizations, they also have crucial roles in responding to climate change.

IPCC work boosts scientific careers. Nomination and appointment as a lead author or review editor, or election as a bureau member, brings international recognition, academic repute and the potential to influence policy. Growing numbers of studies reveal how diverse perspectives can improve decision-making and the quality of studies as well as inspire new generations of researchers^{3–5}.

IPCC GENDER SURVEY: SELECT RESULTS

Of 533 contributors who responded, fewer women than men reported* that they were heard or respected in meetings.

■ Strongly agree ■ Agree or somewhat agree ■ Neutral
■ Somewhat disagree or disagree ■ Strongly disagree ■ No answer



©nature

*See Supplementary information for full survey results.

Source: D. Liverman *et al.*

In the climate sciences, as in other disciplines, women face compounding barriers — from unequal access to training and funding to fewer promotions and citations, lower wages, scarce role models, more harassment and greater family responsibilities compared with men. Women in the Earth sciences are less likely than men to be nominated for awards or to be offered senior leadership positions^{6,7}. They are disproportionately disadvantaged by metrics used to evaluate research. For example, in a Reuters 2021 ranking of the world's 1,000 leading climate scientists, which was based on scientometrics, only 2 women made it into the top 50 and just 12% of the total list were women (see go.nature.com/3rq9mf6).

Since it noted the task group's report in May 2019, the IPCC has established another task group to develop a gender policy and implementation plan. The plan was adopted by the IPCC in February 2020 and sets out three priorities: equal opportunities for participation and leadership; a gender-inclusive environment; and raising awareness through training and guidance. Some lead-author meetings for AR6 have included training on gender and cultural sensitivity. The IPCC website includes a page highlighting gender issues (see go.nature.com/3hd4tpk). The organization spotlights female authors and those from developing nations in its media feeds and events. We hope that it will keep up this momentum.

Such measures have been even more important during the COVID-19 pandemic, because participation in IPCC work can be hindered by unequal technological access, travel restrictions and commitments to family or home institutions. The pandemic emerged in the middle of the sixth assessment cycle. Meetings moved online and the timeline was delayed. Although remote meetings might have helped some authors by avoiding the need to travel, they also raised new challenges, some related to gender bias. Delegates faced difficulties with Internet connections, time zones and the loss of in-person discussions. Efforts to compress meeting times could have limited the opportunity for diverse voices to speak, including those who are less confident or are not as comfortable speaking in English. If the IPCC continues with virtual meetings, it will need to address these challenges.

Survey results

Our survey revealed some positive trends. Women's rising involvement (see 'IPCC gender breakdown') reflects broader shifts in science. Women now constitute about 30% of researchers worldwide, but have lower representation at senior levels (see go.nature.com/3g6ej88). The IPCC's rules of procedure list gender as one criterion for selecting author teams, in addition to disciplinary and regional balance.

More than three-quarters of both women and men who responded to our survey agreed that the gender balance has improved. Overall, 79% felt positive about the transparency of decision-making, and 89% were positive about the learning experience. At least three-quarters had excellent or good

experiences in terms of being respected and listened to, and in making professional connections.



UN climate chief Patricia Espinosa speaks during the COP26 climate summit in Glasgow, UK, in November 2021. Credit: Ian Forsyth/Getty

However, women were 15% less likely than men to agree that everyone has equal opportunities to be nominated, speak, shape content or lead chapters. Members of our task group and several survey respondents commented that increased numbers do not always equate to greater influence if women are excluded or not given voice or power.

Researchers have reported an ongoing imbalance in scientific and regional representation in the IPCC^{8–10}. IPCC authors tend to be drawn more from the natural than the social sciences, and from wealthy nations rather than developing ones.

Gender balance differs across the working groups. Women are best represented in Working Group II, which covers climate impacts, adaptation and vulnerability and draws on the ecological and social sciences (see ‘IPCC

gender breakdown'). They are less well represented in the groups that focus on the physical sciences (I) or mitigation (III). This variation might reflect gender inequalities in the disciplines involved.

Meetings of IPCC authors are fast-paced, with a sense of urgency. IPCC processes and procedures are designed to achieve scientific rigour and excellence but, in practice, nationality, culture, gender, language, seniority and other variables can affect whose voices prevail.

More than one-third of respondents (both men and women) perceived that male scientists dominate discussions and writing. Women were less likely than men (64% versus 78%) to agree that all points of view were represented in discussions. Although most respondents said they had not themselves experienced or observed gender bias and discrimination, women were on average 15% more likely than men to have witnessed discrimination against others.

More women than men reported that they had observed someone else take credit for a woman's idea (38% versus 24%), or had seen a woman being ignored (52% versus 30%) or patronized (41% versus 27%). Around one-third of women reported that someone had implied at least once that they were in the IPCC only because of their gender. Worryingly, some women had experienced (8%) or observed (11.5%) sexual harassment while working with the IPCC.

Unequal barriers

Most nominations to the IPCC are made through government agencies and other national focal points. These can reflect scientific hierarchies and biases in countries and organizations that favour men. Cultural patterns such as a greater reluctance by women to put themselves forward and obligations to family could also be factors. Opportunities to join the IPCC might not be widely publicized, narrowing the pool.

In the survey, the top six barriers that both men and women identified as most inhibiting their own ability to contribute were: lack of time (55%), childcare obligations (33%), not having confidence to challenge others

(32%), problems with accessing computers or research materials (31%), inadequate financial support from their home country (31%) and limited writing skills (24%).



Female authors of the IPCC 1.5°C report in Brazil on International Women's Day in 2017. Credit: Lincoln Alves

Many IPCC authors contribute on top of their full-time jobs. Most bureau members and authors are not paid by the IPCC. They also generally have to fund their own travel, although travel support is provided for people from developing countries. Almost twice as many women (44%) as men (24%) reported childcare responsibilities as a barrier. Also, 40% of women saw their lack of confidence in challenging others as a barrier to inclusion, compared with 26% of men. Respondents saw these barriers as greater for others than themselves, especially lack of time (66%), writing skills (64%), access to computers and materials (44%) and English language proficiency (41%).

The survey highlighted the importance of other dimensions of diversity that intersect with gender, and can be barriers, including ethnicity, race, nationality, religion, disability and age. Several respondents reported seeing

themselves or colleagues be brushed aside owing to a lack of fluency in English, or to youth, race, gender or being from developing countries.

Survey respondents suggested ways to improve gender balance. These included widening the pool of nominations through broader publicity, mentoring and nominating opportunities, and establishing targets for and monitoring of gender balance in nominations and in the IPCC. They suggested that training on gender issues and guidance on group facilitation would help ensure that the voices of women and those with limited fluency in English are included fairly. They proposed formal processes, such as neutral points of contact and regular surveys for monitoring and managing issues related to gender, bias or harassment.

They also asked that IPCC processes be made more sensitive to family issues, including pregnancy, and child- and elder care. Health and travel risks that disproportionately affect women should be addressed by selecting conference locations and transport that pose a low risk to personal safety. Some respondents emphasized that opportunities for remote participation and access to research publications need to be expanded.

We echoed these survey comments in the recommendations in our task-group report (see ‘Six priorities for gender bias’).

Six priorities for gender bias

Recommendations that the Intergovernmental Panel on Climate Change gender task group called for in its 2019 report.

1. National and other contact points should consider and monitor gender balance in soliciting nominations; build diverse capacity; and share best practices.
2. Develop an IPCC gender policy and implementation plan and gender committee, with objectives, actions, monitoring and regular reporting.
3. Increase the share of women in the leadership of the IPCC; mainstream gender concerns into the selection of authors, review editors and reviewers.

4. Provide training on inclusive practices, gender balance and consensus decision-making, especially for those leading chapters and reports.
5. Undertake regular surveys and feedback.
6. Ensure that IPCC meetings take into account travel safety, family issues and pregnancy, with support and options for remote participation.

Next steps

As the sixth assessment cycle concludes over the next year, the IPCC will reflect on its processes and draw lessons. A new gender action team proposes to undertake another survey of experiences in this cycle, and to further develop a code of conduct and formal training on diversity. An expert meeting on diversity and inclusivity is planned for the seventh assessment cycle.

Through continued attention and effort, we all look forward to a more balanced and gender-inclusive environment in the IPCC and in climate science more broadly.

Nature **602**, 30-32 (2022)

doi: <https://doi.org/10.1038/d41586-022-00208-1>

References

1. 1.

Nagel, J. *Gender and Climate Change: Impacts, Science, Policy* (Routledge, 2015).

2. 2.

Pearse, R. *WIREs Clim. Change* **8**, e451 (2017).

3. 3.

Nielsen, M. W. *et al.* *Proc. Natl Acad. Sci. USA* **114**, 1740–1742 (2017).

4. 4.

Campbell, L. G., Mehtani, S., Dozier, M. E. & Rinehart, J. *PLoS ONE* **8**, e79147 (2013).

5. 5.

National Academies of Sciences, Engineering, and Medicine. *Promising Practices for Addressing the Underrepresentation of Women in Science, Engineering, and Medicine* (NASEM, 2020).

6. 6.

Marín-Spiotta, E. *et al.* *Adv. Geosci.* **53**, 117–127 (2020).

7. 7.

Popp, A. L., Lutz, S. R., Khatami, S., van Emmerik, T. H. M. & Knoben, W. J. M. *Earth Space Sci.* **6**, 1460–1468 (2019).

8. 8.

Ho-Lem, C., Zerriffi, H. & Kandlikar, M. *Glob. Environ. Chang.* **21**, 1308–1317 (2011).

9. 9.

Corbera, E., Calvet-Mir, L., Hughes, H. & Paterson, M. *Nature Clim. Change* **6**, 94–99 (2016).

10. 10.

Yamineva, Y. *Environ. Sci. Pol.* **77**, 244–251 (2017).

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

- CORRESPONDENCE
- 01 February 2022

COVID-19: LMICs need antivirals as well as vaccines

- [Simar Singh Bajaj](#) ⁰ &
- [Fatima Cody Stanford](#) ¹

In the battle against COVID-19, there must be global equity in the distribution of antiviral drugs such as molnupiravir and nirmatrelvir. These are potent public-health tools beyond disease mitigation and vaccination (see, for example, [Nature 601, 165; 2022](#)). Cheap and easy to use, these pills are well suited to low- and middle-income countries (LMICs).

In under-vaccinated LMICs, widespread access to antiviral drugs is crucial. Antivirals could be a more enduring defence against emerging SARS-CoV-2 variants than are vaccines and monoclonal antibodies. They have the potential to prevent hundreds of thousands of hospitalizations and deaths globally. However, LMICs currently face significant manufacturing and pricing obstacles, and high-income countries have bought much of the existing supply (see [go.nature.com/3nfz](#)).

Production of these antivirals must be rapidly scaled up through compulsory licensing, the US Defense Production Act and other strategies (see [go.nature.com/3kcxkga](#)). Issues of supply, demand and cost can all be addressed readily, in tandem with vaccine equity, to achieve global pill equity.

Nature **602**, 33 (2022)

doi: <https://doi.org/10.1038/d41586-022-00220-5>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00220-5>

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

- CORRESPONDENCE
- 01 February 2022

Emissions cuts take political and social innovation too

- [Jonas De keersmaecker](#) ORCID: <http://orcid.org/0000-0002-8062-7422>⁰,
- [Katharina Schmid](#) ORCID: <http://orcid.org/0000-0001-6018-9245>¹ &
- [Sander van der Linden](#) ORCID: <http://orcid.org/0000-0002-0269-1744>²

Your Editorial makes no mention of the pivotal role of societal support in combating global warming ([Nature 601, 7; 2022](#)). As has become clear during the COVID-19 pandemic, it is naive to assume that new technologies and scientific recommendations will be unequivocally embraced by policymakers and the public. Fighting climate change will depend crucially on people's support for the transition from fossil fuels to renewable energy.

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 **602**, 33 (2022)

doi: <https://doi.org/10.1038/d41586-022-00219-y>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00219-y>.

- CORRESPONDENCE
- 01 February 2022

FAO should focus on real not nominal food prices

- [Alastair M. Smith](#) 0

The latest report from the Food and Agricultural Organization (FAO) of the United Nations risks conveying an overly optimistic interpretation of global food security to the press and the public when it concludes that, although nominal food prices grew by more than 28% between 2020 and 2021, they remain below their 2012 peak (see go.nature.com/3rrojmc). This is not the case for ‘real’ food prices, which take inflation factors into account. I argue that focusing on real, rather than nominal, food prices is a better indicator of food accessibility.

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 **602**, 33 (2022)

doi: <https://doi.org/10.1038/d41586-022-00221-4>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00221-4>

- CORRESPONDENCE
- 01 February 2022

Early-career satisfaction: industry beats academia

- [Shusheng Geng](#) ✉

Your conclusion that industry scores higher than academia for job satisfaction ([Nature 600, 8; 2021](#)) doesn't factor in how this can vary at different career stages. Inspection of the raw data from the *Nature*-sponsored survey (see [go.nature.com/3eqc](#)) indicates that dissatisfaction is highest among young scientists in academia. Greater financial and institutional support is essential to help this group to overcome their difficulties. It would also be an investment in the future.

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 **602**, 33 (2022)

doi: <https://doi.org/10.1038/d41586-022-00222-3>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00222-3>

Work

- **Why early-career researchers should step up to the peer-review plate** [31 January 2022]

Career Feature • Science benefits when junior scientists sign up as reviewers. It's also good for their careers.

- **How to fix your scientific coding errors** [31 January 2022]

Technology Feature • Software bugs are frustrating. Adopting some simple strategies can help you to avoid them, and fix them when they occur.

- **Grape expectations: making Australian wine more sustainable** [31 January 2022]

Where I Work • Geoff Gurr is working on an eco-friendly approach to controlling pests in New South Wales vineyards.

- CAREER FEATURE
- 31 January 2022

Why early-career researchers should step up to the peer-review plate

Science benefits when junior scientists sign up as reviewers. It's also good for their careers.

- [Amber Dance](#) ✉

[Find a new job](#)



The process of peer review provides a key service to the scientific enterprise. Credit: Getty

Seyedali Mirjalili had just received his master's degree in computer science in 2011 when he received his first invitation to peer review a paper.

"I had no idea what the e-mail was about, to be honest," recalls Mirjalili, then attending the University of Technology Malaysia in Johor Bahru. He accepted anyway, and asked the journal editor for guidance about completing the review. He hasn't stopped reviewing since. "The more I did it, the better I got," says Mirjalili, now a computer scientist at Torrens University Australia in Brisbane. He currently reviews one or two papers a day, and sits on the editorial board of several journals. To manage the workload, he sticks to papers that fit his expertise and therefore don't require any further research on his part.

Reviewers contribute a key service to peer-reviewed science, catching errors or problems, helping authors to improve their work and even rejecting shoddy research. Scientists who review also benefit: they see original research before it is made public, helping them to stay up to date in their fields, and gain insight into the review process so that they can improve their own submissions.

In some ways, it is also a thankless task. "It's unseen labour," says Rebeccah Lijek, a molecular biologist and peer-review scholar at Mount Holyoke College in South Hadley, Massachusetts. "When you have a lot on your plate, it's the kind of thing that can drop down on your priority list."

The result is a system in which one-fifth of researchers contribute up to 94% of reviews¹. "The literature has exploded, and there are vastly more papers than a handful of people can handle," says Randy Schekman, a cell biologist at the University of California, Berkeley, and founding editor of the journal *eLife*. The COVID-19 crisis has exacerbated the problem, with a proliferation of preprints awaiting attention. According to the website RetractionWatch, more than 200 published papers relating to the pandemic have been retracted (see go.nature.com/3rhwnbj).

Journals and editors are eager to add a diverse range of scientists to their talent pools, including early-career researchers from around the world. “It’s important to get different perspectives,” says Schekman. “Senior people might have a broad view of a subject, but not be versed in the details.” Yet [junior researchers can be invisible to editors](#) because they haven’t published much or lack an online presence.

For early-career scientists, there are ways to get noticed and to learn to review. Peer-review training has not been a standard component of postgraduate curricula, but it is possible to gain know-how from a mentor or from online courses offered by several organizations, including Nature Masterclasses (part of Springer Nature, which publishes *Nature*). However, novice reviewers must take care to understand what’s expected and what constitutes a conflict of interest, as well as how the open peer-review movement — which seeks to make reviews and publishing decisions more transparent — might affect the anonymity of their comments.

“Reliable reviewers are always in high demand,” says Thereza Soares, a chemist at the University of São Paulo in Brazil and an editor for the *Journal of Chemical Information and Modeling*.

Peer pressure

Despite the need for diversity, peer review has an ongoing problem in this area. [Publons](#), a service that gives reviewers public credit for their work through their online profiles, analysed this aspect in 2018. It found that researchers in certain nations, including several in the global south, contribute fewer reviews than do those in much of Europe, North America and Japan². The report estimated that women are also likely to be under-represented in peer review. And when a group of researchers analysed the peer-reviewer pool in Frontiers journals, they found that women were under-represented, that male editors were more likely to appoint male reviewers and that female editors more often invited female reviewers³. Junior scientists made up just 3% of the invited reviewers in that sample.

The effects of this older, predominantly male reviewer pool trickle down into manuscript acceptance rates, according to a paper posted on the preprint

server bioRxiv (and thus not yet peer reviewed itself)⁴. Information scientist Cassidy Sugimoto at the Georgia Institute of Technology in Atlanta and her colleagues analysed thousands of *eLife* submissions, and found that women, as well as researchers who were not from North America or Europe, were less likely to be editors, reviewers and authors. When editors and reviewers happened to match the gender or geography of authors, acceptance rates went up.

However, not all studies follow this pattern. The authors of a 2021 paper covering 145 journals found that papers by women were favourably reviewed overall, particularly in biomedicine, health sciences and social sciences⁵. Acceptance rates for manuscripts with a higher proportion of female authors were greater than for those with mostly male authorship in biomedicine, health sciences and physical sciences.

Kathrin Rousk, an ecologist at the University of Copenhagen, sometimes feels that even when she is invited to review, her opinions aren't taken seriously because she is relatively young and female. At times, she says, when she has expressed serious reservations about a manuscript or even recommended it be rejected, the paper was nonetheless approved without many revisions. "I get the impression that the opinion doesn't weigh as much as those of the senior male colleagues," she says. "It's just my feeling."

Even the tone of reviews matters and can influence the diversity of science overall. According to a 2019 survey of more than 1,000 researchers, 58% of respondents had received a review that was unprofessional in content or tone⁶. Women, non-binary people and people of colour were more likely than white men to report a drop in self-confidence and disruptions to their publication rate or career advancement as a result.

As a graduate student at Pennsylvania State University in State College in 2015, Carolyn Trietsch received a harsh review of her first paper. In describing the textured back of a particular wasp, she mistakenly used the word alveolate, which means honeycombed, instead of foveolate, meaning pitted. One reviewer was irate. "He basically wrote a three-page rant," recalls Trietsch, who now coordinates the development of multidisciplinary research proposals at the university. Trietsch was able to laugh it off — and

fixed the offending vocabulary — but other scientists might not be so resilient.

“Personal attacks are never a good look” for reviewers, says Lijek. That is, reviewers should focus on the science, not the scientist.

Yet, Lijek says, the reviewer pool is probably more diverse than it seems because of the practice of ghostwriting, in which junior scientists pen reviews on a senior researcher’s behalf. Lijek and her colleagues surveyed postdocs and other early-career researchers, and found that about half had ghostwritten a review. This happened even though more than 80% of respondents thought ghostwriting without credit was unethical.⁷



Ecologist Kathrin Rousk feels her opinions in reviews are sometimes overlooked. Credit: Sidsel Kretzschmer Henriksen

The practice also means that editors can’t check for any conflicts of interest between reviewers and authors, says Lijek’s colleague, Gary McDowell,

who is chief executive and founder of the research-policy consultancy firm Lightoller in Chicago, Illinois. Furthermore, ghostwriting means that unnamed reviewers don't receive boosts to their profiles that might lead to further review opportunities.

Rousk sometimes asks postdocs or graduate students in her group to collaborate with her on reviews, so they learn how to do it. But she's careful to ensure the journal editor is comfortable with that approach, and to be transparent about who has written the review.

Junior researchers have plenty to offer the peer-review process, McDowell says. In fact, studies find that the best reviews often come from younger scientists^{8–10}. “Early-career people write very thoughtful, long reviews,” says McDowell. One advantage of including reviewers who are not faculty members is that they're often more familiar with current experimental methods and equipment. When McDowell was a postdoc at Tufts University in Boston, Massachusetts, his supervisor looked to him to unravel technical details of papers under review and credited him when journal policy allowed.

A foot in the door

There are no hard guidelines for who's ready to accept their first solo review assignment, although individual journals do sometimes specify reviewer criteria. Critical-thinking skills and expertise in the subject matter are crucial.

Early-career researchers should tell their supervisors they'd like to help with a future review, or express interest to other academics who are journal editors. Scientists can also reach out to editors by e-mail or at conferences. Many will appreciate the contact, says Mirjalili: “They love it.”

Trietsch received her first review invitation a few years after publishing her first paper, when she was midway through her PhD in entomology. She'd taken a professional-development course, but peer review hadn't been covered. “I just wasn't sure how to get started, what I should focus on, or

not,” she recalls. But help was at hand; Trietsch asked a professor in her department for guidance.

“Experience in peer review doesn’t receive much focus as part of academic training,” says Benjamin Mudrak, a senior author-product manager at the American Chemical Society (ACS) who is based in Durham, North Carolina, and oversees the ACS [‘Reviewer Lab’ training course on peer review](#). Launched in 2017, the course offers six free modules on how to understand the peer-review process, assess a paper and write a clear, useful review. On completion, scientists can be flagged as a graduate in the ACS reviewer database.

Training and mentoring are also available through the free [Web of Science Academy](#). “It provides the solid principles in academic peer review,” such as how to evaluate the abstract and how to determine whether the literature review is sufficiently comprehensive, says Diaa Ahmedien, a new-media arts researcher at Helwan University in Cairo.

He took the course, then called Publons Academy, in 2017 on completing his PhD at the University of Bern. For the final examination, he selected three published articles to review, and then a professor in his field testified to Publons that his work was satisfactory. As a result, his Publons profile lists him as an academy graduate. (That ‘test’ has been replaced with the Web of Science module ‘Co-reviewing with a mentor’, but still results in credit on the Publons site.) Since then, Ahmedien has completed more than 240 reviews and has become an associate editor of *Humanities and Social Sciences Communications*, published by Springer Nature.

The Genetics Society of America (GSA) also offers an [online peer-review course and mentorship programme](#) to early-career researchers from around the world. After training, participants enter the reviewer pool at the journal *Genetics*. The course runs annually and is likely to open its next application cycle in the next few months, according to GSA communications manager Jacqueline Treboschi.

The proliferation of preprints, too, provides opportunities to practise reviewing. Researchers with expertise in various fields related to the COVID-19 pandemic, from vaccinology and statistics to history, can

volunteer with *Rapid Reviews: COVID-19*, published by MIT Press in Cambridge, Massachusetts, and edited by a team at the University of California, Berkeley. The project solicits quick-turnaround reviews for COVID-19 articles on preprint servers, such as medRxiv and bioRxiv, that are deemed by the network of reviewers to be important.

Fortunately, peer-review skills are easy to build. Ahmedien says junior researchers are likely to get a feel for the process after completing three or four reviews.

But once researchers get into the reviewer pool, another problem can arise: a deluge of requests. When Sugimoto was an assistant professor, she reviewed about one paper a month, but as her career advanced, the rate of invitations increased. “It’s not uncommon for me to get hundreds of requests per year,” says Sugimoto, now a tenured faculty member. “At a certain point, you just start saying no.”

To avoid burnout, “researchers should try to set realistic goals of how many manuscripts they can review,” says Julia Vilstrup Mouatt, head of the Web of Science Academy, who is based in Auckland, New Zealand. As a rule of thumb, some researchers try to review three papers for every one they publish.

Hazards of reviewing

Indeed, there are times when it’s best to decline or withdraw from a review opportunity — if a researcher lacks relevant expertise, for example.

Avoiding conflicts of interest is also key. Such a conflict would arise if the would-be reviewer had collaborated with or worked in the same department as any author in recent years, or if the reviewer would receive financial or other benefits from the paper or review. Researchers should also avoid reviewing papers authored by friends or family members. “My colleagues, my friends, a deep personal relationship — this is a conflict of interest,” says Ahmedien. A more casual acquaintance, however, is not.

Mirjalili says it’s not acceptable to contact authors during the review process, even to ask questions or clarify a point. Instead, he says, “Get in

touch with the journal”. Editors will want to track queries, and direct discussions between reviewers and authors could break confidentiality and make the review unusable.

Confidentiality policies vary by journal. The advantage of confidentiality, says Rousk, is that reviewers will be less biased by any previous knowledge of that author, their demographic or geographical location.

Aiming to improve transparency, some journals might identify the authors and reviewers to each other, or even publish reviews alongside the final paper. Such ‘open review’ has been growing in popularity over the past five years, with nearly 80% of medical and scientific journals using it at least occasionally¹¹. One advantage is that open review is expected to make referees consider the scientific issues at hand more carefully. Open reviews tend to be shorter and nicer, says Sugimoto.

Most often, says Schekman, the reviewer will know the authors’ identities, but the authors will not know the reviewers’ — although, he adds, “they usually try to guess”.

Nature reviewers are told the names of manuscript authors, and the journal acknowledges reviewers by name in the published paper with the reviewer’s consent. In February 2020, *Nature* also began offering authors the option to [publish anonymous reviews and author responses alongside papers](#).

For junior reviewers, having their name linked to a review can boost visibility. But there can be downsides. A senior scientist might take criticism poorly, and in some cases could even later damage the early-career scientist’s chances of employment, promotion or publishing their own work. “That is something that worries me intensely about the move towards open peer review,” says Sugimoto.

One option is not to take on a review under such open conditions. And it’s fine to withdraw after accepting a review offer, if issues arise. Reviewers can contact the journal editor, outside of their formal review, to share any concerns about the paper or their role in the process, says McDowell. After all, he says, editors should be eager to hear about any issues: they don’t want to approve a paper that will be panned on social media or even retracted.

That quality control, after all, is the point of the peer-review enterprise. “We do it,” says Lijek, “because we want science to be better.”

Nature **602**, 169-171 (2022)

doi: <https://doi.org/10.1038/d41586-022-00216-1>

References

1. 1.

Kovanis, M., Porcher, R., Ravaud, P. & Trinquart, L. *PLoS ONE* **11**, e0166387 (2016).

2. 2.

Publons & Clarivate Analytics. *2018: Global State of Peer Review* (Publons & Clarivate Analytics, 2018).

3. 3.

Helmer, M., Schottdorf, M., Neef, A. & Battaglia, D. *eLife* **6**, e21718 (2017).

4. 4.

Murray, D. *et al.* Preprint at bioRxiv <https://doi.org/10.1101/400515> (2019).

5. 5.

Squazzoni, F. *et al.* *Sci. Adv.* **7**, eabd0299 (2021).

6. 6.

Silbiger, N. J. & Stubler, A. D. *PeerJ* **7**, e8247 (2019).

7. 7.

McDowell, G. S., Knutsen, J. D., Graham, J. M., Oelker, S. K. & Lijek, R. S. *eLife* **8**, e48425 (2019).

8. 8.

Evans, A. T., McNutt, R. A., Fletcher, S. W. & Fletcher, R. H. *J. Gen. Intern. Med.* **8**, 422–428 (1993).

9. 9.

Black, N., van Rooyen, S., Godlee, F., Smith, R. & Evans, S. *JAMA* **280**, 231–233 (1998).

10. 10.

Kliewer, M. A., Freed, K. S., DeLong, D. M., Pickhardt, P. J. & Provenzale, J. M. *Am. J. Roentgenol.* **184**, 1731–1735 (2005).

11. 11.

Wolfram, D., Wang, P., Hembree, A. & Park, H. *Scientometrics* **125**, 1033–1051 (2020).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00216-1>

- TECHNOLOGY FEATURE
- 31 January 2022

How to fix your scientific coding errors

Software bugs are frustrating. Adopting some simple strategies can help you to avoid them, and fix them when they occur.

- [Jeffrey M. Perkel](#)

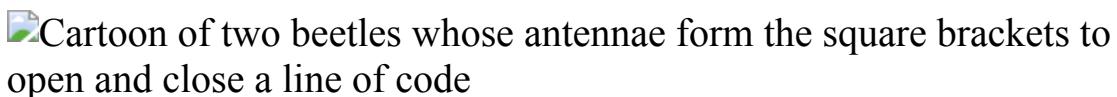


Illustration by The Project Twins

As a graduate student, Steven Weisberg helped to develop a university campus — albeit, a virtual one. Called Virtual Slichton, the software tests spatial navigation skills, teaching people the layout of a virtual campus and then challenging them to point in the direction of specific landmarks¹. It has been used by more than a dozen laboratories, says Weisberg, who is now a cognitive neuroscientist at the University of Florida in Gainesville.

But in February 2020, a colleague who was testing the software identified a problem: it couldn't compute your direction accurately if you were pointing more than 90 degrees from the site. "The first thing I thought was, 'oh, that's weird,'" Weisberg recalls. But it was true: his software was generating errors that could alter its calculations and conclusions.

"We have to retract everything," he thought.

When it comes to software, bugs are inevitable — especially in academia, where code tends to be written by graduate students and postdocs who were

never trained in software development. But simple strategies can minimize the likelihood of a bug, and ease the process of recovering from them.

Avoidance

Julia Strand, a psychologist at Carleton College in Northfield, Minnesota, investigates strategies to help people to engage in conversation in, for example, a noisy, crowded restaurant. In 2018, she reported that a visual cue, such as a blinking dot on a computer screen that coincided with speech, reduced the cognitive effort required to understand what was being said². That suggested that a simple smartphone app could reduce the mental fatigue that sometimes arises in such situations.

But it wasn't true. Strand had inadvertently programmed the testing software to start timing one condition earlier than the other, which, as she wrote in 2020, "is akin to starting a stopwatch before a runner gets to the line".

"I felt physically ill," she wrote — the mistake could have negatively affected her students, her collaborators, her funding and her job. It didn't — she corrected her article, kept her grants and received tenure. But to help others avoid a similar experience, she has created a teaching resource called [Error Tight](#)³.

Error Tight provides practical tips that echo computational reproducibility checklists, such as; use version control; document code and workflows; and adopt standardized file naming and organizational strategies.

Its other recommendations are more philosophical. An 'error tight' laboratory, Strand says, recognizes that even careful researchers make mistakes. As a result, her team adopted a strategy that is common in professional software development: code review. The team proactively looks for bugs by having two people review their work, rather than assuming those bugs don't exist.

Joana Grave, a psychology PhD student at the University of Aveiro, Portugal, also uses code review. In 2021, [Grave retracted a study](#) when she

discovered that the tests she had programmed had been miscoded to show the wrong images. Now, experienced programmers on the team double-check her work, she says, and Grave repeats coding tasks to ensure she gets the same answer.

Scientific software can be difficult to review, warns C. Titus Brown, a bioinformatician at the University of California, Davis. “If we’re operating at the ragged edge of novelty, there may only be one person that understands the code, and it may take a lot of time for another person to understand it. And even then, they may not be asking the right questions.”

Weisberg shared other helpful practices in a Twitter thread about his experience. These include sharing code, data and computational environments on sites such as GitHub and Binder; ensuring computational results dovetail with evidence collected using different methods; and adopting widely used software libraries in lieu of custom algorithms when possible, as these are often extensively tested by the scientific community.

Whatever the origin of your code, validate it before using it — and then again periodically, for instance after upgrading your operating system, advises Philip Williams, a natural-products chemist at the University of Hawaii at Manoa in Honolulu. “If anything changes, the best practice is to go back and just make sure everything’s OK, rather than just assume that these black boxes will always turn out the correct answer,” he says.

Williams and his colleagues identified what they called a ‘glitch’ in another researcher’s published code for interpreting nuclear magnetic resonance data⁴, which resulted in data sets being sorted differently depending on the user’s operating system. Checking their numbers against a model data set with known ‘correct’ answers, could have alerted them that the code wasn’t working as expected, he says.

Recovery

If code cannot be bug-free, it can at least be developed so that any bugs are relatively easy to find. Lorena Barba, a mechanical and aerospace engineer at George Washington University in Washington DC, says that when she

and her then graduate student Natalia Clementi discovered a mistake in code underlying a study⁵ they had published in 2019, “there were some poop emojis being sent by Slack and all sorts of scream emojis and things for a few hours”. But the pair were able to quickly resolve their problem, thanks to the [reproducibility packages](#) (known as repro-packs) that Barba’s lab makes for all their published work.

A repro-pack is an open-access archive of all the scripts, data sets and configuration files required to perform an analysis and reproduce the results published in a paper, which Barba’s team uploads to open-access repositories such as Zenodo and Figshare. Once they realized that their code contained an error — they had accidentally omitted a mathematical term in one of their equations — Clementi retrieved the relevant repro-pack, fixed the code, reran her computations and compared the results. Without a repro-pack, she would have had to remember exactly how those data were processed. “It probably would have taken me months to try to see if this [code] was correct or not,” she says. Instead, it took just two days.

Brown needed significantly more time to resolve a bug he discovered in 2020 when attempting to apply his lab’s metagenome-search tool, called spacegraphcats, towards a new question. The software contained a bad filtering step, which removed some data from consideration. “I started to think, ‘oh dear, this maybe calls into question the original publication’,” he deadpans. Brown fixed the software in less than two weeks. But re-running the computations set the project back by several months.

To minimize delays, good documentation is crucial. Milan Curcic, an oceanographer at the University of Miami, Florida, co-authored a 2020 study⁶ that investigated the impact of hurricane wind speed on ocean waves. As part of that work, Curcic and his colleagues repeated calculations that had been conducted in the same lab in 2004, only to discover that the original code was using the wrong data file to perform some of its calculations, producing an “offset” of about 30%.

According to Google Scholar, the 2004 study⁷ has been cited more than 800 times, and its predictions inform hurricane forecasts today, Curcic says. Yet its code, written in the programming language MATLAB, was never placed

online. And it was so poorly documented that Curcic had to work through it line by line to understand how it worked. When he found the error, he says, “The question was, am I not understanding this correctly, or is this indeed incorrect?”

Strand has team members read each others’ code to familiarize them with programming and encourage good documentation. “Code should be clearly commented enough that even someone who doesn’t know how to code can understand what’s happening and how the data are changing at each step,” she says.

And she encourages students to view errors as part of science rather than personal failings. “Labs that have a culture of ‘people who are smart and careful don’t make mistakes’, are setting themselves up for being a lab that doesn’t admit their mistakes,” she says.

Bugs don’t necessarily mean retraction in any event. Barba, Brown and Weisberg’s errors had only minor impacts on their results, and none required changes to their publications. In 2016, Marcos Gallego Llorente, then a genetics graduate student at the University of Cambridge, UK, identified an error in the code he wrote to study human migratory patterns in Africa 4,500 years ago. When he reanalysed the data, the overall conclusion was unchanged, although the extent of its geographic impact was, and a correction sufficed.

Thomas Hoye, an organic chemist at the University of Minnesota at Minneapolis, co-authored a study that used the software in which Williams discovered a bug. When Williams contacted him, Hoye says, he didn’t have “any particular strong reaction”. He and his colleagues fixed their code, updated their online protocols, and moved on.

“I couldn’t help but at the end think, ‘this is the way science should work’,” he says. “You find a mistake, you go back, you improve, you correct, you advance.”

Nature **602**, 172-173 (2022)

doi: <https://doi.org/10.1038/d41586-022-00217-0>

References

1. 1.

Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F. & Epstein, R. A. *J. Exp. Psychol. Learn. Mem. Cogn.* **40**, 669–682 (2014).

2. 2.

Strand, J. F., Brown, V. A. & Barbour, D. L. *Psychon. Bull. Rev.* **26**, 291–297 (2019).

3. 3.

Strand, J. F. Preprint at PsyArXiv <https://doi.org/10.31234/osf.io/rsn5y> (2021).

4. 4.

Neupane, J. B. *et al. Org. Lett.* **21**, 8449–8453 (2019).

5. 5.

Clementi, N. C., Cooper, C. D. & Barba, L. A. *Phys. Rev. E* **100**, 063305 (2019).

6. 6.

Curcic, M. & Haus, B. K. *Geophys. Res. Lett.* **47**, e2020GL087647 (2020).

7. 7.

Donelan, M. A. *et al. Geophys. Res. Lett.* **31**, L18306 (2004).

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

- WHERE I WORK
- 31 January 2022

Grape expectations: making Australian wine more sustainable

Geoff Gurr is working on an eco-friendly approach to controlling pests in New South Wales vineyards.

- [Benjamin Plackett](#) ⁰



Geoff Gurr is a professor of applied ecology at Charles Sturt University in Orange, New South Wales. Matthew Abbott for *Nature*

This photograph was taken at the Angullong estate in New South Wales, Australia, which hosts some of my field trials. The aim is to study sustainable agriculture in vineyards. You have to dodge the odd brown snake, but, as offices go, this one — among the grapevines of such a picturesque part of the world — makes my job quite a privilege.

It's a November evening, which is springtime here in the Southern Hemisphere, and this time of year is when pests such as the light brown apple moth (*Epiphyas postvittana*) start to emerge. That means that ecologists such as myself, as well as the commercial winemakers we collaborate with, move into data-capture mode to track the presence of the insects. These moths produce multiple generations every year, so they can be quite numerous by harvest time, and can cause real damage by getting into the grapes.

We're conducting experiments to see whether positioning various plant species between and under grapevines can help to reduce the population of

pests by encouraging their predators. Parasitoid wasps, for example, target the eggs of light brown apple moths, injecting them with their own eggs. When the wasp larvae hatch, they eat the moth larvae from the inside out. Although quite gruesome, parasitoid wasps could provide an environmentally friendly way to control moth populations.

In my laboratory at Charles Sturt University in Orange, we're incubating moth eggs that we then put on special cards in the vineyard. Because parasitoids love nectar, we expect to see more attacks on the moth eggs in areas where we've planted flowering shrubs than in the control areas, where grass predominates. We collect the cards after about 48 hours in the field, and incubate the moth eggs to measure the level of parasitism. In the next couple of years, with more data, we hope to identify the optimum mix of plant species to manage pests without resorting to chemicals.

Nature **602**, 176 (2022)

doi: <https://doi.org/10.1038/d41586-022-00218-z>

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00218-z>

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

Research

- **[A subset of immune-system T cells branded as seeds for type 1 diabetes](#)** [25 January 2022]

News & Views • The identification of a specific subpopulation of immune-system T cells that drives type 1 diabetes provides insight into the development of autoimmune disease, and could point the way to new therapies.

- **[Lost magnetism pinned on atomic rotations](#)** [02 February 2022]

News & Views • Crystal-lattice vibrations reveal the mechanism by which laser pulses can strip a metal of its magnetism. The vibrations absorb the angular momentum of electrons in a sample, allowing it to demagnetize.

- **[Important genomic regions mutate less often than do other regions](#)** [12 January 2022]

News & Views • Genomic regions that are crucial for the viability and reproduction of the model plant *Arabidopsis thaliana* are enriched with molecular features that are associated with a reduced rate of mutation.

- **[Search for single magnetic charges in the largest of fields](#)** [02 February 2022]

News & Views • Collisions between lead ions have produced the largest measured magnetic field in the Universe, enabling a search for elusive exotic particles that carry an isolated magnetic charge.

- **[Reproducibility in the fabrication and physics of moiré materials](#)** [02 February 2022]

Review Article • The essential properties of moiré materials and the progress and latest developments in the field are reviewed, and their fabrication and physics are discussed from a reproducibility perspective.

- **[An open science study of ageing in companion dogs](#)** [02 February 2022]

Perspective • The Dog Aging Project is an open-data, community science study to identify genetic, environmental and lifestyle factors associated with canine healthy lifespan, generating knowledge that could readily translate to human ageing.

- **[Microwave background temperature at a redshift of 6.34 from H₂O absorption](#)** [02 February 2022]

Article • Measurement of the cosmic microwave background temperature using H₂O absorption at a redshift of 6.34 is reported, the results of which were consistent with those from standard ΛCDM cosmology.

- **Search for magnetic monopoles produced via the Schwinger mechanism** [02 February 2022]

Article • At the Large Hadron Collider, the MoEDAL experiment shows no evidence for magnetic monopoles generated via the Schwinger mechanism at integer Dirac charges below 3, and suggests a lower mass limit of 75 GeV/c².

- **Domain-wall dynamics in Bose–Einstein condensates with synthetic gauge fields** [02 February 2022]

Article • The creation of density-dependent gauge fields in an ultracold atomic system is described, leading to formation of domain walls that behave as elementary excitations and respond to electric fields, enabling a path to simulating gauge physics.

- **Polarized phonons carry angular momentum in ultrafast demagnetization** [02 February 2022]

Article • Ultrafast electron diffraction is used here to reveal in nickel an almost instantaneous, long-lasting population of anisotropic phonons with angular momentum.

- **Selective sulfidation of metal compounds** [16 December 2021]

Article • A framework for selective sulfidation is proposed for the practical separation and enrichment of numerous metallic elements, to enable environmentally and economically sustainable metal processing.

- **Fluctuation-induced quantum friction in nanoscale water flows** [02 February 2022]

Article • The quantum contribution to friction enables the rationalization of the peculiar friction properties of water on carbon surfaces, and in particular the radius dependence of slippage in carbon nanotubes.

- **Irreversible synthesis of an ultrastrong two-dimensional polymeric material** [02 February 2022]

Article • The solution-phase irreversible synthesis of a two-dimensional polymer with excellent elastic modulus and yield strength is reported.

- **Novel flight style and light wings boost flight performance of tiny beetles** [19 January 2022]

Article • Three-dimensional reconstructions of morphology and flight mechanics of the beetle Paratuposa placentis reveal adaptations that enable extremely small insects to fly at speeds similar to those of much larger insects.

- **Mutation bias reflects natural selection in *Arabidopsis thaliana*** [12 January 2022]

Article • Data on de novo mutations in *Arabidopsis thaliana* reveal that mutations do not occur randomly; instead, epigenome-associated mutation bias reduces the occurrence of deleterious mutations.
- **Malaria protection due to sickle haemoglobin depends on parasite genotype** [09 December 2021]

Article • A strong association has been found between three regions of the *Plasmodium falciparum* genome and sickle haemoglobin in children with severe malaria, suggesting parasites have adapted to overcome natural host immunity.
- **Androgens increase excitatory neurogenic potential in human brain organoids** [19 January 2022]

Article • Experiments in cerebral organoids show that sex hormones have a role in regulating the number of excitatory neurons in the human neocortex, providing insight into the mechanistic basis of sex-related brain differences in humans.
- **A speech planning network for interactive language use** [05 January 2022]

Article • Using intracranial electrocorticography and a series of motor tasks, a speech planning network that is central to natural language generation during social interaction is identified.
- **Toroidal topology of population activity in grid cells** [12 January 2022]

Article • Simultaneous recordings from hundreds of grid cells in rats, combined with topological data analysis, show that network activity in grid cells resides on a toroidal manifold that is invariant across environments and brain states.
- **Brahma safeguards canalization of cardiac mesoderm differentiation** [26 January 2022]

Article • The BAF chromatin-remodelling complex ATPase gene *Brm* safeguards cell identity during directed cardiogenesis of mouse embryonic stem cells.
- **Emergence of methicillin resistance predates the clinical use of antibiotics** [05 January 2022]

Article • Methicillin-resistant strains of *Staphylococcus aureus* appeared in European hedgehogs in the pre-antibiotic era as a co-evolutionary adaptation to antibiotic-producing dermatophytes and have spread within the local hedgehog populations and between hedgehogs and secondary hosts.
- **Petabase-scale sequence alignment catalyses viral discovery** [26 January 2022]

Article • Serratus, an open-source cloud-computing infrastructure, can be used to screen millions of nucleic acid sequencing libraries at the petabase scale, and has enabled many new

RNA viruses to be identified efficiently.

- **Signature of long-lived memory CD8+ T cells in acute SARS-CoV-2 infection** [07 December 2021]
Article • Evidence of a transcriptional signature that marks precursors of long-lived CD8+ memory T cells in SARS-CoV-2 infection.
- **An autoimmune stem-like CD8 T cell population drives type 1 diabetes** [30 November 2021]
Article • A population of β-cell-specific autoimmune stem-like CD8 T cells initiates and sustains β-cell destruction and disease in a mouse model of type 1 diabetes.
- **Life histories of myeloproliferative neoplasms inferred from phylogenies** [20 January 2022]
Article • Whole-genome sequencing of 1,013 clonal haematopoietic colonies from myeloproliferative neoplasms of 12 individuals reveals haematopoietic phylogenies and indicates that driver mutations are acquired sequentially, starting early in life.
- **No freshwater-filled glacial Arctic Ocean** [02 February 2022]
Matters Arising •
- **Reply to: No freshwater-filled glacial Arctic Ocean** [02 February 2022]
Matters Arising •
- **Rainfall an unlikely factor in Kīlauea’s 2018 rift eruption** [02 February 2022]
Matters Arising •
- **Reply to: Rainfall an unlikely factor in Kīlauea’s 2018 rift eruption** [02 February 2022]
Matters Arising •

- NEWS AND VIEWS
- 25 January 2022

A subset of immune-system T cells branded as seeds for type 1 diabetes

The identification of a specific subpopulation of immune-system T cells that drives type 1 diabetes provides insight into the development of autoimmune disease, and could point the way to new therapies.

- [Stephen J. Turner](#) ⁰ &
- [Nicole L. La Gruta](#) ¹

A characteristic of many autoimmune diseases, including type 1 diabetes, is an immune response that persists and does not resolve. [Writing in Nature](#), Gearty *et al.*¹ pinpoint a specific group of T cells in the immune system that sustain this autoimmune response. The finding has implications for treatments that seek to modulate immune responses.

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 **602**, 35-36 (2022)

doi: <https://doi.org/10.1038/d41586-021-03800-z>

References

1. 1.

Gearty, S. V. *et al.* *Nature* **602**, 156–161 (2022).

2. 2.

Khan, O. *et al.* *Nature* **571**, 211–218 (2019).

3. 3.

McLane, L. M., Abdel-Hakeem, M. S. & Wherry, E. J. *Annu. Rev. Immunol.* **37**, 457–495 (2019).

4. 4.

Yao, C. *et al.* *Nature Immunol.* **20**, 890–901 (2019).

5. 5.

Bluestone, J. A., Herold, K. & Eisenbarth, G. *Nature* **464**, 1293–1300 (2010).

6. 6.

Lin, W.-H. W. *et al.* *Cell Rep.* **17**, 1773–1782 (2016).

7. 7.

Zhou, X. *et al.* *Immunity* **33**, 229–240 (2010).

8. 8.

Chen, Z. *et al.* *Immunity* **51**, 840–855 (2019).

9. 9.

Connolly, K. A. *et al.* *Sci. Immunol.* **6**, eabg7836 (2021).

10. 10.

Utzschneider, D. T. *et al.* *Immunity* **45**, 415–427 (2016).

11. 11.

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-021-03800-z>

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

- NEWS AND VIEWS
- 02 February 2022

Lost magnetism pinned on atomic rotations

Crystal-lattice vibrations reveal the mechanism by which laser pulses can strip a metal of its magnetism. The vibrations absorb the angular momentum of electrons in a sample, allowing it to demagnetize.

- [Georg Woltersdorf](#) 

More than two decades ago, physicists noticed a curious thing about thin magnetic films of nickel: their magnetization was reduced almost immediately when the metal was hit with extremely short, intense pulses of light¹. In general, materials are magnetic when the intrinsic angular momentum (or spin) of each electron aligns parallel to that of its neighbours. Because electrons carry an electric charge, this angular momentum gives rise to a magnetic moment similar to the one that arises when an electric current flows through a wire loop. So when the nickel demagnetized rapidly, researchers were understandably puzzled about what happened to the angular momentum. [Writing in Nature](#), Tauchert *et al.*² have found the answer to this long-standing question, showing that the vibrational modes of nickel's crystal lattice carry a large fraction of the lost angular momentum.

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 **602**, 36-38 (2022)

doi: <https://doi.org/10.1038/d41586-022-00189-1>

References

1. 1.
Beaurepaire, E., Merle, J.-C., Daunois, A. & Bigot, J.-Y. *Phys. Rev. Lett.* **76**, 4250–4253 (1996).
2. 2.
Tauchert, S. R. *et al.* *Nature* **602**, 73–77 (2022).
3. 3.
Stamm, C. *et al.* *Nature Mater.* **6**, 740–743 (2007).
4. 4.
Koopmans, B. *et al.* *Nature Mater.* **9**, 259–265 (2010).
5. 5.
Eschenlohr, A. *et al.* *Nature Mater.* **12**, 332–336 (2013).
6. 6.
Rudolf, D. *et al.* *Nature Commun.* **3**, 1037 (2012).
7. 7.
Einstein, A. & de Haas, W. J. *Proc. Sect. Sci. K. Ned. Akad. Wet. Amst.* **18**, 696–711 (1915).
8. 8.
Dornes, C. *et al.* *Nature* **565**, 209–212 (2019).
9. 9.
Radu, I. *et al.* *Nature* **472**, 205–208 (2011).

10. 10.

Seifert, T. *et al.* *Nature Photon.* **10**, 483–488 (2016).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00189-1>

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

- NEWS AND VIEWS
- 12 January 2022

Important genomic regions mutate less often than do other regions

Genomic regions that are crucial for the viability and reproduction of the model plant *Arabidopsis thaliana* are enriched with molecular features that are associated with a reduced rate of mutation.

- [Jianzhi Zhang](#) ⁰

Salvador Luria and Max Delbrück made a profound discovery in 1943 that won them a Nobel prize, shared with Alfred Hershey, 26 years later. What they found was that bacterial mutations that confer resistance to a virus arise at the same rate, regardless of whether the virus is present¹. That the generation of mutations (a process called mutagenesis) is blind to its consequence has since become an established principle of genetics. [Writing in Nature](#), Monroe *et al.*² report that, in stark contrast to this tenet, the rate of mutation in the model plant *Arabidopsis thaliana* is lower in genomic regions that are functionally more important, and in regions where mutations are more frequently harmful.

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 **602**, 38-39 (2022)

doi: <https://doi.org/10.1038/d41586-022-00017-6>

References

1. 1.

Luria, S. E. & Delbrück, M. *Genetics* **28**, 491–511 (1943).

2. 2.

Monroe, J. G. *et al.* *Nature* **602**, 101–105 (2022).

3. 3.

Martincorena, I., Seshasayee, A. S. N. & Luscombe, N. M. *Nature* **485**, 95–98 (2012).

4. 4.

Zhang, J. & Yang, J. R. *Nature Rev. Genet.* **16**, 409–420 (2015).

5. 5.

Xia, B. *et al.* *Cell* **180**, 248–262 (2020).

6. 6.

Chen, X. & Zhang, J. *Mol. Biol. Evol.* **30**, 1559–1562 (2013).

7. 7.

Maddamsetti, R. *et al.* *Mol. Biol. Evol.* **32**, 2897–2904 (2015).

8. 8.

Liu, H. & Zhang, J. *Mol. Biol. Evol.* **37**, 3225–3231 (2020).

9. 9.

Kimura, M. *Genet. Res.* **9**, 23–34 (1967).

10. 10.

Lynch, M. *Genome Biol. Evol.* **3**, 1107–1118 (2011).

11. 11.

Chen, X., Yang, J.-R. & Zhang, J. *Genome Res.* **26**, 50–59 (2016).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00017-6>

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

- NEWS AND VIEWS
- 02 February 2022

Search for single magnetic charges in the largest of fields

Collisions between lead ions have produced the largest measured magnetic field in the Universe, enabling a search for elusive exotic particles that carry an isolated magnetic charge.

- [Sonia Kabana](#) 0

Break a magnet into smaller pieces, and each part will have its own north and south pole. But in the subatomic realm, an exotic particle called a magnetic monopole can possess an isolated magnetic charge — existing as only a north pole or a south pole. [Writing in Nature](#), Acharya *et al.*¹ report the results of a search for these extraordinary particles using a very strong magnetic field.

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 **602**, 39-40 (2022)

doi: <https://doi.org/10.1038/d41586-022-00188-2>

References

1. 1.

Acharya, B. *et al.* *Nature* **602**, 63–67 (2022).

2. 2.

Curie, P. *Seanc. Soc. Fr. Phys.* **76**, 1 (1894).

3. 3.

Dirac, P. A. M. *Proc. R. Soc. A* **133**, 60–72 (1931).

4. 4.

Schwinger, J. *Phys. Rev.* **82**, 664–679 (1951).

5. 5.

Mavromatos, N. E. & Mitsou, V. A. *Int. J. Mod. Phys. A* **35**, 2030012 (2020).

6. 6.

Huang, X.-G. *Rep. Prog. Phys.* **79**, 076302 (2016).

This article was downloaded by **calibre** from <https://www.nature.com/articles/d41586-022-00188-2>

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

- Review Article
- [Published: 02 February 2022](#)

Reproducibility in the fabrication and physics of moiré materials

- [Chun Ning Lau](#) [ORCID: orcid.org/0000-0003-2159-6723¹](#),
- [Marc W. Bockrath¹](#),
- [Kin Fai Mak](#) [ORCID: orcid.org/0000-0002-5768-199X²](#) &
- [Fan Zhang](#) [ORCID: orcid.org/0000-0003-4623-4200³](#)

[Nature](#) volume **602**, pages 41–50 (2022)

- 1720 Accesses
- 12 Altmetric
- [Metrics details](#)

Subjects

- [Electronic properties and materials](#)
- [Two-dimensional materials](#)

Abstract

Overlaying two atomic layers with a slight lattice mismatch or at a small rotation angle creates a moiré superlattice, which has properties that are markedly modified from (and at times entirely absent in) the ‘parent’ materials. Such moiré materials have progressed the study and engineering of strongly correlated phenomena and topological systems in reduced dimensions. The fundamental understanding of the electronic phases, such

as superconductivity, requires a precise control of the challenging fabrication process, involving the rotational alignment of two atomically thin layers with an angular precision below 0.1 degrees. Here we review the essential properties of moiré materials and discuss their fabrication and physics from a reproducibility perspective.

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: Various factors and phenomena affecting the reproducibility of moiré materials studies.

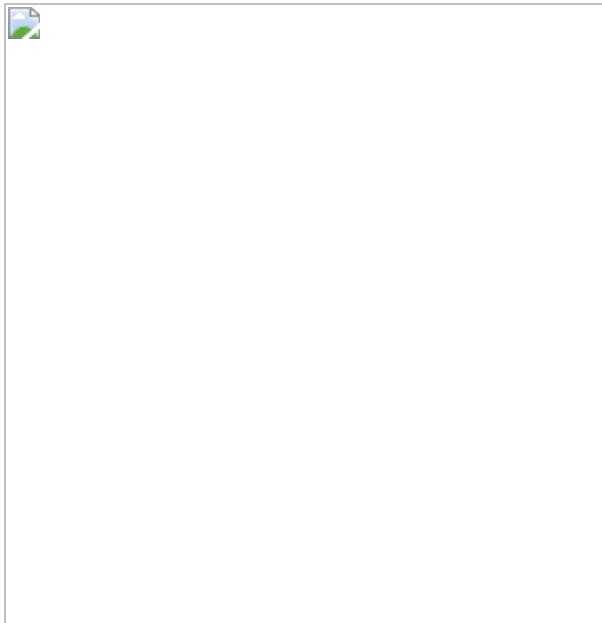
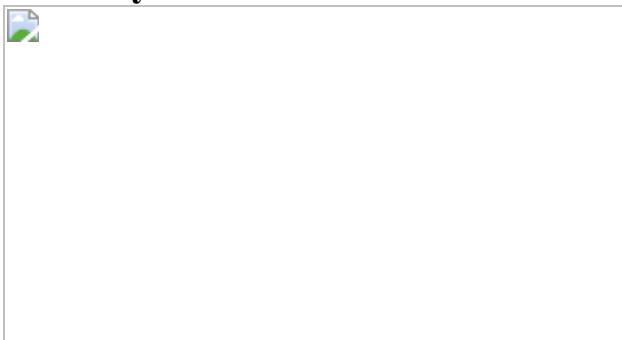


Fig. 2: Fabrication of twisted bilayer graphene and various disorders in moiré systems.



Fig. 3: Schematics of various electronic phases and the systems in which they are observed.



References

1. 1.

Cao, Y. et al. Correlated insulator behaviour at half-filling in magic-angle graphene superlattices. *Nature* **556**, 80–84 (2018).

2. 2.

Cao, Y. et al. Unconventional superconductivity in magic-angle graphene superlattices. *Nature* **556**, 43–50 (2018).

References ^{1,2} report the observation of a correlated insulating state and superconductivity in twisted bilayer graphene.

3. 3.

Liu, X. et al. Tunable spin-polarized correlated states in twisted double bilayer graphene. *Nature* **583**, 221–225 (2020).

4. 4.

Burg, G. W. et al. Correlated insulating states in twisted double bilayer graphene. *Phys. Rev. Lett.* **123**, 197702 (2019).

5. 5.

He, M. et al. Symmetry breaking in twisted double bilayer graphene. *Nat. Phys.* **17**, 26–30 (2021).

6. 6.

Shen, C. et al. Correlated states in twisted double bilayer graphene. *Nat. Phys.* **16**, 520–525 (2020).

7. 7.

Cao, Y. et al. Tunable correlated states and spin-polarized phases in twisted bilayer–bilayer graphene. *Nature* **583**, 215–220 (2020).

8. 8.

Chen, S. et al. Electrically tunable correlated and topological states in twisted monolayer–bilayer graphene. *Nat. Phys.* **17**, 374–380 (2020).

9. 9.

Xu, S. et al. Tunable van Hove singularities and correlated states in twisted monolayer–bilayer graphene. *Nat. Phys.* **17**, 619–626 (2021).

10. 10.

Polshyn, H. et al. Electrical switching of magnetic order in an orbital Chern insulator. *Nature* **588**, 66–70 (2020).

11. 11.

Wu, F., Lovorn, T., Tutuc, E., Martin, I. & MacDonald, A. H. Topological insulators in twisted transition metal dichalcogenide homobilayers. *Phys. Rev. Lett.* **122**, 086402 (2019).

12. 12.

Shimazaki, Y. et al. Strongly correlated electrons and hybrid excitons in a moiré heterostructure. *Nature* **580**, 472–477 (2020).

13. 13.

Wang, L. et al. Correlated electronic phases in twisted bilayer transition metal dichalcogenides. *Nat. Mater.* **19**, 861–866 (2020).

14. 14.

Wu, F., Lovorn, T., Tutuc, E. & MacDonald, A. H. Hubbard model physics in transition metal dichalcogenide moiré bands. *Phys. Rev. Lett.* **121**, 026402 (2018). **This reference shows that the flat bands in moiré TMD structures can be described by generalized Hubbard models, and that a number of many-body ground states are possible.**

15. 15.

Regan, E. C. et al. Mott and generalized Wigner crystal states in WSe₂/WS₂ moiré superlattices. *Nature* **579**, 359–363 (2020).

16. 16.

Tang, Y. et al. Simulation of Hubbard model physics in WSe₂/WS₂ moiré superlattices. *Nature* **579**, 353–358 (2020). **References 15,16 report a Mott insulator in WSe₂/WS₂ bilayer superlattices at half filling, and generalized Wigner crystal and ferromagnetic states at fractional fillings.**

17. 17.

Xu, Y. et al. Correlated insulating states at fractional fillings of moiré superlattices. *Nature* **587**, 214–218 (2020).

18. 18.

Chen, G. et al. Evidence of a gate-tunable Mott insulator in a trilayer graphene moiré superlattice. *Nat. Phys.* **15**, 237–241 (2019). **This reference reports a gate-tunable insulator state in an r-TLG/hBN moiré superlattice at half filling.**

19. 19.

Chen, G. et al. Tunable correlated Chern insulator and ferromagnetism in a moiré superlattice. *Nature* **579**, 56–61 (2020).

20. 20.

Chen, G. et al. Signatures of tunable superconductivity in a trilayer graphene moiré superlattice. *Nature* **572**, 215–219 (2019).

21. 21.

Park, J. M., Cao, Y., Watanabe, K., Taniguchi, T. & Jarillo-Herrero, P. Tunable strongly coupled superconductivity in magic-angle twisted trilayer graphene. *Nature* **590**, 249–255 (2021).

22. 22.

Hao, Z. et al. Electric field–tunable superconductivity in alternating-twist magic-angle trilayer graphene. *Science* **371**, 1133–1138 (2021).

23. 23.

Hofstadter, D. R. Energy levels and wave functions of Bloch electrons in rational and irrational magnetic fields. *Phys. Rev. B* **14**, 2239–2249 (1976).

24. 24.

Yankowitz, M. et al. Emergence of superlattice Dirac points in graphene on hexagonal boron nitride. *Nat. Phys.* **8**, 382–386 (2012).

25. 25.

Ponomarenko, L. A. et al. Cloning of Dirac fermions in graphene superlattices. *Nature* **497**, 594–597 (2013).

26. 26.

Dean, C. R. et al. Hofstadter’s butterfly and the fractal quantum Hall effect in moire superlattices. *Nature* **497**, 598–602 (2013).

27. 27.

Hunt, B. et al. Massive Dirac fermions and Hofstadter butterfly in a van der Waals heterostructure. *Science* **340**, 1427–1430 (2013).

28. 28.

Wang, P. et al. Topological winding number change and broken inversion symmetry in a Hofstadter’s butterfly. *Nano Lett.* **15**, 6395–6399 (2015).

29. 29.

Kim, K. et al. Tunable moiré bands and strong correlations in small-twist-angle bilayer graphene. *Proc. Natl Acad. Sci. USA* **114**, 3364–3369 (2017).

30. 30.

Kim, K. et al. Van der Waals heterostructures with high accuracy rotational alignment. *Nano Lett.* **16**, 1989–1995 (2016).

31. 31.

Cao, Y. et al. Superlattice-induced insulating states and valley-protected orbits in twisted bilayer graphene. *Phys. Rev. Lett.* **117**, 116804 (2016).

32. 32.

Saito, Y., Ge, J., Watanabe, K., Taniguchi, T. & Young, A. F. Independent superconductors and correlated insulators in twisted bilayer graphene. *Nat. Phys.* **16**, 926–930 (2020).

33. 33.

Chen, X.-D. et al. High-precision twist-controlled bilayer and trilayer graphene. *Adv. Mater.* **28**, 2563–2570 (2016).

34. 34.

Pizzocchero, F. et al. The hot pick-up technique for batch assembly of van der Waals heterostructures. *Nat. Commun.* **7**, 11894 (2016).

35. 35.

Borodin, B. R., Benimetskiy, F. A. & Alekseev, P. A. Study of local anodic oxidation regimes in MoSe₂. *Nanotechnology* **32**, 155304 (2021).

36. 36.

Zhou, H. et al. Half and quarter metals in rhombohedral trilayer graphene. Preprint at <https://arxiv.org/abs/2104.00653> (2021).

37. 37.

Rosenberger, M. R. et al. Nano-“squeegee” for the creation of clean 2D material interfaces. *ACS Appl. Mater. Interfaces* **10**, 10379–10387 (2018).

38. 38.

Wang, D. et al. Thermally induced graphene rotation on hexagonal boron nitride. *Phys. Rev. Lett.* **116**, 126101 (2016).

39. 39.

Gustafsson, M. V. et al. Ambipolar Landau levels and strong band-selective carrier interactions in monolayer WSe₂. *Nat. Mater.* **17**, 411–415 (2018).

40. 40.

Larentis, S. et al. Large effective mass and interaction-enhanced Zeeman splitting of K-valley electrons in MoSe₂. *Phys. Rev. B* **97**, 201407 (2018).

41. 41.

Fallahazad, B. et al. Shubnikov-de Haas oscillations of high-mobility holes in monolayer and bilayer WSe₂ Landau level degeneracy, effective mass, and negative compressibility. *Phys. Rev. Lett.* **116**, 086601 (2016).

42. 42.

Cui, X. et al. Multi-terminal transport measurements of MoS₂ using a van der Waals heterostructure device platform. *Nat. Nanotechnol.* **10**, 534–540 (2015).

43. 43.

Ghiotto, A. et al. Quantum criticality in twisted transition metal dichalcogenides. Preprint at <https://arxiv.org/abs/2103.09796> (2021).

44. 44.

Li, T. et al. Continuous Mott transition in semiconductor moiré superlattices. Preprint at <https://arxiv.org/abs/2103.09779> (2021).

45. 45.

Li, T. et al. Quantum anomalous Hall effect from intertwined moiré bands. Preprint at <https://arxiv.org/abs/2107.01796> (2021).

46. 46.

Lui, C. H. et al. Imaging stacking order in few-layer graphene. *Nano Lett.* **11**, 164–169 (2011).

47. 47.

Cong, C. et al. Raman characterization of ABA- and ABC-stacked trilayer graphene. *ACS Nano* **5**, 8760–8768 (2012).

48. 48.

Dienwiebel, M. et al. Superlubricity of graphite. *Phys. Rev. Lett.* **92**, 126101 (2004).

49. 49.

Zheng, Q. et al. Self-retracting motion of graphite microflakes. *Phys. Rev. Lett.* **100**, 067205 (2008).

50. 50.

Liu, Z. et al. Observation of microscale superlubricity in graphite. *Phys. Rev. Lett.* **108**, 205503 (2012).

51. 51.

Yang, J. et al. Observation of high-speed microscale superlubricity in graphite. *Phys. Rev. Lett.* **110**, 255504 (2013).

52. 52.

Woods, C. R. et al. Macroscopic self-reorientation of interacting two-dimensional crystals. *Nat. Commun.* **7**, 10800 (2016).

53. 53.

Zhu, M. et al. Stacking transition in bilayer graphene caused by thermally activated rotation. *2D Mater.* **4**, 011013 (2016).

54. 54.

Feng, X., Kwon, S., Park, J. Y. & Salmeron, M. Superlubric sliding of graphene nanoflakes on graphene. *ACS Nano* **7**, 1718–1724 (2013).

55. 55.

Woods, C. R. et al. Commensurate–incommensurate transition in graphene on hexagonal boron nitride. *Nat. Phys.* **10**, 451–456 (2014).
This reference reports the commensurate–incommensurate transition in monolayer graphene at very small angles to the underlying hBN substrates.

56. 56.

Brown, L. et al. Twinning and twisting of tri- and bilayer graphene. *Nano Lett.* **12**, 1609–1615 (2012).

57. 57.

Alden, J. S. et al. Strain solitons and topological defects in bilayer graphene. *Proc. Natl Acad. Sci. USA* **110**, 11256–11260 (2013).

58. 58.

Xu, S. G. et al. Giant oscillations in a triangular network of one-dimensional states in marginally twisted graphene. *Nat. Commun.* **10**, 4008 (2019).

59. 59.

Yoo, H. et al. Atomic and electronic reconstruction at the van der Waals interface in twisted bilayer graphene. *Nat. Mater.* **18**, 448–453 (2019).

60. 60.

Kazmierczak, N. P. et al. Strain fields in twisted bilayer graphene. *Nat. Mater.* **20**, 956–963 (2021). **This reference maps the strain fields and structural relaxation in t-BLG using Bragg interferometry.**

61. 61.

Li, H. et al. Imaging moiré flat bands in three-dimensional reconstructed WSe₂/WS₂ superlattices. *Nat. Mater.* **20**, 945–950 (2021). **This reference demonstrates strong three-dimensional buckling reconstruction and large in-plane strain redistribution in WSe₂/WS₂ moiré heterostructures.**

62. 62.

Bistritzer, R. & MacDonald, A. H. Moiré bands in twisted double-layer graphene. *Proc. Natl Acad. Sci. USA* **108**, 12233–12237 (2011). **This reference theoretically shows the presence of flat bands in t-BLG and predicts the magic angle of 1.05°.**

63. 63.

Ramires, A. & Lado, J. L. Electrically tunable gauge fields in tiny-angle twisted bilayer graphene. *Phys. Rev. Lett.* **121**, 146801 (2018).

64. 64.

De Sanctis, A. et al. Strain-engineering of twist-angle in graphene/hBN superlattice devices. *Nano Lett.* **18**, 7919–7926 (2018).

65. 65.

Beechem, T. E., Ohta, T., Diaconescu, B. & Robinson, J. T. Rotational disorder in twisted bilayer graphene. *ACS Nano* **8**, 1655–1663 (2014).

66. 66.

Kerelsky, A. et al. Maximized electron interactions at the magic angle in twisted bilayer graphene. *Nature* **572**, 95–100 (2019).

67. 67.

Zhang, C. et al. Interlayer couplings, moiré patterns, and 2D electronic superlattices in MoS₂/WSe₂ hetero-bilayers. *Sci. Adv.* **3**, e1601459 (2017).

68. 68.

Xie, Y. et al. Spectroscopic signatures of many-body correlations in magic-angle twisted bilayer graphene. *Nature* **572**, 101–105 (2019).

69. 69.

Jiang, Y. et al. Charge order and broken rotational symmetry in magic-angle twisted bilayer graphene. *Nature* **573**, 91–95 (2019).

70. 70.

Choi, Y. et al. Electronic correlations in twisted bilayer graphene near the magic angle. *Nat. Phys.* **15**, 1174–1180 (2019).

71. 71.

Uri, A. et al. Mapping the twist-angle disorder and Landau levels in magic-angle graphene. *Nature* **581**, 47–52 (2020). **This reference maps the twist-angle disorder in t-BLG using scanning SQUID microscopy.**

72. 72.

Zondiner, U. et al. Cascade of phase transitions and Dirac revivals in magic-angle graphene. *Nature* **582**, 203–208 (2020).

73. 73.

McGilly, L. J. et al. Visualization of moiré superlattices. *Nat. Nanotechnol.* **15**, 580–584 (2020).

74. 74.

Sainz-Cruz, H., Cea, T., Pantaleón, P. A. & Guinea, F. High transmission in twisted bilayer graphene with angle disorder. Preprint at <https://arxiv.org/abs/2105.03383> (2021).

75. 75.

Lee, S. et al. Graphene transfer in vacuum yielding a high quality graphene. *Carbon* **93**, 286–294 (2015).

76. 76.

Masubuchi, S. et al. Autonomous robotic searching and assembly of two-dimensional crystals to build van der Waals superlattices. *Nat. Commun.* **9**, 1413 (2018).

77. 77.

Gong, Y. et al. Vertical and in-plane heterostructures from WS₂/MoS₂ monolayers. *Nat. Mater.* **13**, 1135–1142 (2014).

78. 78.

Liu, J. & Dai, X. Orbital magnetic states in moiré graphene systems. *Nat. Rev. Phys.* **3**, 367–382 (2021).

79. 79.

Andrei, E. Y. & MacDonald, A. H. Graphene bilayers with a twist. *Nat. Mater.* **19**, 1265–1275 (2020).

80. 80.

Balents, L., Dean, C. R., Efetov, D. K. & Young, A. F. Superconductivity and strong correlations in moiré flat bands. *Nat.*

Phys. **16**, 725–733 (2020).

81. 81.

Liang, L. et al. Band geometry, Berry curvature, and superfluid weight. *Phys. Rev. B* **95**, 024515 (2017).

82. 82.

Peotta, S. & Törmä, P. Superfluidity in topologically nontrivial flat bands. *Nat. Commun.* **6**, 8944 (2015).

83. 83.

Ma, C. et al. Moiré band topology in twisted bilayer graphene. *Nano Lett.* **20**, 6076–6083 (2020).

84. 84.

Song, Z. et al. All magic angles in twisted bilayer graphene are topological. *Phys. Rev. Lett.* **123**, 036401 (2019).

85. 85.

Po, H. C., Zou, L., Senthil, T. & Vishwanath, A. Faithful tight-binding models and fragile topology of magic-angle bilayer graphene. *Phys. Rev. B* **99**, 195455 (2019).

86. 86.

Ahn, J., Park, S. & Yang, B.-J. Failure of Nielsen–Ninomiya theorem and fragile topology in two-dimensional systems with space-time inversion symmetry: application to twisted bilayer graphene at magic angle. *Phys. Rev. X* **9**, 021013 (2019).

87. 87.

Hu, X., Hyart, T., Pikulin, D. I. & Rossi, E. Geometric and conventional contribution to the superfluid weight in twisted bilayer

graphene. *Phys. Rev. Lett.* **123**, 237002 (2019).

88. 88.

Julku, A., Peltonen, T. J., Liang, L., Heikkilä, T. T. & Törmä, P. Superfluid weight and Berezinskii–Kosterlitz–Thouless transition temperature of twisted bilayer graphene. *Phys. Rev. B* **101**, 060505 (2020).

89. 89.

Xie, F., Song, Z., Lian, B. & Bernevig, B. A. Topology-bounded superfluid weight in twisted bilayer graphene. *Phys. Rev. Lett.* **124**, 167002 (2020).

90. 90.

Tian, H. et al. Evidence for flat band Dirac superconductor originating from quantum geometry. Preprint at <https://arxiv.org/abs/2112.13401> (2021).

91. 91.

Yankowitz, M. et al. Tuning superconductivity in twisted bilayer graphene. *Science* **363**, 1059–1064 (2019).

92. 92.

Lu, X. et al. Superconductors, orbital magnets and correlated states in magic-angle bilayer graphene. *Nature* **574**, 653–657 (2019). **This reference reports the observation of a correlated insulating state, orbital magnetism and superconductivity at every integer filling of t-BLG.**

93. 93.

Stepanov, P. et al. Competing zero-field Chern insulators in superconducting twisted bilayer graphene. Preprint at <https://arxiv.org/abs/2012.15126> (2020).

94. 94.

An, L. et al. Interaction effects and superconductivity signatures in twisted double-bilayer WSe₂. *Nanoscale Horiz.* **5**, 1309–1316 (2020).

95. 95.

Rodan-Legrain, D. et al. Highly tunable junctions and nonlocal Josephson effect in magic angle graphene tunneling devices. Preprint at <https://arxiv.org/abs/2011.02500> (2020).

96. 96.

Vries, F. K. D. et al. Gate-defined Josephson junctions in magic-angle twisted bilayer graphene. Preprint at <https://arxiv.org/abs/2011.00011> (2020).

97. 97.

Stepanov, P. et al. Untying the insulating and superconducting orders in magic-angle graphene. *Nature* **583**, 375–378 (2020).

98. 98.

Codecido, E. et al. Correlated insulating and superconducting states in twisted bilayer graphene below the magic angle. *Sci. Adv.* **5**, eaaw9770 (2019).

99. 99.

Wu, S., Zhang, Z., Watanabe, K., Taniguchi, T. & Andrei, E. Y. Chern insulators, van Hove singularities and topological flat bands in magic-angle twisted bilayer graphene. *Nat. Mater.* **20**, 488–494 (2021).

100. 100.

Saito, Y. et al. Hofstadter subband ferromagnetism and symmetry-broken Chern insulators in twisted bilayer graphene. *Nat. Phys.* **17**, 478–481 (2021).

101. 101.

Wong, D. et al. Cascade of electronic transitions in magic-angle twisted bilayer graphene. *Nature* **582**, 198–202 (2020).

102. 102.

Nuckolls, K. P. et al. Strongly correlated Chern insulators in magic-angle twisted bilayer graphene. *Nature* **588**, 610–615 (2020).

103. 103.

Serlin, M. et al. Intrinsic quantized anomalous Hall effect in a moiré heterostructure. *Science* **367**, 900–903 (2020). **This reference reports the observation of an intrinsic QAH state (that is, without magnetic dopants) in t-BLG aligned to hBN.**

104. 104.

Pierce, A. T. et al. Unconventional sequence of correlated Chern insulators in magic-angle twisted bilayer graphene. Preprint at <https://arxiv.org/abs/2101.04123> (2021).

105. 105.

Sharpe, A. L. et al. Emergent ferromagnetism near three-quarters filling in twisted bilayer graphene. *Science* **365**, 605–608 (2019).

106. 106.

Tschirhart, C. L. et al. Imaging orbital ferromagnetism in a moiré Chern insulator. Preprint at <https://arxiv.org/abs/2006.08053> (2020).

107. 107.

Zhang, F., Jung, J., Fiete, G. A., Niu, Q. A. & MacDonald, A. H. Spontaneous quantum Hall states in chirally stacked few-layer graphene systems. *Phys. Rev. Lett.* **106**, 156801 (2011).

108. 108.

Velasco, J. et al. Transport spectroscopy of symmetry-broken insulating states in bilayer graphene. *Nat. Nanotechnol.* **7**, 156–160 (2012).

109. 109.

Shi, Y. et al. Electronic phase separation in multilayer rhombohedral graphite. *Nature* **584**, 210–214 (2020).

110. 110.

Lee, Y. et al. Gate tunable magnetism and giant magnetoresistance in ABC-stacked few-layer graphene. Preprint at <https://arxiv.org/abs/1911.04450> (2019).

111. 111.

Geisenhof, F. R. et al. Quantum anomalous Hall octet driven by orbital magnetism in bilayer graphene. *Nature* **598**, 53–58 (2021).

112. 112.

Jin, C. et al. Stripe phases in WSe₂/WS₂ moiré superlattice. Preprint at <https://arxiv.org/abs/2007.12068> (2020).

113. 113.

Huang, X. et al. Correlated insulating states at fractional fillings of the WS₂/WSe₂ moiré lattice. Preprint at <https://arxiv.org/abs/2007.11155> (2020).

114. 114.

Li, T. et al. Charge-order-enhanced capacitance in semiconductor moiré superlattices. Preprint at <https://arxiv.org/abs/2102.10823> (2021).

115. 115.

Fradkin, E., Kivelson, S. A., Lawler, M. J., Eisenstein, J. P. & Mackenzie, A. P. Nematic Fermi fluids in condensed matter physics. *Annu. Rev. Condens. Matter Phys.* **1**, 153–178 (2010).

116. 116.

Li, S.-Y. et al. Splitting of van Hove singularities in slightly twisted bilayer graphene. *Phys. Rev. B* **96**, 155416 (2017).

117. 117.

Cao, Y. et al. Nematicity and competing orders in superconducting magic-angle graphene. Preprint at <https://arxiv.org/abs/2004.04148> (2020).

118. 118.

Jin, C. et al. Stripe phases in WSe₂/WS₂ moire superlattices. Preprint at <https://arxiv.org/abs/2007.12068> (2020).

119. 119.

Xian, L. et al. Realization of nearly dispersionless bands with strong orbital anisotropy from destructive interference in twisted bilayer MoS₂. Preprint at <https://arxiv.org/abs/2004.02964> (2020).

120. 120.

Zhou, H., Xie, T., Taniguchi, T., Watanabe, K. & Young, A. F. Superconductivity in rhombohedral trilayer graphene. Preprint at <https://arxiv.org/abs/2106.07640> (2021).

121. 121.

Qiao, J.-B., Yin, L.-J. & He, L. Twisted graphene bilayer around the first magic angle engineered by heterostrain. *Phys. Rev. B* **98**, 235402

(2018).

122. 122.

Huder, L. et al. Electronic spectrum of twisted graphene layers under heterostrain. *Phys. Rev. Lett.* **120**, 156405 (2018).

123. 123.

Shi, H. et al. Large-area, periodic, and tunable intrinsic pseudo-magnetic fields in low-angle twisted bilayer graphene. *Nat. Commun.* **11**, 371 (2020).

124. 124.

Kariyado, T. & Vishwanath, A. Flat band in twisted bilayer Bravais lattices. *Phys. Rev. Res.* **1**, 033076 (2019).

125. 125.

Hejazi, K., Luo, Z.-X. & Balents, L. Noncollinear phases in moiré magnets. *Proc. Natl Acad. Sci. USA* **117**, 10721–10426 (2020).

126. 126.

Xu, Y. et al. Emergence of a noncollinear magnetic state in twisted bilayer CrI₃. Preprint at <https://arxiv.org/abs/2103.09850> (2021).

127. 127.

Chittari, B. L., Chen, G., Zhang, Y., Wang, F. & Jung, J. Gate-tunable topological flat bands in trilayer graphene boron-nitride moiré superlattices. *Phys. Rev. Lett.* **122**, 016401 (2019).

128. 128.

Tarnopolsky, G., Kruchkov, A. J. & Vishwanath, A. Origin of magic angles in twisted bilayer graphene. *Phys. Rev. Lett.* **122**, 106405 (2019).

129. 129.

San-Jose, P., González, J. & Guinea, F. Non-Abelian gauge potentials in graphene bilayers. *Phys. Rev. Lett.* **108**, 216802 (2012).

130. 130.

Liu, J., Ma, Z., Gao, J. & Dai, X. Quantum valley Hall effect, orbital magnetism, and anomalous Hall effect in twisted multilayer graphene systems. *Phys. Rev. X* **9**, 031021 (2019).

131. 131.

Khalaf, E., Kruchkov, A. J., Tarnopolsky, G. & Vishwanath, A. Magic angle hierarchy in twisted graphene multilayers. *Phys. Rev. B* **100**, 085109 (2019).

132. 132.

Bao, W. et al. Stacking-dependent band gap and quantum transport in trilayer graphene. *Nat. Phys.* **7**, 948–952 (2011).

133. 133.

Angeli, M. & MacDonald, A. H. Γ valley transition metal dichalcogenide moiré bands. *Proc. Natl Acad. Sci. USA* **118**, e2021826118 (2021).

Acknowledgements

C.N.L. acknowledges support from the Department of Energy under grant number DOE DE-SC0020187. M.W.B. acknowledges support from the National Science Foundation under grant numbers DMR-2004801 and DMR-2105028. F.Z. acknowledges support from the Army Research Office under grant number W911NF-18-1-0416 and the National Science Foundation under grant numbers DMR-1945351, DMR-2105139 and DMR-1921581. K.F.M. acknowledges support from the Air

Force Office of Scientific Research under award number FA9550-20-1-0219.

Author information

Affiliations

1. Department of Physics, The Ohio State University, Columbus, OH, USA

Chun Ning Lau & Marc W. Bockrath

2. Department of Physics, and School of Applied and Engineering Physics, Cornell University, Ithaca, NY, USA

Kin Fai Mak

3. Department of Physics, University of Texas at Dallas, Dallas, TX, USA

Fan Zhang

Contributions

C.N.L., M.W.B., K.F.M. and F.Z. discussed and co-wrote the manuscript.

Corresponding author

Correspondence to [Chun Ning Lau](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Roman Gorbachev 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.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Lau, C.N., Bockrath, M.W., Mak, K.F. *et al.* Reproducibility in the fabrication and physics of moiré materials. *Nature* **602**, 41–50 (2022). <https://doi.org/10.1038/s41586-021-04173-z>

- Received: 16 March 2021
- Accepted: 21 October 2021
- Published: 02 February 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04173-z>

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/s41586-021-04173-z>

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

- Perspective
- [Published: 02 February 2022](#)

An open science study of ageing in companion dogs

- [Kate E. Creevy](#) [ORCID: orcid.org/0000-0003-4169-374X¹](#),
- [Joshua M. Akey](#) [ORCID: orcid.org/0000-0002-4411-1330²](#),
- [Matt Kaeberlein](#) [ORCID: orcid.org/0000-0002-1311-3421³](#),
- [Daniel E. L. Promislow](#) [ORCID: orcid.org/0000-0001-7088-4495^{3,4}](#)
&
- [The Dog Aging Project Consortium](#)

Nature volume **602**, pages 51–57 (2022)

- 1950 Accesses
- 161 Altmetric
- [Metrics details](#)

Subjects

- [Ageing](#)
- [Epidemiology](#)
- [Genome-wide association studies](#)
- [Predictive markers](#)
- [Translational research](#)

Abstract

The Dog Aging Project is a long-term longitudinal study of ageing in tens of thousands of companion dogs. The domestic dog is among the most variable mammal species in terms of morphology, behaviour, risk of age-related disease and life expectancy. Given that dogs share the human environment and have a sophisticated healthcare system but are much shorter-lived than people, they offer a unique opportunity to identify the genetic, environmental and lifestyle factors associated with healthy lifespan. To take advantage of this opportunity, the Dog Aging Project will collect extensive survey data, environmental information, electronic veterinary medical records, genome-wide sequence information, clinicopathology and molecular phenotypes derived from blood cells, plasma and faecal samples. Here, we describe the specific goals and design of the Dog Aging Project and discuss the potential for this open-data, community science study to greatly enhance understanding of ageing in a genetically variable, socially relevant species living in a complex environment.

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: Structure of the DAP cohorts.

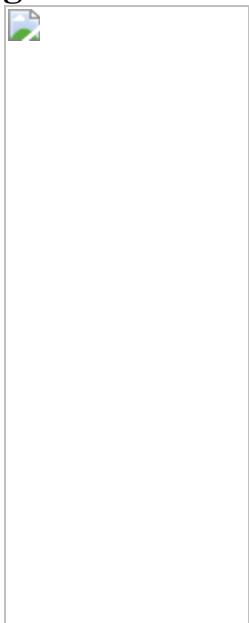


Fig. 2: DAP integration.

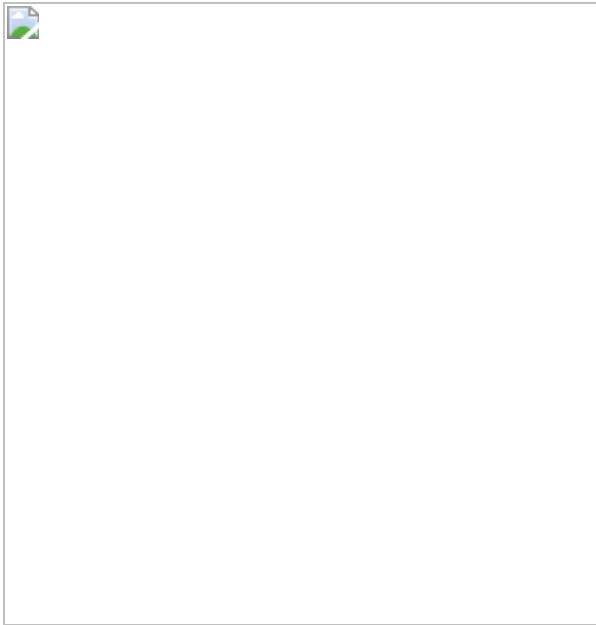
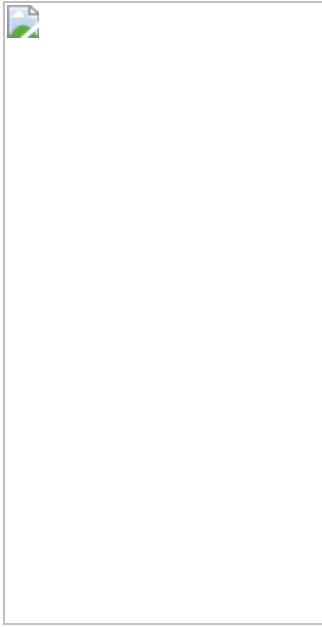


Fig. 3: Biospecimen and environmental measures.



Data availability

The data used to generate Fig. 1b,c are freely available for download at <https://data.dogagingproject.org>.

References

1. 1.

Kaeberlein, M., Rabinovitch, P. S. & Martin, G. M. Healthy aging: the ultimate preventative medicine. *Science* **350**, 1191–1193 (2015). **This paper makes a compelling argument that treatments that target the underlying causes of ageing could ameliorate the effects of multiple age-related diseases.**

2. 2.

Melzer, D., Hurst, A. J. & Frayling, T. Genetic variation and human aging: progress and prospects. *J. Gerontol. A Biol. Sci. Med. Sci.* **62**, 301–307 (2007).

3. 3.

Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

4. 4.

Kaeberlein, M., Creevy, K. E. & Promislow, D. E. L. The Dog Aging Project: translational geroscience in companion animals. *Mamm. Genome* **27**, 279–288 (2016).

5. 5.

Boyko, A. R. et al. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol.* **8**, e1000451 (2010). **This paper demonstrated the tremendous power of the domestic dog as a model for mapping natural variation for complex traits.**

6. 6.

Minnema, L. et al. Correlation of artemin and GFR α 3 with osteoarthritis pain: early evidence from naturally occurring osteoarthritis-associated chronic pain in dogs. *Front. Neurosci.* **14**, 77 (2020).

7. 7.

Harris, P. A. et al. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42**, 377–381 (2009).

8. 8.

Harris, P. A. et al. The REDCap consortium: building an international community of software platform partners. *J. Biomed. Inform.* **95**, 103208 (2019).

9. 9.

Li, J. H., Mazur, C. A., Berisa, T. & Pickrell, J. K. Low-pass sequencing increases the power of GWAS and decreases measurement error of polygenic risk scores compared to genotyping arrays. *Genome Res.* **31**, 529–537 (2021).

10. 10.

Franceschi, C. & Campisi, J. Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. *J. Gerontol. A Biol. Sci. Med. Sci.* **69**, S4–S9 (2014).

11. 11.

Lippi, G. et al. Preanalytical challenges—time for solutions. *Clin. Chem. Lab. Med.* **57**, 974–981 (2019).

12. 12.

Haumann, R. & Verspaget, H. W. Quality-assured biobanking: the Leiden University Medical Center model. *Methods Mol. Biol.* **1730**, 361–370 (2018).

13. 13.

Simeon-Dubach, D., Zeisberger, S. M. & Hoerstrup, S. P. Quality assurance in biobanking for pre-clinical research. *Transfus Med. Hemother.* **43**, 353–357 (2016).

14. 14.

US Department of the Census. American Community Survey (ACS): Public Use Microdata Sample (PUMS), 2009. <https://doi.org/10.3886/ICPSR33802.v1> (Inter-university Consortium for Political and Social Research, 2013).

15. 15.

Kim, S.-Y. et al. Concentrations of criteria pollutants in the contiguous U.S., 1979–2015: role of prediction model parsimony in integrated empirical geographic regression. *PLoS ONE* **15** e0228535 (2020).

16. 16.

Vose, R. S. et al. NOAA Monthly U.S. Climate Divisional Database (NClimDiv). <https://doi.org/10.7289/V5M32STR> (NOAA National Climatic Data Center, 2014).

17. 17.

Mooney, S. J. et al. Residential neighborhood features associated with objectively measured walking near home: revisiting walkability using the Automatic Context Measurement Tool (ACMT). *Health Place* **63**, 102332 (2020).

18. 18.

Wilfond, B. S., Porter, K. M., Creevy, K. E., Kaeberlein, M. & Promislow, D. Research to promote longevity and health span in companion dogs: a pediatric perspective. *Am. J. Bioeth.* **18** 64–65 (2018).

19. 19.

Taylor, H. A., Morales, C., Johnson, L.-M. & Wilfond, B. S. A randomized trial of rapamycin to increase longevity and healthspan in companion animals: navigating the boundary between protections for animal research and human subjects research. *Am. J. Bioeth.* **18**, 58–59 (2018).

20. 20.

Bisong, E. in *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners* (ed. Bisong, E.) 7–10 (Apress, 2019).

21. 21.

Salvin, H. E., McGreevy, P. D., Sachdev, P. S. & Valenzuela, M. J. The canine cognitive dysfunction rating scale (CCDR): a data-driven and ecologically relevant assessment tool. *Vet. J.* **188**, 331–336 (2011).

Acknowledgements

All dog research described here, including informed owner consent, is approved by the Texas A&M University Institutional Animal Care and Use Committee, under AUPs 2018-0401 CAM and 2018-0368 CAM. The DAP is supported by grant U19AG057377 from the National Institute on Aging, a part of the National Institutes of Health, and by private donations. We thank S. Moon for help in preparing figures.

Author information

Affiliations

1. Department of Small Animal Clinical Sciences, Texas A&M University College of Veterinary Medicine & Biomedical Sciences, College Station, TX, USA

Kate E. Creevy, Brian G. Barnett, Lucy Chou, Jeremy Evans, Jonathan M. Levine, Kellyn E. McNulty & M. Katherine Tolbert

2. Lewis-Sigler Institute for Integrative Genomics, Princeton University,
Princeton, NJ, USA

Joshua M. Akey & William Thistlethwaite

3. Department of Laboratory Medicine and Pathology, University of
Washington School of Medicine, Seattle, WA, USA

Matt Kaeberlein, Daniel E. L. Promislow, Brooke Benton, Erica C.
Jonlin & Silvan R. Urfer

4. Department of Biology, University of Washington, Seattle, WA, USA

Daniel E. L. Promislow

5. Department of Clinical Microbiology and Immunology, Sackler
Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Elhanan Borenstein

6. Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv,
Israel

Elhanan Borenstein

7. Santa Fe Institute, Santa Fe, NM, USA

Elhanan Borenstein

8. Cornell Veterinary Biobank, College of Veterinary Medicine, Cornell
University, Ithaca, NY, USA

Marta G. Castelhano

9. Department of Sociology, University of Washington, Seattle, WA,
USA

Devin Collins, Kyle Crowder & Hannah Lee

10. Department of Small Animal Medicine and Surgery, University of Georgia College of Veterinary Medicine, Athens, GA, USA

Amanda E. Coleman

11. Center for Studies in Demography and Ecology, Seattle, WA, USA

Kyle Crowder & Matthew D. Dunbar

12. Department of Veterinary Physiology and Pharmacology, Texas A&M University College of Veterinary Medicine & Biomedical Sciences, College Station, TX, USA

Virginia R. Fajt

13. Department of Family Medicine, University of Washington, Seattle, WA, USA

Annette L. Fitzpatrick

14. Department of Epidemiology, University of Washington, Seattle, WA, USA

Annette L. Fitzpatrick & Stephen M. Schwartz

15. Department of Global Health, University of Washington, Seattle, WA, USA

Annette L. Fitzpatrick

16. Department of Veterinary Pathobiology, Texas A&M University College of Veterinary Medicine & Biomedical Sciences, College Station, TX, USA

Unity Jeffery

17. Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA, USA

Erica C. Jonlin

18. Bioinformatics and Integrative Biology, University of Massachusetts Chan Medical School, Worcester, MA, USA

Elinor K. Karlsson & Kathleen Morrill

19. Broad Institute of MIT and Harvard, Cambridge, MA, USA

Elinor K. Karlsson

20. Department of Biostatistics, University of Washington, Seattle, WA, USA

Kathleen F. Kerr, Robyn L. McClelland & Yunbi Nam

21. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Jing Ma

22. Department of Population Health Sciences, Virginia-Maryland College of Veterinary Medicine, Virginia Tech, Blacksburg, VA, USA

Audrey Ruple

23. Epidemiology Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Stephen M. Schwartz

24. Collaborative Health Studies Coordinating Center, Department of Biostatistics, University of Washington, Seattle, WA, USA

Sandi Shrager

25. School of Life Sciences, Arizona State University, Tempe, AZ, USA

Noah Snyder-Mackler

26. Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA

Noah Snyder-Mackler

27. School for Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA

Noah Snyder-Mackler

28. Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA, USA

Benjamin S. Wilfond

29. Division of Bioethics and Palliative Care, Department of Pediatrics, University of Washington School of Medicine, Seattle, WA, USA

Benjamin S. Wilfond

Consortia

The Dog Aging Project Consortium

- , Kate E. Creevy
- , Joshua M. Akey
- , Matt Kaeberlein
- , Daniel E. L. Promislow
- , Brian G. Barnett
- , Brooke Benton
- , Elhanan Borenstein
- , Marta G. Castelhano
- , Lucy Chou
- , Devin Collins
- , Amanda E. Coleman
- , Kyle Crowder
- , Matthew D. Dunbar
- , Jeremy Evans

- , Virginia R. Fajt
- , Annette L. Fitzpatrick
- , Unity Jeffery
- , Erica C. Jonlin
- , Elinor K. Karlsson
- , Kathleen F. Kerr
- , Hannah Lee
- , Jonathan M. Levine
- , Jing Ma
- , Robyn L. McClelland
- , Kellyn E. McNulty
- , Kathleen Morrill
- , Yunbi Nam
- , Audrey Ruple
- , Stephen M. Schwartz
- , Sandi Shrager
- , Noah Snyder-Mackler
- , William Thistlethwaite
- , M. Katherine Tolbert
- , Silvan R. Urfer
- & Benjamin S. Wilfond

Contributions

K.E.C., M.K. and D.E.L.P. conceived of the DAP; J.M.A., K.E.C., M.K. and D.E.L.P. wrote the initial draft of this paper. All authors, including consortium authors, have been involved in the design and implementation of DAP goals, infrastructure and activities, and they have had the opportunity to participate in editing both form and content of this paper and have approved the final version.

Corresponding author

Correspondence to [Daniel E. L. Promislow](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Steven Austad, Dario Valenzano and Eric Verdin 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.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Creevy, K.E., Akey, J.M., Kaeberlein, M. *et al.* An open science study of ageing in companion dogs. *Nature* **602**, 51–57 (2022).
<https://doi.org/10.1038/s41586-021-04282-9>

- Received: 27 May 2021
- Accepted: 24 November 2021
- Published: 02 February 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04282-9>

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

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04282-9>

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

- Article
- Open Access
- [Published: 02 February 2022](#)

Microwave background temperature at a redshift of 6.34 from H₂O absorption

- [Dominik A. Riechers](#) ORCID: orcid.org/0000-0001-9585-1462¹,
- [Axel Weiss](#)²,
- [Fabian Walter](#)^{3,4},
- [Christopher L. Carilli](#)⁴,
- [Pierre Cox](#)⁵,
- [Roberto Decarli](#)⁶ &
- [Roberto Neri](#) ORCID: orcid.org/0000-0002-7176-4046⁷

[Nature](#) volume 602, pages 58–62 (2022)

- 533 Accesses
- 132 Altmetric
- [Metrics details](#)

Subjects

- [Cosmology](#)
- [Early universe](#)

Abstract

Distortions of the observed cosmic microwave background provide a direct measurement of the microwave background temperature at redshifts from 0 to 1 (refs. [1,2](#)). Some additional background temperature estimates exist at redshifts from 1.8 to 3.3 based on molecular and atomic line-excitation temperatures in quasar absorption-line systems, but are model dependent³. No deviations from the expected $(1 + z)$ scaling behaviour of the microwave background temperature have been seen⁴, but the measurements have not extended deeply into the matter-dominated era of the Universe

at redshifts $z > 3.3$. Here we report observations of submillimetre line absorption from the water molecule against the cosmic microwave background at $z = 6.34$ in a massive starburst galaxy, corresponding to a lookback time of 12.8 billion years (ref. 5).

Radiative pumping of the upper level of the ground-state ortho-H₂O(1₁₀–1₀₁) line due to starburst activity in the dusty galaxy HFLS3 results in a cooling to below the redshifted microwave background temperature, after the transition is initially excited by the microwave background. This implies a microwave background temperature of 16.4–30.2 K (1σ range) at $z = 6.34$, which is consistent with a background temperature increase with redshift as expected from the standard Λ CDM cosmology⁴.

[Download PDF](#)

Main

We used the Northern Extended Millimeter Array (NOEMA) to obtain a sensitive scan across the 3-mm atmospheric window towards the $z = 6.34$ massive dusty starburst galaxy HFLS3 (also known as 1HERMES S350 J170647.8+584623; see Methods)⁵. These observations reveal a broad range of emission features dominated by the CO, H₂O and H₂O⁺ molecules and atomic carbon, on top of thermal dust continuum emission that is rising with frequency consistent with a dust temperature of $T_{\text{dust}} = \langle \{63.3\} - \{-5.8\}^{+5.4} \rangle \text{ K}$ (refs. 5,6) (Fig. 1). The spectrum also shows a deep absorption feature due to the ortho-H₂O(1₁₀–1₀₁) ground-state transition at rest-frame 538 μm (observed at 3.95 mm, or 75.9 GHz). This absorption is about two times stronger than the continuum emission from the starburst at the same frequency (Fig. 2). For this effect to occur, a substantial population of the ortho-H₂O 1₁₀ level (which lies 26.7 K above the 1₀₁ ground state) has to be excited by cosmic microwave background (CMB) photons as a basis for pumping of this level by the starburst infrared radiation field (see Extended Data Fig. 1). The effect becomes observable towards HFLS3 because of the warm CMB at this redshift, which is predicted to be $T_{\text{CMB}} = 20.0 \text{ K}$ at $z = 6.34$ based on the standard Λ CDM cosmology (where $T_{\text{CMB}}(z) = T_{\text{CMB}}(z=0)*(1+z)^{(1-\beta)}$, $T_{\text{CMB}}(z=0) = 2.72548 \pm 0.00057 \text{ K}$ (ref. 7) and the power-law index $\beta = 0$). The absorption of photons from the CMB radiation field appreciably populates the H₂O 1₁₀ level. The intense infrared radiation field from the starburst then preferentially de-populates the 1₁₀ level through radiative pumping, resulting in a deficit in the 1₁₀ level compared with 1₀₁ relative to a thermal distribution. In combination, these two processes result in an excitation temperature T_{ex} of the H₂O(1₁₀–1₀₁) line that is lower than T_{CMB} , such that the line becomes observable in absorption against the CMB. As the effect depends on the strength of the CMB

radiation field, it can be used to measure T_{CMB} for galaxies that have well-measured dust spectral energy distributions and dust continuum sizes, as is the case for HFLS3.

Fig. 1: Broad-band, 3-mm spectroscopy of the starburst galaxy HFLS3 at a redshift of 6.34 with NOEMA.

 figure 1



Black/yellow histogram, NOEMA spectroscopy data, binned to 40-MHz (158 km s⁻¹ at 75.9 GHz) spectral resolution. Expected frequencies of molecular and atomic lines at the redshift of HFLS3 are indicated, with the dominant species labelled in black. The dashed red box indicates the frequency range of the ortho-H₂O(1₁₀–1₀₁) line, which is detected in absorption against the CMB.

Fig. 2: H₂O line and continuum properties and modelling of HFLS3.

 **figure 2**

Ortho-H₂O energy-level structure (**a**, red solid arrows are detected transitions, grey dotted lines are upper limits and blue dashed arrows are pumping transitions, with observed and model-predicted absorption/emission lines indicated as upward/downward arrows, respectively; percentages are the level populations in the model) and zoom-in on the H₂O line at the same spectral resolution as in Fig. 1 to show that the line absorbs into the CMB (**b**, blue shading added for emphasis). The black curve is a fit to the spectrum. The red dashed curve is the best-fit radiative transfer model.

Source data

To understand the effect, we have calculated a series of spherically symmetric RADEX⁸ models over a wide range of H₂O column densities, assuming purely radiative excitation (Figs. 2 and 3; see Methods for additional details). Exposing a cold, H₂O-bearing region associated with HFLS3 to the black-body CMB radiation field at $T_{\text{CMB}}(z=6.34)$, the models suggest that 77.2% of the molecules will be in the 1₀₁ ground state and 20.3% will be in the upper 1₁₀ state, and all H₂O transitions have an excitation temperature T_{ex} equal to T_{CMB} . As a result of this zero temperature contrast, no H₂O emission or absorption would be observable, despite the fact that the H₂O rotational ladder is excited by the CMB radiation. However, this picture changes when the same region is also exposed to the infrared radiation field emitted by the starburst nucleus of HFLS3, as the latter does not follow a single-black-body radiation pattern. Indeed, the infrared spectral energy distribution of HFLS3 reaches its peak intensity at $\sim 73.3 \mu\text{m}$ and can be approximated by a grey body

with a Rayleigh–Jeans slope of $\beta_{\text{IR}} = \langle \{1.94\} _ \{-0.09\}^{\{+0.07\}} \rangle$. This is due to the presence of dust at multiple temperatures and an increasing dust optical depth towards shorter wavelengths^{5,6}. In this case, the level populations of the 1_{01} and 1_{10} states will deviate from the single-temperature thermal equilibrium population and change to 68.0% and 14.6%, respectively, for the solution shown in Fig. 2, resulting in an excitation temperature T_{ex} of only 17.4 K for this transition. Owing to the $\Delta J=1$ selection rule for photon emission/absorption, only three ortho-H₂O transitions contribute to the modification of populations in the 1_{01} and 1_{10} levels, namely, the 538-μm $1_{10}-1_{01}$ and 180-μm $2_{12}-1_{01}$ transitions affecting the former, and the 108-μm $2_{21}-1_{10}$ transition affecting the latter (see Fig. 2; the $2_{21}-1_{01}$ transition is forbidden). The over-proportional de-population of the 1_{10} level occurs because the H₂O($2_{21}-1_{10}$) transition at 108 μm dominates the modification of the level population^{9,10}. This transition lies near the peak of the dust spectral energy distribution, where the dust emission has a higher optical depth than at 538 μm, where the H₂O($1_{10}-1_{01}$) transition occurs. The increase in dust optical depth with wavelength leads to an increased availability of 108-μm photons relative to 538-μm photons compared with the thermal equilibrium case of a single-black-body radiation field. This implies that the $2_{21}-1_{10}$ transition at 108 μm is exposed to a more intense infrared radiation field than the $1_{10}-1_{01}$ transition at 538 μm. For the infrared radiation field of HFLS3, the models therefore suggest that both H₂O transitions should be found in absorption, but that the $2_{21}-1_{10}$ transition (which is not covered by our observations) will contribute a larger fraction to the 1_{10} level de-population than that found for the thermal equilibrium case. The CMB photons, on the other hand, only provide the base population expected for the thermal black-body equilibrium case. This causes a reduced excitation temperature in the $1_{10}-1_{01}$ transition compared with the thermal equilibrium case, unless the 1_{01} level is even more substantially de-populated due to the H₂O($2_{12}-1_{01}$) transition at 180 μm. This, however, does not occur, as our models show that this transition is expected to be seen in emission (at a line strength approximately 3–5 times below a previously reported upper limit⁵). This is because the upper-level population of the $2_{12}-1_{01}$ transition is also affected by the population of the 2_{12} level through the $2_{21}-2_{12}$ line, which appears in emission due to the pumping of its upper 2_{21} state by the $2_{21}-1_{10}$ absorption line. As such, the models suggest a net deficit in the upper-level population of the $1_{10}-1_{01}$ transition compared with the thermal equilibrium case, which causes the T_{ex} of the line to end up below the CMB temperature.

Fig. 3: Radiative transfer models for HFLS3 and constraints on the CMB temperature.

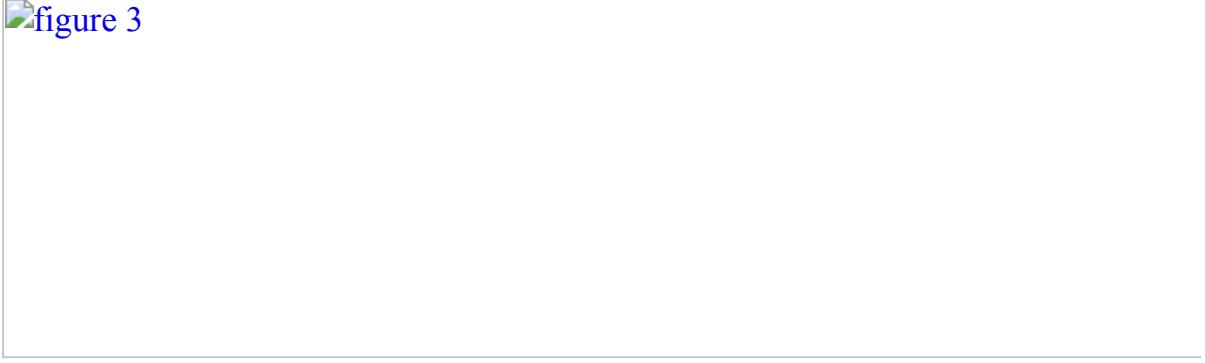


figure 3

a, Model grid for the predicted line-absorption strength for $T_{\text{CMB}}(z=6.34)=20.0 \text{ K}$ (greyscale) as a function of H_2O column density (y axis) and radius of the dust-emission region at $108 \mu\text{m}$ (x axis). The white curves show the parameter space allowed by the measurement (solid line) and the -1σ r.m.s. uncertainty region (dotted line). The dashed black lines show the measured continuum size (left) and $+1\sigma$ r.m.s. uncertainty region (right). The overlapping region between the white boundary (that is, the minimum allowed absorption strength) and the size measurement (that is, the minimum required emitting area at 100% covering fraction) is the allowed parameter space for the absorption strength within 1σ r.m.s. The minimum required radius at $N(\text{H}_2\text{O}) \sim 10^{17} \text{ cm}^{-2}$ is due to a minimum in T_{ex} in the models. **b**, Constraints on T_{CMB} for the observed absorption strength (green line and shaded region) at the minimum size compatible with the observations (**a**), based on the same models (red/blue shaded regions are the allowed ranges within the source radius $+1\sigma/+2\sigma$ r.m.s.). The source radius at face value (black line), as well as the -1σ and -2σ r.m.s. regions (not shown), are ruled out by the observations. The minimum filling factor of the dust emission region CF_{\min} is indicated for the $+1\sigma$ and $+2\sigma$ r.m.s. regions. The grey dashed line shows a model assuming a continuum radius of 5 kpc, which provides a conservative lower limit on T_{CMB} . **c**, Observability of the H_2O absorption as a function of redshift for three solutions allowed by the data without and with collisional excitation. The effect becomes observable at $z \sim 4.5$ and remains visible at similar strength to $z > 12$. The lower-redshift limit is higher in cases where collisional excitation is important, but the impact is minor below $n(\text{H}_2) = 10^5 \text{ cm}^{-3}$.

The RADEX models yield $T_{\text{ex}} = 17.4 \text{ K}$ due to this level-population modification. To translate model-predicted temperature differences into an absorption-line flux that can be compared with the observations, the size of the emitting/absorbing region needs to be known. Based on NOEMA observations at rest-frame $122 \mu\text{m}$ (Extended Data Fig. 2), we estimate the dust continuum size of the emitting region at $108 \mu\text{m}$ (the wavelength of the pumping transition) of HFLS3 to be $r_{108\mu\text{m}} = 1.62 \pm 0.45 \text{ kiloparsecs (kpc)}$. Within the uncertainties of the size estimate, the RADEX models suggest that the strength of the observed H_2O absorption can be reproduced over about two orders

of magnitude in H₂O column density, with a lower limit of around 10¹⁶ cm⁻². The minimum covering fraction of the dust continuum is about 60% when conservatively leaving T_{CMB} as a free parameter (100% is assumed for the grid shown in Fig. 3a). The upper limit for the H₂ column density implied by the gas mass of HFLS3 of (1.04 ± 0.09) × 10¹¹ M_{sun} (ref. 5) provides a lower limit to the gas phase [H₂O]/[H₂] abundance of >2 × 10⁻⁷, which falls within the range of 10⁻⁹–10⁻⁵ found for nearby starbursts¹¹. The small difference $\Delta T = T_{\text{ex}} - T_{\text{CMB}} = -2.6$ K is, therefore, sufficient to explain the observed strength of the H₂O(1₁₀–1₀₁) absorption line towards the CMB in HFLS3 when assuming a layer of cold, diffuse H₂O-bearing gas with a high covering fraction in front of the warm dust continuum source associated with the H₂O emission lines.

As the absorption line is observed in contrast to the CMB, we can use the strength of the absorption line to obtain a measurement of T_{CMB} at the redshift of HFLS3. The RADEX models suggest that, to detect the H₂O(1₁₀–1₀₁) line in absorption against the CMB, $T_{\text{CMB}}(z=6.34)$ must be >7–8 K, independent of the model assumptions (see Methods). The observed strength of the signal suggests 16.4 K < $T_{\text{CMB}}(z=6.34)$ < 30.2 K (1 σ , or 12.8 K < $T_{\text{CMB}}(z=6.34)$ < 34.0 K 2 σ) for HFLS3 when treating T_{dust} , β_{IR} and the wavelength where the dust optical depth reaches unity as free-fitting parameters for each dust continuum size sampled by the models. This explains why the effect has not been previously seen. T_{CMB} must be sufficiently high to satisfy the requirement of a notable H₂O 1₁₀ level population due to the CMB, such that a de-population by the infrared radiation field of the starburst will lead to a sufficiently important decrement to be observable in absorption against the CMB. This limits observability to $z > 4.5$ for dust spectral energy distribution shapes and dust continuum sizes of star-forming galaxies like HFLS3 (Fig. 3c), where only a few spectra at rest-frame 538 μm with sufficient signal-to-noise ratio to detect the effect exist. This differs from molecules like H₂CO, for which absorption against the CMB has been predicted to occur at any redshift up to the present day¹², but for which no detections at high redshift currently exist¹³. For starbursts with dust as warm as HFLS3, its relative strength is expected to continue to increase with redshift all the way up to $z \sim 7$ –8, and to remain observable back to the earliest epochs when such galaxies existed.

The thermal Sunyaev–Zel'dovich (SZ) effect entails the modification of the thermal distribution of CMB photons by Thomson scattering off electrons at high temperatures in the intergalactic medium of galaxy (proto)clusters. The effect observed here requires the CMB photons to excite the H₂O rotational ladder to create a thermal distribution of the lower energy levels, which is then modified through the absorption of far-infrared photons from the starburst radiation field permeating the interstellar medium. In both cases, the CMB is responsible for establishing unperturbed thermal

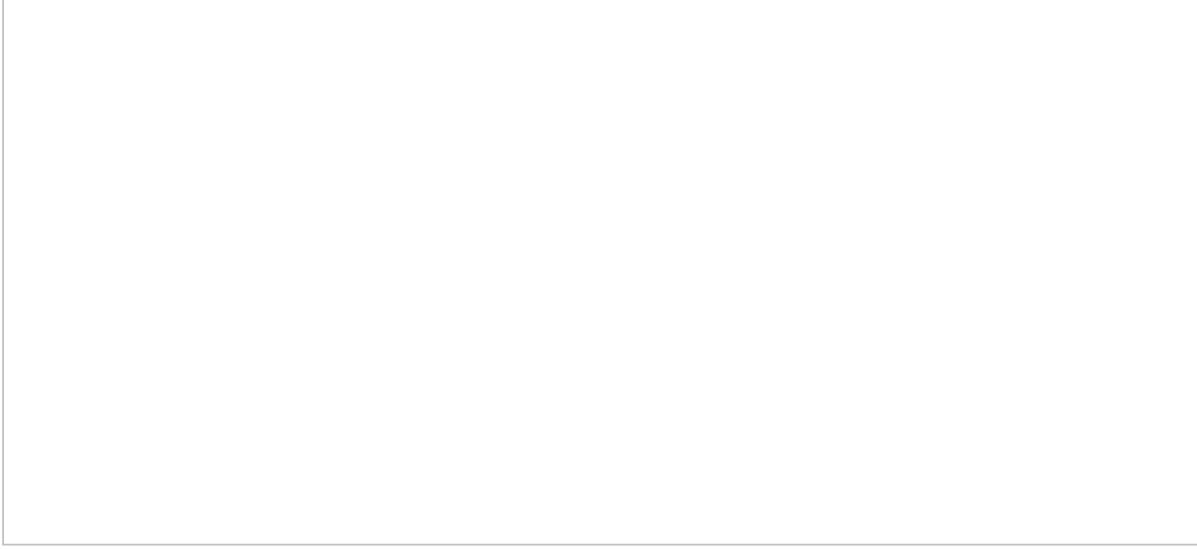
distributions (of photons and H₂O excitation, respectively), which are then modified by local conditions. The SZ effect is a broad-band modification of the thermal distribution of CMB photons via scattering, with an expected signal strength (relative to the CMB) that is independent of redshift. In contrast, the H₂O absorption signal described herein is a narrow-band (spectral-line) absorption process of the CMB photons, catalysed (in part) by the CMB itself, with an absorption-line strength that increases with redshift owing to the increasing temperature of the CMB, relative to the fixed excitation temperature of the H₂O(1₁₀–1₀₁) transition.

Standard Λ CDM cosmologies predict a linear increase of T_{CMB} with $(1+z)$. However, there are hypothetical physical mechanisms that could lead to departures from this linear behaviour, including the evolution of physical constants¹⁴, decaying dark-energy models¹⁵ and axion-photon-like coupling processes^{16,17}. Direct measurements of T_{CMB} are, thus, a crucial test of cosmology, but they are currently limited to $z < 1$, due to the lack of sufficiently precise measurements of the thermal SZ effect in galaxy clusters at higher redshifts (Fig. 4; see Methods for further details). A limited sample of additional constraints exists at $z = 1.8$ – 3.3 based on measurements of T_{ex} for the ultraviolet transitions of CO, [CI] and [CII] in absorption-line systems along the lines of sight to quasars. These lines are not directly observed in contrast to the CMB, and they use the T_{ex} of these lines as a proxy of T_{CMB} , such that the resulting T_{CMB} estimates are subject to model-dependent excitation corrections^{3,17,18,19,20,21,22,23,24,25,26,27,28}. As an example, for the CO molecule, T_{ex} typically already exceeds T_{CMB} in the diffuse interstellar medium in the Milky Way due to collisional excitation, showing a rising excess with increasing CO optical depth due to photon trapping¹⁸. In contrast, our models suggest that collisional excitation of H₂O becomes important only at very high densities, such that H₂O-based measurements are probably only minimally affected by this effect. The H₂O absorption against the CMB at $z = 6.34$ reported here thus provides the most direct constraint on T_{CMB} currently available at $z > 1$. Indeed, the existence of this effect on its own directly implies that the CMB is warmer than at low redshift, because T_{CMB} must be sufficiently high to notably excite the H₂O 1₁₀ level, which lies 26.7 K above ground, as a basis for the observed decrement due to de-population of this level by the starburst radiation field. A combined fit to the available data (Fig. 4) is consistent with the redshift scaling expected from Λ CDM. Fitting for the adiabatic index γ in the equation of state between pressure P and energy density ρ for the sum of baryonic and dark matter and radiation—that is, $P_{\text{rm}} = (\gamma - 1)\rho_{\text{rm}}$ with a standard formalism (see Methods)—we find $\gamma = \langle \{1.328\} \rangle \{-0.007\}^{+0.008\}$, which agrees with the standard value of $\gamma = 4/3$ expected in Λ CDM. At the same time, we find an effective dark energy equation of state parameter $w_{\text{eff}} = P_{\text{de}}/\rho_{\text{de}} = -1$.

$(\{1.011\}_{-0.017}^{+0.018})$, which is consistent with the $w = -1$ expectation for a dark energy density that does not evolve with time.

Fig. 4: Measurements of the CMB temperature as a function of redshift^{3,17,19,20,21,22,23,24,25,26,27,28}.

 figure 4



a, 1σ (black) and 2σ r.m.s. (grey) uncertainties are shown for HFLS3 and 1σ r.m.s. uncertainties elsewhere. **b**, Same data but dividing out the $(1+z)$ redshift scaling of the CMB expected from Λ CDM. Previous direct measurements are from CMB mapping at $z=0$ and SZ effect measurements of galaxy clusters in contrast to the CMB out to $z \sim 1$ (left zoom-in panel in **b**). Additional measurements are from ultraviolet absorption systems along the lines of sights to quasars out to $z \sim 3$. The downward (upward) triangles are not corrected (corrected) for the contribution of collisional excitation in the diffuse interstellar medium to the excitation temperature T_{ex} of the tracer (right zoom-in panel in **b**; green dots show an alternative proposed correction^{3,21,27}). The separation of these pairs of points for the same sources is an indication of the systematic uncertainties on top of the statistical uncertainties indicated by the error bars. The H₂O-based measurement of HFLS3, like those up to $z \sim 1$, is in contrast to the CMB, but—as a line measurement—it is more precise in redshift. It is not subject to the same uncertainties in T_{ex} as the intermediate redshift measurements, because collisions can only decrease (rather than boost) the resulting absorption strength into the CMB for the H₂O-based measurement. They are also unlikely to play an important role due to the high density required to collisionally excite the relevant H₂O lines. Ignoring collisions results in the most conservative estimate of T_{CMB} for HFLS3. The orange shaded region shows a $T_{\text{CMB}} = T_{\text{CMB}}(z=0)^*(1+z)^{1-\beta}$ fit to the data in Extended Data Table 1 and its uncertainty (where $T_{\text{CMB}}(z=$

$0) = 2.72548 \pm 0.00057$ K (ref. 5) and $\beta = ((3.4)_{-7.3}^{+8.1}) \times 10^{-3}$), the orange dashed line indicates the $\beta = 0$ case corresponding to the standard cosmology and the dotted lines indicate a $\pm 10\%$ deviation in $1 - \beta$.

Methods

NOEMA observations

The target was observed in the 3-mm wavelength band 1 (rest-frame 400 μm) with NOEMA as part of project S20DA (Principal Investigators: D. A. Riechers, F. Walter). Three partially overlapping spectral setups were observed under good weather conditions between 26 July 2020 and 25 August 2020 with ten antennas in the most compact D configuration, using a bandwidth of 7.7 GHz (dual polarization) at 2-MHz spectral resolution per sideband. We also included previously published⁵ observations between 6 February 2012 and 31 May 2012 in the A and D configurations tuned to 110.128 and 113.819 GHz, respectively, and previously unpublished observations between 1 June 2012 and 4 June 2012 and on 10 July 2017 in the D configuration tuned to 78.544 and 101.819 GHz taken as part of projects V0BD, W058, and S17CC (Principal Investigator: D. A. Riechers), all using 3.6 GHz of bandwidth (dual polarization), yielding 21 observing runs in total. Nearby radio quasars were used for complex gain, bandpass and absolute flux calibration. The target was also observed in the 0.87-mm wavelength band 4 (rest-frame 122 μm) with NOEMA as part of project X0CC (Principal Investigator: D. A. Riechers). Observations were carried out during three observing runs with six antennas in the A and C configurations under good weather conditions between 4 December 2013 and 12 March 2015, with the band 4 receivers tuned to 335.5 GHz and using a bandwidth of 3.6 GHz (dual polarization). Nearby radio quasars were used for complex gain, bandpass and absolute flux calibration. The GILDAS package was used for data calibration and imaging. All 3-mm data were combined to a single visibility cube before imaging. Imaging was carried out with natural baseline weighting. The band 4 data were also imaged with Briggs robust weighting to increase the spatial resolution. A map of the continuum emission at the frequency of the H₂O line was created by averaging the visibility data over a bandwidth of 2.04 GHz centred on the line. This range was chosen to avoid other lines in the bandpass. Continuum emission was subtracted from the H₂O line cube in the visibility plane. Moment 0 images of the line absorption were created before and after continuum subtraction by integrating the signal over a bandwidth of 100 MHz, corresponding to 395 km s⁻¹. The resulting r.m.s. noise levels are provided in Extended Data Fig. 2. We also make use of previously published⁵ rest-frame 158- μm NOEMA data, which were adopted without further modification.

Line and continuum parameters

The flux of the H₂O(1₁₀–1₀₁) line was extracted by simultaneous Gaussian fitting of the line and continuum emission (including a linear term for the continuum) in the one-dimensional spectrum shown in Fig. 1, which was extracted from the image cube. The source is unresolved at the frequency of the H₂O(1₁₀–1₀₁) line, such that the main uncertainties are due to the slope of the continuum emission and the appropriate fitting of other nearby lines, in particular, CO(5–4). The uncertainties in these parameters are part of the quoted uncertainties. We find a line peak flux of $-818 \pm 145 \mu\text{Jy}$ at a line full width half maximum (FWHM) of $507 \pm 111 \text{ km s}^{-1}$, centred at a frequency of 75.8948 GHz ($\pm 46 \text{ km s}^{-1}$; the calibration uncertainties on the line FWHM and centre frequency are negligible and that on the line peak flux is $< 10\%$ —that is, minor compared with the measurement uncertainty). Given the rest frequency of the line of 556.9359877 GHz, this corresponds to a redshift of 6.3383, which is consistent with the systemic redshift of HFLS3 ($z = 6.3335$ and 6.3427 with uncertainties of ± 14 and $\pm 54 \text{ km s}^{-1}$ at Gaussian FWHM of 243 ± 39 and $760 \pm 152 \text{ km s}^{-1}$, respectively, for the two velocity components detected in the 158-μm [CII] line)⁵. For comparison, the H₂O(2₀₂–1₁₁) and H₂O(2₁₁–2₀₂) emission lines in HFLS3 have FWHM of 805 ± 129 and $927 \pm 330 \text{ km s}^{-1}$, respectively⁵—that is, only marginally broader than the 1₁₀–1₀₁ line at the current measurement uncertainties. The continuum flux at the line frequency is $396 \pm 15 \mu\text{Jy}$, corresponding to $48\% \pm 9\%$ of the absorption-line flux (the relative flux calibration uncertainty between the line and continuum emission is negligible). We also measured the 335.5-GHz continuum flux by two-dimensional fitting to the continuum emission in the visibility plane. We find a flux of $33.9 \pm 1.1 \text{ mJy}$, which agrees with previous lower-resolution observations at the same wavelength⁵. The major (minor) axis FWHM diameter of the source is $0.617 \pm 0.074 \text{ arcsec}$ ($0.37 \pm 0.20 \text{ arcsec}$). This yields the physical source size quoted in the main text at the redshift of HFLS3.

Brightness temperature contrast

The H₂O(1₁₀–1₀₁) line leads to a decrement in continuum photons from the starburst and, as such, is observed as a lack of continuum emission at its frequency at the position of the starburst. It therefore appears as negative flux in an image where starburst continuum emission has been subtracted. In addition, (sub)millimetre-wavelength interferometric images reveal structure against a flat sky background defined by the large-scale CMB surface brightness, which the interferometer does not detect itself due to its limited spatial sampling. Therefore the fraction of the signal due to the decrement in CMB photons at the position of the starburst not only appears as negative flux without subtracting any further signal but it also corresponds to a lack of continuum emission at the line frequency in practice. As the mere presence of an absorption-line signal stronger than the measured continuum emission implies

absorption against the CMB, this interpretation is not limited by uncertainties in the galaxy continuum flux or uncertainties in the absolute flux calibration.

Line-excitation modelling

RADEX is a radiative transfer program to analyse interstellar line spectra by calculating the intensities of atomic and molecular lines, assuming statistical equilibrium and considering collisional and radiative processes, as well as radiation from background sources. Optical depth effects are treated with an escape probability method⁸. Studies of nearby star-forming galaxies show that the observed absorption strengths of the ground-state H₂O and H₂O⁺ transitions are due to cooler gas that is located in front of, and irradiated by, a warmer background source that is emitting the infrared continuum light that also excites the higher-level H₂O emission lines^{11,29}. We therefore adopt the same geometry for the modelling in this work, which is adequately treated within RADEX (that is, treating the dust continuum plus the CMB as background fields for the absorbing material)⁸. The dust continuum emission is modelled as a grey body with treating T_{dust} , β_{IR} and the wavelength where the dust optical depth reaches unity as free-fitting parameters for each dust continuum size and T_{CMB} sampled by the models. The observed spectral energy distribution of HFLS3, including all literature⁵ photometry and the measurements presented in this work, is then treated as the contrast between the dust continuum and CMB background fields, such that the resulting fit parameters for the dust continuum source change with T_{CMB} in a self-consistent manner. In the RADEX models, we derive the H₂O peak absorption depth into the CMB. We then multiply the best matching peak absorption depth found by RADEX with a Gaussian matched to the fitted line centroid and line width obtained from the observed line profile in Fig. 2 to determine the model line profile. In this approach, the shallower absorption in the line wings either corresponds to a lower filling factor of the H₂O layer at the corresponding velocities or to lower H₂O column densities. Although collisions of H₂O molecules with H₂ is another mechanism that can modify the level populations especially at very high gas densities (which is an important mechanism for the cooling of low-excitation-temperature transitions of molecules like H₂CO to below T_{CMB})^{12,30}, the RADEX models show that they do not affect our findings (see Fig. 3c). We therefore adopt models with essentially no collisions by assuming a very low gas density of $n(\text{H}_2) = 10 \text{ cm}^{-3}$. We then compare our findings to those obtained when adopting conditions that are similar to those found in local starburst galaxies¹¹ and to those found for high-density environments with $n(\text{H}_2) > 10^5 \text{ cm}^{-3}$. The cross sections for collisions out of the 1₀₁ level are always larger than those out of the 1₁₀ level, independent of the collision partner and the temperature at which the collisions take place^{31,32,33}. Therefore

collisions cannot be responsible for an over-proportional de-population of the 1_{10} level relative to the 1_{01} ground state, and the net effect of including collisions is a decrease in the absorption depth into the CMB by reducing the $T_{\text{CMB}} - T_{\text{ex}}$ temperature difference at very high gas densities compared with cases without collisions. For reference, the effect of collisions on the determination of T_{CMB} is negligible for the typical conditions found in local starbursts (that is, $n(\text{H}_2) \sim 10^4 \text{ cm}^{-3}$; $T_{\text{kin}} = 20\text{--}180 \text{ K}$)¹¹ and only starts to have an impact for very high densities $n(\text{H}_2) > 10^5 \text{ cm}^{-3}$. For a given continuum source size, the constraints on T_{CMB} would therefore be tighter (that is, would more quickly become inconsistent with the observations) for the high-density case than for the case without collisions, such that the latter approach is more conservative (see Fig. 3b). The overall impact of collisional excitation would therefore be more stringent requirements on the source size, covering fraction and water column, such that their inclusion would only further strengthen our conclusions. We note that this is the opposite effect to the case of the studies of ultraviolet lines^{3,17,19,20,21,22,23,24,25,26,27,28}, where neglecting collisional excitation results in less conservative constraints on T_{CMB} . If we were to assume that the H_2O absorption were to emerge from within the infrared continuum-emitting region, a larger source size would probably be required to obtain the same absorption-line strength due to a reduced effective radiation field strength from the starburst. Previous modelling attempts of nearby galaxies assuming such geometries have not been able to produce $\text{H}_2\text{O}(1_{10}\text{--}1_{01})$ line absorption on the scales necessary to explain the observations of HFLS3, which may indicate that even more complex assumptions would be required¹¹. Thus, the resulting constraints would, once again, be less conservative, perhaps acting in a similar manner as the high-density case. Excluding both of these effects from the models leads to a maximally conservative estimate of T_{CMB} and its uncertainties. Assuming a plane-parallel or similar geometry instead of a spherical geometry would only have a minor impact on our findings⁸. The models shown in Fig. 3 assume a filling factor of unity, which is the most conservative possible assumption. A more clumpy geometry with a lower covering fraction remains possible for all T_{CMB} values for which the predicted absorption strength exceeds the observed value (see shaded regions in Fig. 3b). For reference, the minimum covering fractions consistent with the continuum size at the observed signal strength are shown for the different cases considered in Fig. 3. The line absorption is also found to be optically thick, with an optical depth of $\tau_{\text{H}_2\text{O}} = 21.1$ for the solution shown in Fig. 2b. To determine the redshift above which the effect becomes observable (Fig. 3c), we fixed $r_{108\mu\text{m}}$, T_{dust} , β_{IR} and M_{dust} to the observed values and the H_2O column density to the value corresponding to the model spectrum. H_2O line absorption into the dust continuum of HFLS3 would already become visible at $z > 2.9$, but absorption into the CMB only becomes observable at $z > 4.5$ (or higher for H_2 densities of $> 10^5 \text{ cm}^{-3}$). These values

account for changes in the shape of the dust grey-body spectrum (that is, changes in the relative availability of 538- μm and 108- μm photons) due to changes in T_{CMB} with redshift. To better quantify the impact of different modelling parameters, we have varied T_{dust} and β_{IR} beyond their previously estimated uncertainties (nominal reference values without considering variations in T_{CMB} from the literature are $T_{\text{dust}} = \langle \{63.3\} \{-5.8\}^{+5.4} \rangle$ K and $\beta_{\text{IR}} = \langle \{1.94\} \{-0.09\}^{+0.07} \rangle$)^{5,6}. This is necessary because both parameters are dependent on the varying T_{CMB} in our models (and therefore are changing parameters in Fig. 3b,c), such that their true uncertainties need to be re-evaluated. We independently varied β_{IR} in the 1.6–2.4 range and T_{dust} in the ± 20 K range as functions of T_{CMB} around the best-fit values. This shows that $\beta_{\text{IR}} > 2.0$ and T_{dust} lower by more than 10 K from the best fits yield very poor fits to the spectral energy distribution data, whereas $\Delta\beta_{\text{IR}} > -0.1$ below the best-fit value would require a larger continuum size than the measured $r_{108\mu\text{m}} + 1\sigma$ and therefore are disfavoured by the size constraint. Excluding these ranges, the extrema across this entire range would extend the uncertainty range in the predicted T_{CMB} by only -1.7 and $+5.4$ K and -0.8 and $+4.4$ K for the $r_{108\mu\text{m}} + 1\sigma$ and $r_{108\mu\text{m}} + 2\sigma$ cases, respectively. For comparison, the difference between the $+1\sigma$ and $+2\sigma$ uncertainty ranges is -3.6 and $+3.8$ K). This shows that the impact of the uncertainties in the dust spectral energy distribution fitting parameters on those in T_{CMB} are subdominant to those in the continuum size measurement. Conversely, we have studied the impact of changes in T_{CMB} on the best-fit T_{dust} and β_{IR} . For the values corresponding to $r_{108\mu\text{m}} + 1\sigma$ and $r_{108\mu\text{m}} + 2\sigma$ ranges, T_{dust} typically changes by <0.5 K and β_{IR} typically changes by <0.1 –0.2 when varying the parameters independently. These changes are larger than the actual uncertainties, because the fit to the dust spectral energy distribution becomes increasingly poorer with these single-parameter variations. At the same time, these changes are subdominant to those induced by changes in dust continuum size within the $+1\sigma$ and $+2\sigma$ uncertainty ranges, which is consistent with our other findings.

Other H₂O transitions in HFLS3

Five H₂O lines were previously detected towards HFLS3 ($2_{02}-1_{11}$, $2_{11}-2_{02}$, $3_{12}-2_{21}$, $3_{12}-3_{03}$ and $3_{21}-3_{12}$) and two additional lines were tentatively detected ($4_{13}-4_{04}$ and $4_{22}-4_{13}$)⁵. The $J_{\text{up}} = 3$ transitions are due to ortho-H₂O and all other transitions are due to para-H₂O. All of these transitions appear in emission. Given the high critical densities of these transitions, our RADEX models cannot reproduce the strength of these lines as the same time as the observed ortho-H₂O($1_{10}-1_{01}$) absorption strength, which suggests that they emerge from different gas components. For reference, to reproduce the strength of the H₂O($2_{11}-2_{02}$) in Fig. 1 alone with collisional excitation,

$n(\text{H}_2) = 2 \times 10^7 \text{ cm}^{-3}$ and $T_{\text{kin}} = 200 \text{ K}$ would be required, but the $\text{H}_2\text{O}(1_{10}-1_{01})$ would no longer appear in absorption against the CMB if it were to emerge from the same gas component. This is consistent with the picture that the H_2O absorption is due to a cold gas component along the line of sight to the warm gas that gives rise to the emission lines¹¹. Observations of the para- $\text{H}_2\text{O}(1_{11}-0_{00})$ ground state do not currently exist for HFLS3, but our models do not show this line in absorption towards the CMB.

Origin of the lower and upper limits on T_{CMB}

Our models show that the lower limit on T_{CMB} at a given redshift based on the observed H_2O absorption is due to the minimum ‘seed’ level population due to the CMB black-body radiation field. To determine a conservative lower limit, we have calculated models with continuum sizes up to $r_{108\mu\text{m}} = 5 \text{ kpc}$ (see Fig. 3b), corresponding to a $+7.5\sigma$ deviation from the observed continuum size, and recorded the temperatures at which such weakly constrained models turn into absorption. We find that this results in a lower limit of $T_{\text{CMB}} > 7\text{--}8 \text{ K}$, independent of the model assumptions. This finding alone does not explain the existence of an upper limit in Fig. 3b. For a given size of the dust continuum emission, an increase in T_{CMB} also requires an increase in M_{dust} to still reproduce the observed dust spectral energy distribution, which leads to an effective increase in the dust optical depth at a given wavelength. The result of a rising optical depth is that the grey-body spectrum between 538 and $108 \mu\text{m}$ increasingly resembles a black-body spectrum and, hence, a decrease in the H_2O absorption against the CMB. This effect is responsible for the upper limit in allowed T_{CMB} for a given dust continuum size and absorption strength.

Uncertainties of T_{CMB} measurements

The uncertainties shown for the literature data in Fig. 4 are adopted from the literature sources without modification, and they typically represent the statistical uncertainties from the individual measurements or sample averages. Individual cluster measurements of the thermal SZ effect may be affected by dust associated with foreground galaxies or the Milky Way, the galaxy clusters or background galaxies that may be amplified by gravitational lensing, uncertainties in the reconstruction of the Compton- γ parameter maps due to flux uncertainties, radio emission due to active galactic nuclei and/or relics, the kinetic and relativistic SZ effects, and general bandpass and calibration uncertainties¹⁷. Furthermore, uncertainties on the cluster geometry—and therefore line-of-sight travel distance of the CMB photons through the cluster—and on the temperature of the intra-cluster gas limit the precision of individual SZ measurements. Sample averages may also be affected by systematics in

the stacking procedures. Individual data points deviate by up to at least two standard deviations from the trend, which may indicate residual uncertainties beyond the statistical error bars provided, such that the error bars shown in Fig. 4 are underestimated. The main source of uncertainty for the ultraviolet absorption-line-based measurements are due to the assumption of no collisional excitation, which is not taken into account in the statistical uncertainties shown in Fig. 4. Attempts to take this effect into account appear to suggest substantially larger uncertainties than indicated by individual error bars²⁷ (Fig. 4). To expand on earlier estimates²¹, we have calculated RADEX models for typical T_{kin} , $n(\text{H})$ and column densities found from [CI] measurements in the diffuse interstellar medium³⁴, which suggests that collisional excitation contributes to the predicted T_{ex} of the lower fine-structure transition.

Although we show the original unmodified data, the ultraviolet-based measurements are therefore subject to uncertainties due to model-dependent excitation corrections in addition to the statistical uncertainties. Furthermore, the fine-structure levels of tracers like the [CI] lines can be excited by ultraviolet excitation and following cascades. To constrain T_{CMB} based on these measurements, the kinetic temperature, particle density and local ultraviolet radiation field must be known, and are typically determined based on tracers other than the species used to constrain T_{CMB} . Also, some measurements are based on spectrally unresolved lines, which limits the precision of kinetic temperature measurements based on thermal broadening²¹. Owing to these uncertainties, the ultraviolet absorption-line-based measurements are probably consistent with the standard Λ CDM value, but they do not constitute a direct measurement of T_{CMB} without notable further assumptions. For reference, the median $T_{\text{CMB}}/(1+z)$ estimate based on the [CI] measurements alone (excluding upper limits) is 3.07 K, with a median absolute deviation of 0.09 K and a standard deviation of 0.31 K. Therefore the current sample median deviates from the Λ CDM value by about one standard deviation. A combination of the (uncorrected) measurements based on CO, [CI] and [CII] provides a median value of 2.84 K, with a median absolute deviation of 0.15 K and a standard deviation of 0.25 K. This highlights the importance of the corrections discussed above and in the literature and the value of measurements with systematic uncertainties that differ from this method to obtain a more complete picture. The main source of uncertainty of the H₂O-based measurements, beyond the caveats stated in the line-excitation-modelling section, are the statistical uncertainties on the source size, the lack of a direct measurement of the absorbing H₂O column density, variations in the dust mass absorption coefficient and the filling factor. Given the high metallicity suggested by other molecular line detections, the limitation to high filling factors due to the source size and the constraint on the gas mass from dynamical mass measurements, the main source of uncertainty resides in the source size due to limited spatial resolution in the current data. As such, major improvements should be possible by obtaining higher, (sub-)kpc resolution (that is, <0.2'') imaging with the Atacama Large Millimeter/submillimeter Array (ALMA; for other targets) and planned

upgrades to NOEMA, and, in the future, with the Next Generation Very Large Array (ngVLA). Statistical uncertainties will also be greatly reduced by observing larger samples of massive star-forming galaxies over the entire redshift range where measurements are possible, closing the gap to SZ-based studies, which are currently limited to $z < 1$. The resulting improvement in precision will provide the constraints that are necessary to confirm or challenge the evolution of the CMB temperature with redshift predicted by standard cosmological models.

Accessibility of the line signal

The frequency range currently covered by NOEMA is 70.4–119.3, 127.0–182.9 and 196.1–276.0 GHz (with greatly reduced sensitivity above about 115 and 180 GHz in the first two frequency ranges). ALMA covers the 84–500-GHz range with gaps at 116–125 and 373–385 GHz, with a future extension down to 65 GHz (with greatly reduced sensitivity below about 67 GHz). The ngVLA is envisioned to cover the 70–116-GHz range. Excluding regions of poor atmospheric transparency, the H₂O(1₁₀–1₀₁) line is therefore observable in these frequency ranges at redshifts of $z = 0.1$ –0.4, 0.5–2.0, 2.1–3.4 and 3.8–6.9 in principle, but the detectability of the line in absorption against the CMB is probably limited to the $z \sim 4.5$ –6.9 range if the spectral energy distribution shape of HFLS3 is representative. At lower frequencies, the Karl G. Jansky Very Large Array and, in the future, ALMA and the ngVLA also provide access to the <52-GHz range, such that the signal also becomes observable at $z > 9.7$ in principle. In conclusion, the absorption of the ground-state H₂O transition against the CMB identified here could be traced from the ground towards star-forming galaxies across most of the first approximately 1.5 billion years of cosmic history.

Detectability of the line signal for different spectral energy distribution shapes

To investigate whether the effect is expected to be detectable for different galaxy populations, we have applied our modelling to the $z = 3.9$ quasar APM 08279+5255, for which the dust spectral energy distribution is composed of a dominant 220-K dust component and a weaker 65-K dust component, contributing only 10–15% to the far-infrared luminosity^{35,36,37,38,39,40,41,42,43,44,45,46}. The models suggest that the line is expected to occur in emission and that it would not be expected to be detectable in absorption at any redshift out to at least $z = 12$ in galaxies with similar dust spectral energy distributions. Other far-infrared-luminous, high-redshift, active galactic nucleus host galaxies typically show a stronger relative contribution of their lower-temperature dust components, such that the effect may remain detectable in less extreme cases. For galaxies with lower dust temperatures than HFLS3, the effect may be present even at lower redshifts, but is typically expected to be weaker in general

and to disappear at redshifts where T_{CMB} approaches their T_{dust} . For a dust spectral energy distribution shape resembling the central region of the Milky Way but otherwise similar properties, the effect is expected to be reduced by more than two orders of magnitude at its redshift peak, and to become virtually unobservable at the redshift of HFLS3. Thus, dusty starburst galaxies appear to be some of the best environments to detect the effect.

Derivation of equation of state parameters

To determine the adiabatic index, we assume a standard Friedmann–Lemaître–Robertson–Walker cosmology with zero curvature and a matter–radiation fluid that follows the standard adiabatic equation of state quoted in the main text. This would correspond to a redshift scaling $T_{\text{CMB}}(z) = T_{\text{CMB}}(z=0)*(1+z)^{3(\gamma-1)}$ in the presence of a dark energy density that does not scale with redshift. The dark energy density is parameterized to scale with a power law $(1+z)^m$, where $m=0$ corresponds to a cosmological constant. With standard assumptions, this yields a redshift scaling of T_{CMB} (ref. [15](#)):

$$\text{\$}\$\{T\}_{\text{CMB}}(z)=\{T\}_{\text{CMB}}(z=0)\{(1+z)\}^{3(\gamma-1)}\{\left[\frac{(m-3)\varOmega_{m,0}+m(1+z)^{(m-3)}(\varOmega_{m,0}-1)}{\varOmega_{m,0}}\right]^{(m-3)}\}^{(\gamma-1)}\$\$$$

and an effective dark energy equation of state $P_{\text{de}} = w_{\text{eff}}\rho_{\text{de}}$, where the effective equation of state parameter $w_{\text{eff}} = (m/3) - 1$. This fitting function is used here with a canonical value of $\varOmega_{m,0} = 0.315$ (ref. [4](#)). The uncertainty of $\varOmega_{m,0}$ is small compared with all other sources of uncertainty and, hence, is neglected. All data used in the fitting are provided in Extended Data Table 1 (refs. [36,37,38,39,40,41,42,43,44,45,46](#)).

Data availability

The spectral line data and model generated and analysed during this study as shown in Fig. [2](#) are linked to this manuscript in spreadsheet form. Additional versions of the NOEMA datasets (visibilities, images and spectra) are available from the corresponding author (D.A.R.) on reasonable request. All data are also available in the IRAM Science Data Archive (isda@iram.fr) under project IDs V0BD, W058, X0CC, S17CC, and S20DA. [Source data](#) are provided with this paper.

Code availability

The RADEX code used for the modelling presented in this work and shown in Fig. 3 is available at <https://home.strw.leidenuniv.nl/~moldata/radex.html>.

References

1. 1. Sunyaev, R. A., & Zel'dovich, I. A. B. Microwave background radiation as a probe of the contemporary structure and history of the universe. *Annu. Rev. Astron. Astrophys.* **18**, 537–560 (1980).
2. 2. Rephaeli, Y. Comptonization of the cosmic microwave background: the Sunyaev-Zeldovich effect. *Annu. Rev. Astron. Astrophys.* **33**, 541–580 (1995).
3. 3. Klimenko, V. V. et al. Estimation of the cosmic microwave background temperature from atomic C I and molecular CO lines in the interstellar medium of early galaxies. *Astron. Lett.* **46**, 715–725 (2020).
4. 4. Planck Collaboration. Planck 2018 results. I. Overview and the cosmological legacy of Planck. *Astron. Astrophys.* **641**, A1 (2020).
5. 5. Riechers, D. A. et al. A dust-obscured massive maximum-starburst galaxy at a redshift of 6.34. *Nature* **496**, 329–333 (2013).
6. 6. Riechers, D. A. et al. Rise of the titans: gas excitation and feedback in a binary hyper-luminous dusty starburst galaxy at $z \sim 6$. *Astrophys. J.* **907**, 62 (2021).
7. 7. Fixsen, D. J. The temperature of the cosmic microwave background. *Astrophys. J.* **707**, 916 (2009).
8. 8.

van der Tak, F. F. S. et al. A computer program for fast non-LTE analysis of interstellar line spectra. With diagnostic plots to interpret observed line intensity ratios. *Astron. Astrophys.* **468**, 627–635 (2007).

9. 9.

Gonzalez-Alfonso, E. et al. Herschel/PACS spectroscopy of NGC 4418 and Arp 220: H₂O, H₂¹⁸O, OH, ¹⁸OH, O I, HCN, and NH₃. *Astron. Astrophys.* **541**, A4 (2012).

10. 10.

Gonzalez-Alfonso, E. et al. Modeling the H₂O submillimeter emission in extragalactic sources. *Astron. Astrophys.* **567**, A91 (2014).

11. 11.

Liu, L. et al. HIFI spectroscopy of H₂O submillimeter lines in nuclei of actively star-forming galaxies. *Astrophys. J.* **846**, 5 (2017).

12. 12.

Darling, J. & Zeiger, B. Formaldehyde silhouettes against the cosmic microwave background: a mass-limited, distance-independent, extinction-free tracer of star formation across the epoch of galaxy evolution. *Astrophys. J. Lett.* **749**, L33 (2012).

13. 13.

Pavesi, R. et al. The CO Luminosity Density at High-*z* (COLDz) survey: a sensitive, large-area blind search for low-J CO emission from cold gas in the early universe with the Karl G. Jansky Very Large Array. *Astrophys. J.* **864**, 49 (2018).

14. 14.

Uzan, J.-P. Varying constants, gravitation and cosmology. *Living Rev. Relativ.* **14**, 2 (2011).

15. 15.

Jetzer, P. et al. Limits on decaying dark energy density models from the CMB temperature–redshift relation. *Gen. Relativ. Gravitation* **43**, 1083–1093 (2011).

16. 16.

Jaeckel, J. & Ringwald, A. The low-energy frontier of particle physics. *Annu. Rev. Nucl. Part. Sci.* **60**, 405–437 (2010).

17. 17.

Hurier, G. et al. Measurement of the T_{CMB} evolution from the Sunyaev-Zel'dovich effect. *Astron. Astrophys.* **561**, A143 (2014).

18. 18.

Burgh, E. B. et al. Direct measurement of the ratio of carbon monoxide to molecular hydrogen in the diffuse interstellar medium. *Astrophys. J.* **658**, 446–454 (2007).

19. 19.

Songaila, A. et al. Measurement of the microwave background temperature at a redshift of 1.776. *Nature* **371**, 43–45 (1994).

20. 20.

Ge, J., Bechtold, J., & Black, J. H. A new measurement of the cosmic microwave background radiation temperature at $z = 1.97$. *Astrophys. J.* **474**, 67 (1997).

21. 21.

Srianand, R., Petitjean, P., & Ledoux, C. The cosmic microwave background radiation temperature at a redshift of 2.34. *Nature* **408**, 931–935 (2000).

22. 22.

Molaro, P. et al. The cosmic microwave background radiation temperature at $z_{\text{abs}} = 3.025$ toward QSO 0347-3819. *Astron. Astrophys. Lett.* **381**, L64–L67 (2002).

23. 23.

Noterdaeme, P. et al. The evolution of the cosmic microwave background temperature. Measurements of T_{CMB} at high redshift from carbon monoxide excitation. *Astron. Astrophys.* **526**, L7 (2011).

24. 24.

Saro, A. et al. Constraints on the CMB temperature evolution using multiband measurements of the Sunyaev–Zel'dovich effect with the South Pole Telescope. *Mon. Not. R. Astron. Soc.* **440**, 2610–2615 (2014).

25. 25.

de Martino, I. et al. Constraining the redshift evolution of the cosmic microwave background blackbody temperature with Planck data. *Astrophys. J.* **808**, 128 (2015).

26. 26.

Avgoustidis, A. et al. Subpercent constraints on the cosmological temperature evolution. *Phys. Rev. D* **93**, 043521 (2016).

27. 27.

Maeder, A. Scale-invariant cosmology and CMB temperatures as a function of redshifts. *Astrophys. J.* **847**, 65 (2017).

28. 28.

Noterdaeme, P. et al. Discovery of a Perseus-like cloud in the early Universe. H I-to-H₂ transition, carbon monoxide and small dust grains at $z_{\text{abs}} \approx 2.53$ towards the quasar J0000+0048. *Astron. Astrophys.* **597**, A82 (2017).

29. 29.

Weiss, A. et al. HIFI spectroscopy of low-level water transitions in M 82. *Astron. Astrophys.* **521**, L1 (2010).

30. 30.

Townes, C. H. Astronomical masers and lasers. *Quantum Electron.* **27** 1031 (1997).

31. 31.

Mueller, H. S. P. et al. The Cologne database for molecular spectroscopy, CDMS. *Astron. Astrophys.* **370**, L49–L52 (2001).

32. 32.

Dubernet, M.-L. et al. Rotational excitation of ortho-H₂O by para-H₂ ($j_z = 0, 2, 4, 6, 8$) at high temperature. *Astron. Astrophys.* **497**, 911–925 (2009).

33. 33.

Daniel, F. et al. Rotational excitation of 45 levels of ortho/para-H₂O by excited ortho/para-H₂ from 5 K to 1500 K: state-to-state, effective, and thermalized rate coefficients. *Astron. Astrophys.* **536**, A76 (2011).

34. 34.

Jenkins, E. B., & Tripp, T. M. The distribution of thermal pressures in the diffuse, cold neutral medium of our galaxy. II. An expanded survey of interstellar C I fine-structure excitations. *Astrophys. J.* **734**, 65 (2011).

35. 35.

Weiss, A. et al. Highly-excited CO emission in APM 08279+5255 at $z = 3.9$. *Astron. Astrophys.* **467**, 955–969 (2007).

36. 36.

Cui, J. et al. Molecular hydrogen in the damped Ly α absorber of Q1331+170. *Astrophys. J.* **633**, 649 (2005).

37. 37.

Ledoux, C., Petitjean, P., & Srianand, R. The Very Large Telescope Ultraviolet and Visible Echelle Spectrograph survey for molecular hydrogen in high-redshift damped Lyman α systems. *Mon. Not. R. Astron. Soc.* **346**, 209–228 (2003).

38. 38.

Balashev, S. A. et al. Partial coverage of the broad-line region of Q1232+082 by an intervening H₂-bearing cloud. *Mon. Not. R. Astron. Soc.* **418**, 357–369 (2011).

39. 39.

Srianand, R. et al. First detection of CO in a high-redshift damped Lyman- α system. *Astron. Astrophys. Lett.* **482**, L39–L42 (2008).

40. 40.

Ranjan, A. et al. Molecular gas and star formation in an absorption-selected galaxy: hitting the bull's eye at $z \approx 2.46$. *Astron. Astrophys.* **618**, A184 (2018).

41. 41.

Noterdaeme, P. et al. Spotting high-z molecular absorbers using neutral carbon: results from a complete spectroscopic survey with the VLT. *Astron. Astrophys.* **612**, A58 (2018).

42. 42.

Balashev, S. A., Ivanchik, A. V., & Varshalovich, D. A. HD/H₂ molecular clouds in the early Universe: the problem of primordial deuterium. *Astron. Lett.* **36**, 761–772 (2010).

43. 43.

Noterdaeme, P. et al. A translucent interstellar cloud at $z = 2.69$: CO, H₂, and HD in the line-of-sight to SDSS J123714.60 + 064759.5. *Astron. Astrophys.* **523**, A80 (2010).

44. 44.

Balashev, S. A. et al. CO-dark molecular gas at high redshift: very large H₂ content and high pressure in a low-metallicity damped Lyman alpha system. *Mon. Not. R. Astron. Soc.* **470**, 2890–2910 (2017).

45. 45.

Jorgenson, R. A., Wolfe, A. M., & Prochaska, J. X. Understanding physical conditions in high-redshift galaxies through C I fine structure lines: data and methodology. *Astrophys. J.* **722**, 460 (2010).

46. 46.

Guimaraes, R. et al. Metallicities, dust, and molecular content of a QSO-damped Ly α system reaching $\log N(\text{H I}) = 22$: an analog to GRB-DLAs. *Astron. J* **143**, 147 (2012).

Acknowledgements

D.A.R. acknowledges support from the National Science Foundation under grant numbers AST-1614213 and AST-1910107. D.A.R. also acknowledges support from

the Alexander von Humboldt Foundation through a Humboldt Research Fellowship for Experienced Researchers. This work is based on observations carried out under project numbers V0BD, W058, X0CC, S17CC, and S20DA with the IRAM NOEMA interferometer. IRAM is supported by INSU/CNRS (France), MPG (Germany) and IGN (Spain).

Funding

Open access funding provided by Universität zu Köln.

Author information

Affiliations

1. I. Physikalisches Institut, Universität zu Köln, Köln, Germany
Dominik A. Riechers
2. Max-Planck-Institut für Radioastronomie, Bonn, Germany
Axel Weiss
3. Max-Planck-Institut für Astronomie, Heidelberg, Germany
Fabian Walter
4. National Radio Astronomy Observatory, Pete V. Domenici Array Science Center, Socorro, NM, USA
Fabian Walter & Christopher L. Carilli
5. Sorbonne Université, UPMC Université Paris 6 and CNRS, UMR 7095, Institut d’Astrophysique de Paris, Paris, France
Pierre Cox
6. INAF - Osservatorio di Astrofisica e Scienza dello Spazio, Bologna, Italy
Roberto Decarli
7. Institut de Radioastronomie Millimétrique, Saint Martin d’Héres, France
Roberto Neri

Contributions

D.A.R. led the project and writing of the manuscript. A.W. produced the models used in this work. F.W. was a co-Principal Investigator of the main observing proposal and contributed to the interpretation of the results. R.N. calibrated the data. All authors have reviewed, discussed and commented on the manuscript.

Corresponding author

Correspondence to [Dominik A. Riechers](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Jeremy Darling, Eduardo González-Alfonso and the other, anonymous, reviewer 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 Combined effect of CMB absorption and starburst radiation field on the strength of the H₂O\(1₁₀–1₀₁\) line in HFLS3.](#)

Top, the cold H₂O vapour is exposed to the CMB radiation field⁴, which has the shape of a black-body function ($T_{\text{CMB}} = 20.0 \text{ K}$ at $z = 6.34$), and the starburst infrared radiation field, which has the shape of a grey-body function ($T_{\text{dust}} = 63 \text{ K}$). NOEMA observed the signal in contrast to the CMB and therefore detects only the dust

emission from the starburst and the H₂O line, but not the CMB itself (which therefore fills the region below zero flux density, as seen by the telescope). Bottom left, as the energy-level difference for the H₂O(1₁₀–1₀₁) line is only 26.7 K, there are sufficient CMB photons at $z = 6.34$ to thermalize the level population between both levels, such that T_{ex} is the same as T_{CMB} in equilibrium. Therefore no H₂O emission or absorption will be observed, despite the presence of a ‘seed’ population in the upper level. Bottom right, the radiation field of the starburst alters the level populations towards increased higher-level populations. Owing to the grey-body shape of its spectral energy distribution, more photons are available at 108 μm to increase the 2₂₁ level population from the 1₁₀ state than there are 538-μm photons available to increase the 1₁₀ level population from the 1₀₁ state, relative to the ‘seed’ population provided by the absorption of CMB photons. Therefore the relative population of the 1₁₀ and 1₀₁ levels is lower than in thermal equilibrium, such that the resulting T_{ex} is lower than T_{CMB} . As a result, the H₂O(1₁₀–1₀₁) line is observed in absorption towards the CMB due to the negative temperature contrast—as observed towards HFLS3.

Extended Data Fig. 2 H₂O line emission integrated moment 0 and continuum maps of HF3.

a–c, H₂O contour maps (blue) before (**b**) and after (**c**) continuum subtraction, and local continuum (**a**, green contours) at the wavelength of the H₂O line, overlaid on the 158-μm continuum (intensity scale)⁵. H₂O emission is integrated over the central 395 km s⁻¹ (100 MHz). **d–e**, Rest-frame 122-μm continuum emission (orange contours and intensity scale) as a proxy for the 108-μm continuum size, showing the full emission (**d**), and the compact nuclear region that accounts for two-thirds of the emission at higher resolution (**e**), overlaid with 158-μm contours (red) for orientation. **f**, Radially averaged visibility amplitude as a function of interferometer baseline length for the data in **d** and **e**. The radial profile of the visibility amplitude (binned to 50-m steps, with 1 σ error bars) shows that the 122-μm dust emission is clearly resolved. Observed-frame 538-μm continuum contours (**a**) are shown in steps of $1\sigma = 22.5 \mu\text{Jy beam}^{-1}$, starting at $\pm 3\sigma$. H₂O contours (**b**, **c**) are shown in steps of $1\sigma = 0.0375 \text{ Jy km s}^{-1} \text{ beam}^{-1}$, starting at $\pm 2\sigma$. Contours of 122 μm (**d**, **e**) are shown in steps of $\pm 10\sigma$ and $\pm 5\sigma$, where $1\sigma = 229$ and $374 \mu\text{Jy beam}^{-1}$, respectively. Contours of 158 μm (**d**, **e**) are shown in steps of 3σ , starting at $\pm 5\sigma$, where $1\sigma = 400 \mu\text{Jy beam}^{-1}$ (all uncertainties are r.m.s.). Negative intensity contours are dashed.

Extended Data Table 1 Summary of $T_{\text{CMB}}(z)$ measurements

Source data

Source Data Fig. 2

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

Riechers, D.A., Weiss, A., Walter, F. *et al.* Microwave background temperature at a redshift of 6.34 from H₂O absorption. *Nature* **602**, 58–62 (2022).
<https://doi.org/10.1038/s41586-021-04294-5>

- Received: 12 February 2021
- Accepted: 30 November 2021
- Published: 02 February 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04294-5>

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/s41586-021-04294-5>

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

- Article
- [Published: 02 February 2022](#)

Search for magnetic monopoles produced via the Schwinger mechanism

- [B. Acharya¹](#),
- [J. Alexandre¹](#),
- [P. Benes²](#),
- [B. Bergmann²](#),
- [S. Bertolucci³](#),
- [A. Bevan⁴](#),
- [H. Branzas⁵](#),
- [P. Burian²](#),
- [M. Campbell⁶](#),
- [Y. M. Cho⁷](#),
- [M. de Montigny⁸](#),
- [A. De Roeck⁶](#),
- [J. R. Ellis^{1,9}](#),
- [M. El Sawy⁶](#),
- [M. Fairbairn¹](#),
- [D. Felea⁵](#),
- [M. Frank¹⁰](#),
- [O. Gould^{11,12}](#),
- [J. Hays⁴](#),
- [A. M. Hirt¹³](#),
- [D. L.-J. Ho¹⁴](#),
- [P. Q. Hung¹⁵](#),
- [J. Janecek²](#),

- [M. Kalliokoski](#)¹²,
- [A. Korzenev](#)¹⁶,
- [D. H. Lacarrère](#)⁶,
- [C. Leroy](#)¹⁷,
- [G. Levi](#)^{3,18},
- [A. Lionti](#)¹⁶,
- [A. Maulik](#)^{3,8},
- [A. Margiotta](#)¹⁸,
- [N. Mauri](#)³,
- [N. E. Mavromatos](#)¹,
- [P. Mermod](#)^{16 na1},
- [L. Millward](#)⁴,
- [V. A. Mitsou](#)¹⁹,
- [I. Ostrovskiy](#) ORCID: orcid.org/0000-0003-4939-0225²⁰,
- [P.-P. Ouimet](#)⁸,
- [J. Papavassiliou](#)¹⁹,
- [B. Parker](#)²¹,
- [L. Patrizii](#)³,
- [G. E. Pāvālaš](#)⁵,
- [J. L. Pinfold](#)⁸,
- [L. A. Popa](#)⁵,
- [V. Popa](#)⁵,
- [M. Pozzato](#)³,
- [S. Pospisil](#)²,
- [A. Rajantie](#)¹⁴,
- [R. Ruiz de Austri](#)¹⁹,
- [Z. Sahnoun](#)³,
- [M. Sakellariadou](#)¹,
- [A. Santra](#)¹⁹,
- [S. Sarkar](#)¹,
- [G. Semenoff](#)²²,
- [A. Shaa](#)⁸,
- [G. Sirri](#)³,
- [K. Sliwa](#)²³,

- [R. Soluk⁸](#),
- [M. Spurio¹⁸](#),
- [M. Staelens⁸](#),
- [M. Suk²](#),
- [M. Tenti²⁴](#),
- [V. Togo³](#),
- [J. A. Tuszyn'ski⁸](#),
- [A. Upreti²⁰](#),
- [V. Vento¹⁹](#) &
- [O. Vives¹⁹](#)

[*Nature*](#) volume **602**, pages 63–67 (2022)

- 1239 Accesses
- 125 Altmetric
- [Metrics details](#)

Subjects

- [Experimental particle physics](#)
- [Quantum mechanics](#)

Abstract

Electrically charged particles can be created by the decay of strong enough electric fields, a phenomenon known as the Schwinger mechanism¹. By electromagnetic duality, a sufficiently strong magnetic field would similarly produce magnetic monopoles, if they exist². Magnetic monopoles are hypothetical fundamental particles that are predicted by several theories beyond the standard model^{3,4,5,6,7} but have never been experimentally detected. Searching for the existence of magnetic monopoles via the Schwinger mechanism has not yet been attempted, but it is advantageous, owing to the possibility of calculating its rate through semi-classical techniques without perturbation theory, as well as that the production of the

magnetic monopoles should be enhanced by their finite size^{8,9} and strong coupling to photons^{2,10}. Here we present a search for magnetic monopole production by the Schwinger mechanism in Pb–Pb heavy ion collisions at the Large Hadron Collider, producing the strongest known magnetic fields in the current Universe¹¹. It was conducted by the MoEDAL experiment, whose trapping detectors were exposed to 0.235 per nanobarn, or approximately 1.8×10^9 , of Pb–Pb collisions with 5.02-teraelectronvolt center-of-mass energy per collision in November 2018. A superconducting quantum interference device (SQUID) magnetometer scanned the trapping detectors of MoEDAL for the presence of magnetic charge, which would induce a persistent current in the SQUID. Magnetic monopoles with integer Dirac charges of 1, 2 and 3 and masses up to 75 gigaelectronvolts per speed of light squared were excluded by the analysis at the 95% confidence level. This provides a lower mass limit for finite-size magnetic monopoles from a collider search and greatly extends previous mass bounds.

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: Schematic diagram for the search for Schwinger MMs with MoEDAL.

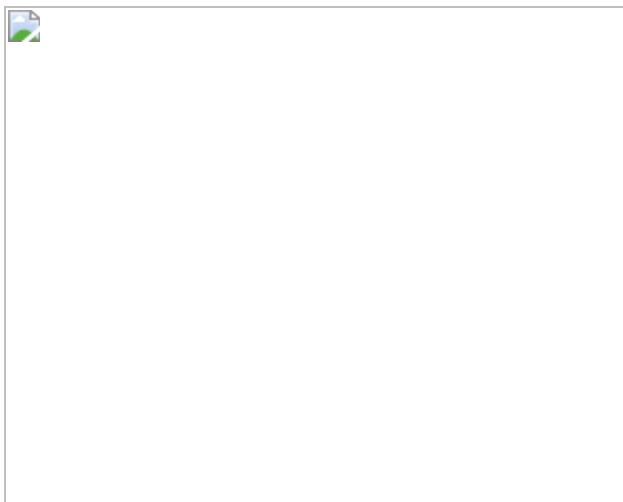
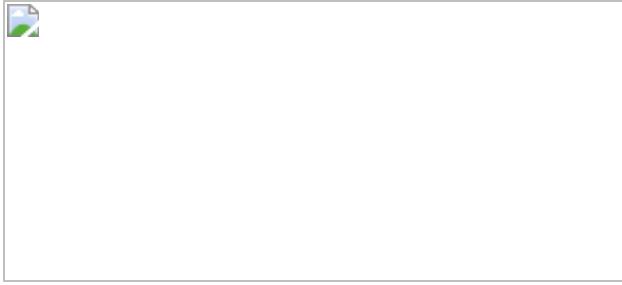


Fig. 2: 95% confidence level exclusion regions.



Data availability

All data used to produce the results of this work, along with the data points shown in the main figures of the paper are stored either on CERN lxplus server or on CERN’s GitLab. They are available upon request to the corresponding author without specific conditions. [Source data](#) are provided with this paper.

Code availability

All code used to produce the results of this work, including code to perform statistical analysis and produce the figures, is stored on CERN’s GitLab server and is available upon request to the corresponding author without specific conditions.

References

1. 1.

Schwinger, J. On gauge invariance and vacuum polarization. *Phys. Rev.* **82**, 664–679 (1951).

2. 2.

Affleck, I. K. & Manton, N. S. Monopole pair production in a magnetic field. *Nucl. Phys. B* **194**, 38–64 (1982).

3. 3.

Dirac, P. A. M. Quantised singularities in the electromagnetic field. *Proc. R. Soc. London A* **133**, 60–72 (1931).

4. 4.

't Hooft, G. Magnetic monopoles in unified gauge theories. *Nucl. Phys. B* **79**, 276–284 (1974).

5. 5.

Polyakov, A. M. Particle spectrum in quantum field theory. *JETP Lett.* **20**, 194–195 (1974).

6. 6.

Wen, X.-G. & Witten, E. Electric and magnetic charges in superstring models. *Nucl. Phys. B* **261**, 651–677 (1985).

7. 7.

Mavromatos, N. E. & Mitsou, V. A. Magnetic monopoles revisited: models and searches at colliders and in the cosmos. *Int. J. Mod. Phys. A* **35**, 2030012 (2020).

8. 8.

Ho, D. L.-J. & Rajantie, A. Classical production of 't Hooft–Polyakov monopoles from magnetic fields. *Phys. Rev. D* **101**, 055003 (2020).

9. 9.

Ho, D. L.-J. & Rajantie, A. Instanton solution for Schwinger production of 't Hooft–Polyakov monopoles. *Phys. Rev. D* **103**, 115033 (2021).

10. 10.

Gould, O., Ho, D. L.-J. & Rajantie, A. Towards Schwinger production of magnetic monopoles in heavy-ion collisions. *Phys. Rev. D* **100**,

015041 (2019).

11. 11.

Huang, X.-G. Electromagnetic fields and anomalous transports in heavy-ion collisions—a pedagogical review. *Rep. Prog. Phys.* **79**, 076302 (2016).

12. 12.

MoEDAL Collaboration. Magnetic monopole search with the full MoEDAL trapping detector in 13 TeV pp collisions interpreted in photon-fusion and Drell–Yan production. *Phys. Rev. Lett.* **123**, 021802 (2019).

13. 13.

Guth, A. H. Inflationary universe: a possible solution to the horizon and flatness problems. *Phys. Rev. D* **23**, 347–356 (1981).

14. 14.

Witten, E. Baryons in the $1/N$ expansion. *Nucl. Phys. B* **160**, 57–115 (1979).

15. 15.

Drukier, A. K. & Nussinov, S. Monopole pair creation in energetic collisions: is it possible? *Phys. Rev. Lett.* **49**, 102–105 (1982).

16. 16.

Blagojević, M. & Senjanović, P. The quantum field theory of electric and magnetic charge. *Phys. Rep.* **157**, 233–346 (1988).

17. 17.

Cho, Y. & Maison, D. Monopole configuration in Weinberg–Salam model. *Phys. Lett. B* **391**, 360–365 (1997).

18. 18.

Kimm, K., Yoon, J. H. & Cho, Y. M. Finite energy electroweak dyon. *Eur. Phys. J. C* **75**, 67 (2015).

19. 19.

Ellis, J., Mavromatos, N. E. & You, T. The price of an electroweak monopole. *Phys. Lett. B* **756**, 29–35 (2016).

20. 20.

Mavromatos, N. E. & Sarkar, S. Magnetic monopoles from global monopoles in the presence of a Kalb–Ramond field. *Phys. Rev. D* **95**, 104025 (2017).

21. 21.

Arunasalam, S. & Kobakhidze, A. Electroweak monopoles and the electroweak phase transition. *Eur. Phys. J. C* **77**, 444 (2017).

22. 22.

Mavromatos, N. E. & Sarkar, S. Regularized Kalb–Ramond magnetic monopole with finite energy. *Phys. Rev. D* **97**, 125010 (2018).

23. 23.

Hung, P. Q. Topologically stable, finite-energy electroweak-scale monopoles. *Nucl. Phys. B* **962**, 115278 (2021).

24. 24.

Sauter, F. Über das Verhalten eines Elektrons im homogenen elektrischen Feld nach der relativistischen Theorie Diracs. *Z. Phys.* **69**, 742–764 (1931).

25. 25.

Heisenberg, W. & Euler, H. Consequences of Dirac's theory of positrons. *Z. Phys.* **98**, 714–732 (1936).

26. 26.

Kaspi, V. M. & Beloborodov, A. M. Magnetars. *Ann. Rev. Astron. Astrophys.* **55**, 261–301 (2017).

27. 27.

Gould, O., Ho, D. L.-J. & Rajantie, A. Schwinger pair production of magnetic monopoles: momentum distribution for heavy-ion collisions. *Phys. Rev. D* **104**, 015033 (2021).

28. 28.

MoEDAL Collaboration. First search for dyons with the full MoEDAL trapping detector in 13 TeV pp collisions. *Phys. Rev. Lett.* **126**, 071801 (2021).

29. 29.

Milton, K. A. Theoretical and experimental status of magnetic monopoles. *Rep. Prog. Phys.* **69**, 1637–1711 (2006).

30. 30.

The MoEDAL Collaboration. The physics programme of the MoEDAL experiment at the LHC. *Int. J. Mod. Phys. A* **29**, 1430050 (2014).

31. 31.

Gamberg, L., Kalbfleisch, G. R. & Milton, K. A. Direct and indirect searches for low-mass magnetic monopoles. *Found. Phys.* **30**, 543–565 (2000).

32. 32.

Agostinelli, S. et al. Geant4—a simulation toolkit. *Nucl. Instrum. Meth. A* **506**, 250–303 (2003).

33. 33.

The MoEDAL Collaboration. Search for magnetic monopoles with the MoEDAL prototype trapping detector in 8 TeV proton–proton collisions at the LHC. *J. High Energy Phys.* **2016**, 67 (2016).

34. 34.

He, Y. D. Search for a Dirac magnetic monopole in high energy nucleus–nucleus collisions. *Phys. Rev. Lett.* **79**, 3134–3137 (1997).

35. 35.

Gould, O. & Rajantie, A. Magnetic monopole mass bounds from heavy-ion collisions and neutron stars. *Phys. Rev. Lett.* **119**, 241601 (2017).

36. 36.

ATLAS Collaboration. Search for magnetic monopoles in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS detector. *Phys. Rev. Lett.* **109**, 261803 (2012).

37. 37.

ATLAS Collaboration. Search for magnetic monopoles and stable particles with high electric charges in 8 TeV pp collisions with the ATLAS detector. *Phys. Rev. D* **93**, 052009 (2016).

38. 38.

ATLAS Collaboration. Search for magnetic monopoles and stable high-electric-charge objects in 13 TeV proton–proton collisions with the ATLAS Detector. *Phys. Rev. Lett.* **124**, 031802 (2020).

39. 39.

Kobayashi, T. Monopole–antimonopole pair production in primordial magnetic fields. *Phys. Rev. D* **104**, 043501 (2021).

40. 40.

Clemencic, M. et al. The LHCb simulation application, Gauss: design, evolution and experience. *J. Phys. Conf. Ser.* **331**, 032023 (2011).

41. 41.

King, M. Simulation of the MoEDAL experiment. *Nucl. Part. Phys. Proc.* **273–275**, 2560–2562 (2016).

42. 42.

Kharzeev, D. E., McLerran, L. D. & Warringa, H. J. The effects of topological charge change in heavy ion collisions: “event by event P and CP violation”. *Nucl. Phys. A* **803**, 227–253 (2008).

43. 43.

Gursoy, U., Kharzeev, D. & Rajagopal, K. Magnetohydrodynamics, charged currents and directed flow in heavy ion collisions. *Phys. Rev. C* **89**, 054905 (2014).

44. 44.

ALICE Collaboration. Centrality determination of Pb–Pb collisions at $\sqrt{\{s\}_{NN}} = 2.76$ TeV with ALICE. *Phys. Rev. C* **88**, 044909 (2013).

45. 45.

ALICE Collaboration. Centrality dependence of particle production in p –Pb collisions at $\sqrt{\{s\}_{NN}} = 5.02$ TeV. *Phys. Rev. C* **91**, 064905 (2015).

46. 46.

Deng, W.-T. & Huang, X.-G. Event-by-event generation of electromagnetic fields in heavy-ion collisions. *Phys. Rev. C* **85**, 044907 (2012).

47. 47.

Baltz, A. J. The physics of ultraperipheral collisions at the LHC. *Phys. Rep.* **458**, 1–171 (2008).

48. 48.

Kruglov, S. I. Pair production and vacuum polarization of vector particles with electric dipole moments and anomalous magnetic moments. *Eur. Phys. J. C* **22**, 89–98 (2001).

49. 49.

Gould, O. & Rajantie, A. Thermal Schwinger pair production at arbitrary coupling. *Phys. Rev. D* **96**, 076002 (2017).

50. 50.

Wolschin, G. Aspects of relativistic heavy-ion collisions. *Universe* **6**, 61 (2020).

51. 51.

Tuchin, K. Time and space dependence of the electromagnetic field in relativistic heavy-ion collisions. *Phys. Rev. C* **88**, 024911 (2013).

52. 52.

Inghirami, G. et al. Magnetic fields in heavy ion collisions: flow and charge transport. *Eur. Phys. J. C* **80**, 293 (2020).

53. 53.

Cecchini, S., Patrizii, L., Sahnoun, Z., Sirri, G. & Togo, V. Energy losses of magnetic monopoles in aluminum, iron and copper. Preprint

at <https://arxiv.org/abs/1606.01220> (2016).

54. 54.

Alvarez, L. W. et al. A magnetic monopole detector utilizing superconducting elements. *Rev. Sci. Instrum.* **42**, 326–330 (1971).

55. 55.

De Roeck, A. et al. Development of a magnetometer-based search strategy for stopped monopoles at the large hadron collider. *Eur. Phys. J. C* **72**, 2212 (2012).

56. 56.

Malkus, W. V. R. The interaction of the Dirac magnetic monopole with matter. *Phys. Rev.* **83**, 899–905 (1951).

57. 57.

Bracci, L. & Fiorentini, G. Binding of magnetic monopoles and atomic nuclei. *Phys Lett. B* **124**, 493–496 (1983).

58. 58.

Bracci, L. & Fiorentini, G. Interactions of magnetic monopoles with nuclei and atoms: formation of bound states and phenomenological consequences. *Nucl. Phys. B* **232**, 236–262 (1984).

59. 59.

Bracci, L. & Fiorentini, G. On the capture of protons by magnetic monopoles. *Nucl. Phys. B* **249**, 519–532 (1985).

60. 60.

Olaussen, K. & Sollie, R. Form factor effects on nucleus–magnetic monopole binding. *Nucl. Phys. B* **255**, 465–479 (1985).

61. 61.

Olaussen, K., Olsen, H. A., Osland, P. & Øverbø, I. Proton capture by magnetic monopoles. *Phys. Rev. Lett.* **52**, 325–328 (1984).

62. 62.

Goebel, C. Binding of monopole to nuclei. In *Monopole '83* (ed. Stone, J. L.) 333–337 (Plenum, 1984).

63. 63.

Ruijgrok, Th. W., Tjon, J. A. & Wu, T. T. Monopole chemistry. *Phys. Lett. B* **129**, 209–212 (1983).

64. 64.

Ruijgrok, T. Binding of matter to a magnetic monopole. *Acta Phys. Pol. B* **15**, 305–314 (1983).

65. 65.

Lipkin, H. J. Effects of magnetic monopoles on nuclear wave functions and possible catalysis of nuclear beta decay and spontaneous fission. *Phys. Lett. B* **133**, 347–350 (1983).

66. 66.

Lipkin, H. J. Monoponucleosis — the wonderful things that monopoles can do to nuclei if they are there. In *Monopole '83* (ed. Stone, J. L.) 347–358 (Plenum, 1984).

Acknowledgements

We thank CERN for the LHC's successful Run-2 operation, as well as the support staff from our institutions without whom MoEDAL could not be operated. We acknowledge the invaluable assistance of particular members of the LHCb Collaboration: G. Wilkinson, R. Lindner, E. Thomas and G.

Corti. Computing support was provided by the GridPP Collaboration, in particular by the Queen Mary University of London and Liverpool grid sites. This work was supported by grant PP00P2 150583 of the Swiss NSF; by the UK Science and Technology Facilities Council via the grants ST/L000326/1, ST/L00044X/1, ST/N00101X/1, ST/P000258/1, ST/P000762/1, ST/T000732/1, ST/T000759/1 and ST/T000791/1; by the Generalitat Valenciana via a special grant for MoEDAL and via the projects PROMETEO-II/2017/033 and PROMETEO/2019/087; by MCIU/AEI/FEDER, UE via the grants FPA2016-77177-C2-1-P, FPA2017-85985-P, FPA2017-84543-P and PGC2018-094856-B-I00; by the Physics Department of King’s College London; by NSERC via a project grant; by the V-P Research of the University of Alberta (UofA); by the Provost of the UofA; by UEFISCDI (Romania); by the INFN (Italy); by the Estonian Research Council via a Mobilitas Pluss grant MOBTT5; by the Research Funds of the University of Helsinki; and by the NSF grant 2011214 to the University of Alabama MoEDAL group. A.R. was also supported by Institute for Particle Physics Phenomenology Associateship.

Author information

Author notes

1. Deceased: P. Mermod

Affiliations

1. Theoretical Particle Physics & Cosmology Group, Physics Department, King’s College London, London, UK
B. Acharya, J. Alexandre, J. R. Ellis, M. Fairbairn, N. E. Mavromatos, M. Sakellariadou & S. Sarkar
2. IEAP, Czech Technical University in Prague, Prague, Czech Republic
P. Benes, B. Bergmann, P. Burian, J. Janecek, S. Pospisil & M. Suk
3. INFN, Section of Bologna, Bologna, Italy

S. Bertolucci, G. Levi, A. Maulik, N. Mauri, L. Patrizii, M. Pozzato, Z. Sahnoun, G. Sirri & V. Togo

4. School of Physics and Astronomy, Queen Mary University of London, London, UK

A. Bevan, J. Hays & L. Millward

5. Institute of Space Science, Măgurele, Romania

H. Branzas, D. Felea, G. E. Păvălaş, L. A. Popa & V. Popa

6. Experimental Physics Department, CERN, Geneva, Switzerland

M. Campbell, A. De Roeck, M. El Sawy & D. H. Lacarrère

7. Center for Quantum Spacetime, Sogang University, Seoul, Korea

Y. M. Cho

8. Physics Department, University of Alberta, Edmonton, Alberta, Canada

M. de Montigny, A. Maulik, P.-P. Ouimet, J. L. Pinfold, A. Shaa, R. Soluk, M. Staelens & J. A. Tuszyński

9. Theoretical Physics Department, CERN, Geneva, Switzerland

J. R. Ellis

10. Department of Physics, Concordia University, Montreal, Quebec, Canada

M. Frank

11. University of Nottingham, Nottingham, UK

O. Gould

12. Helsinki Institute of Physics, University of Helsinki, Helsinki, Finland

O. Gould & M. Kalliokoski

13. Department of Earth Sciences, Swiss Federal Institute of Technology,
Zurich, Switzerland

A. M. Hirt

14. Department of Physics, Imperial College London, London, UK

D. L.-J. Ho & A. Rajantie

15. Department of Physics, University of Virginia, Charlottesville, VA,
USA

P. Q. Hung

16. Département de Physique Nucléaire et Corpusculaire, Université de
Genève, Geneva, Switzerland

A. Korzenev, A. Lioni & P. Mermod

17. Département de Physique, Université de Montréal, Montreal, Quebec,
Canada

C. Leroy

18. Department of Physics and Astronomy, University of Bologna,
Bologna, Italy

G. Levi, A. Margiotta & M. Spurio

19. IFIC, Universitat de València, CSIC, Valencia, Spain

V. A. Mitsou, J. Papavassiliou, R. Ruiz de Austri, A. Santra, V.
Vento & O. Vives

20. Department of Physics and Astronomy, University of Alabama,
Tuscaloosa, AL, USA
I. Ostrovskiy & A. Upreti
21. Institute for Research in Schools, Canterbury, UK
B. Parker
22. Department of Physics, University of British Columbia, Vancouver,
British Columbia, Canada
G. Semenoff
23. Department of Physics and Astronomy, Tufts University, Medford,
MA, USA
K. Sliwa
24. CNAF, INFN, Bologna, Italy
M. Tenti

Contributions

The Monopole and Exotics Detector at the LHC was constructed and is maintained by the MoEDAL collaboration. A large number of authors contributed to the data processing, detector calibration and Monte Carlo simulations used in this work. The MoEDAL collaboration acknowledges the substantial contributions to this manuscript from A.U. and I.O. (simulation, statistical analysis, result plots, paper writing); O.G., D.L.-J.H. and A.R. (theoretical calculations, paper writing); and N.E.M. and J.P. (paper writing). The manuscript was reviewed and edited by the collaboration and all authors approved the final version of the manuscript.

Corresponding author

Correspondence to [I. Ostrovskiy](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Muneto Nitta, Steve Ahlen 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 Mean expected rate of Swinger MMs \(\$R_{\text{exp}}\$ \)](#)

The mean expected rate of MMs with $1g_D$ (left) and $2g_D$ (right) magnetic charge in the MMT as a function of the MM mass in the FPA model. The black line corresponds to the default geometry. The grey region corresponds to the systematic error, which is dominated by the material budget. The 95% confidence level mass exclusion region is shown in blue.

[Source data](#)

[Extended Data Fig. 2 Transverse momentum distribution of Swinger MMs.](#)

The transverse momentum distribution for Swinger MMs derived from the FPA, as a function of MM mass (M) plotted versus MM β .

Source data

Extended Data Table 1 Expected rate of MM trapping in the MoEDAL MMTs for the $1g_D$ FPA model, where ϵ is MMT trapping efficiency and R_{exp} is the mean expected rate of trapped MMs

Extended Data Table 2 Expected rate of MM trapping in the MoEDAL MMTs for the $2g_D$ FPA model

Extended Data Table 3 Expected rate of MM trapping in the MoEDAL MMTs for the $3g_D$ FPA model

Extended Data Table 4 Expected rate of MM trapping in the MoEDAL MMTs for the $4g_D$ FPA model

Extended Data Table 5 Expected rate of MM trapping in the MoEDAL MMTs for the $5g_D$ FPA model

Source data

Source Data Fig. 2

Source Data Extended Data Fig. 1

Source Data Extended Data Fig. 2

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Acharya, B., Alexandre, J., Benes, P. *et al.* Search for magnetic monopoles produced via the Schwinger mechanism. *Nature* **602**, 63–67 (2022).
<https://doi.org/10.1038/s41586-021-04298-1>

- Received: 18 June 2021
- Accepted: 01 December 2021
- Published: 02 February 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04298-1>

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

[Searches for magnetic monopoles](#)

- Iulia Georgescu

Research Highlight 02 Feb 2022

[Search for single magnetic charges in the largest of fields](#)

- Sonia Kabana

News & Views 02 Feb 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04298-1>

- Article
- [Published: 02 February 2022](#)

Domain-wall dynamics in Bose–Einstein condensates with synthetic gauge fields

- [Kai-Xuan Yao](#) ORCID: [orcid.org/0000-0002-4071-7060^{1,2,3}](https://orcid.org/0000-0002-4071-7060),
- [Zhendong Zhang](#) ORCID: [orcid.org/0000-0001-7221-1729^{1,2,3}](https://orcid.org/0000-0001-7221-1729) &
- [Cheng Chin](#) ORCID: [orcid.org/0000-0003-0278-4630^{1,2,3}](https://orcid.org/0000-0003-0278-4630)

[Nature](#) volume **602**, pages 68–72 (2022)

- 670 Accesses
- 15 Altmetric
- [Metrics details](#)

Subjects

- [Bose–Einstein condensates](#)
- [Quantum simulation](#)

Abstract

Interactions in many-body physical systems, from condensed matter to high-energy physics, lead to the emergence of exotic particles. Examples are mesons in quantum chromodynamics and composite fermions in fractional quantum Hall systems, which arise from the dynamical coupling between matter and gauge fields^{1,2}. The challenge of understanding the

complexity of matter–gauge interaction can be aided by quantum simulations, for which ultracold atoms offer a versatile platform via the creation of artificial gauge fields. An important step towards simulating the physics of exotic emergent particles is the synthesis of artificial gauge fields whose state depends dynamically on the presence of matter. Here we demonstrate deterministic formation of domain walls in a stable Bose–Einstein condensate with a gauge field that is determined by the atomic density. The density-dependent gauge field is created by simultaneous modulations of an optical lattice potential and interatomic interactions, and results in domains of atoms condensed into two different momenta. Modelling the domain walls as elementary excitations, we find that the domain walls respond to synthetic electric field with a charge-to-mass ratio larger than and opposite to that of the bare atoms. Our work offers promising prospects to simulate the dynamics and interactions of previously undescribed excitations in quantum systems with dynamical gauge fields.

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: Illustration of a Bose–Einstein condensate with a density-dependent gauge field.

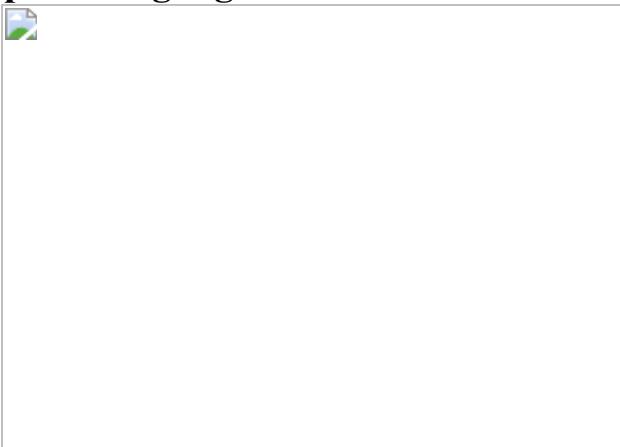


Fig. 2: Creation of static and density-dependent gauge fields.

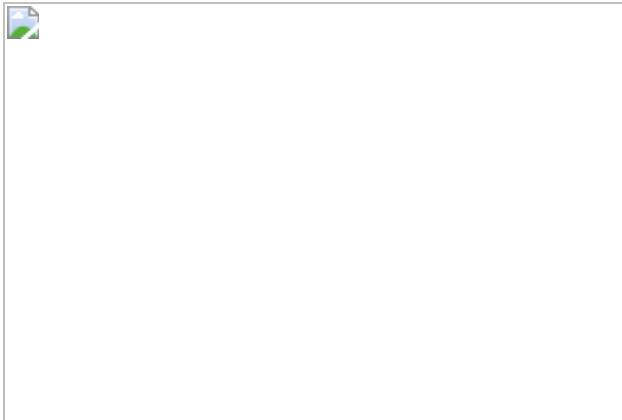


Fig. 3: Domains and domain walls in the presence of a density-dependent gauge field.

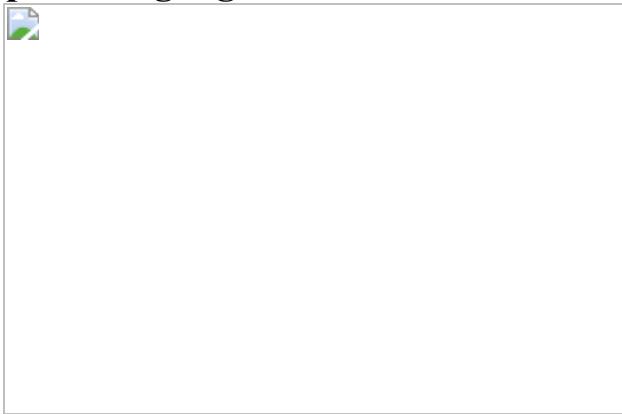


Fig. 4: Dynamics of the domain wall in response to a synthetic electric field $\{\boldsymbol{E}\}$.



Data availability

The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request. [Source data](#) are provided with this paper.

Code availability

The codes for the analysis of data shown within this paper are available from the corresponding author upon reasonable request.

References

1. 1.
Griffiths, D. *Introduction to Elementary Particles* (Wiley, 2008).
2. 2.
Stormer, H. L., Tsui, D. C. & Gossard, A. C. The fractional quantum Hall effect. *Rev. Mod. Phys.* **71**, S298–S305 (1999).
3. 3.
Kogut, J. B. An introduction to lattice gauge theory and spin systems. *Rev. Mod. Phys.* **51**, 659–713 (1979).
4. 4.
Wilson, K. G. Confinement of quarks. *Phys. Rev. D* **10**, 2445–2459 (1974).
5. 5.
Alford, M. G., Schmitt, A., Rajagopal, K. & Schäfer, T. Color superconductivity in dense quark matter. *Rev. Mod. Phys.* **80**, 1455–1515 (2008).
6. 6.

Troyer, M. & Wiese, U.-J. Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations. *Phys. Rev. Lett.* **94**, 170201 (2005).

7. 7.

Goldman, N., Juzeliūnas, G., Öhberg, P. & Spielman, I. B. Light-induced gauge fields for ultracold atoms. *Rep. Prog. Phys.* **77**, 126401 (2014).

8. 8.

Zohar, E., Cirac, J. I. & Reznik, B. Quantum simulations of lattice gauge theories using ultracold atoms in optical lattices. *Rep. Prog. Phys.* **79**, 014401 (2015).

9. 9.

Cooper, N. R., Dalibard, J. & Spielman, I. B. Topological bands for ultracold atoms. *Rev. Mod. Phys.* **91**, 015005 (2019).

10. 10.

Lin, Y., Compton, R., Jiménez-García, K., Porto, J. V. & Spielman, I. B. Synthetic magnetic fields for ultracold neutral atoms. *Nature* **462**, 628–632 (2009).

11. 11.

Jotzu, G. et al. Experimental realization of the topological Haldane model with ultracold fermions. *Nature* **515**, 237–240 (2014).

12. 12.

Aidelsburger, M. et al. Realization of the Hofstadter Hamiltonian with ultracold atoms in optical lattices. *Phys. Rev. Lett.* **111**, 185301 (2013).

13. 13.

Miyake, H., Siviloglou, G. A., Kennedy, C. J., Burton, W. C. & Ketterle, W. Realizing the Harper Hamiltonian with laser-assisted tunneling in optical lattices. *Phys. Rev. Lett.* **111**, 185302 (2013).

14. 14.

Baskaran, G. & Anderson, P. W. Gauge theory of high-temperature superconductors and strongly correlated Fermi systems. *Phys. Rev. B* **37**, 580–583 (1988).

15. 15.

Cheng, T.-P. & Li, L.-F. *Gauge Theory of Elementary Particle Physics* (Oxford Univ. Press, 1994).

16. 16.

Levin, M. & Wen, X.-G. Colloquium: photons and electrons as emergent phenomena. *Rev. Mod. Phys.* **77**, 871–879 (2005).

17. 17.

Wiese, U.-J. Ultracold quantum gases and lattice systems: quantum simulation of lattice gauge theories. *Ann. Phys.* **525**, 777–796 (2013).

18. 18.

Savary, L. & Balents, L. Quantum spin liquids: a review. *Rep. Prog. Phys.* **80**, 016502 (2016).

19. 19.

Clark, L. W. et al. Observation of density-dependent gauge fields in a Bose-Einstein condensate based on micromotion control in a shaken two-dimensional lattice. *Phys. Rev. Lett.* **121**, 030402 (2018).

20. 20.

Görg, F. et al. Realization of density-dependent Peierls phases to engineer quantized gauge fields coupled to ultracold matter. *Nat. Phys.* **15**, 1161–1167 (2019).

21. 21.

Lienhard, V. et al. Realization of a density-dependent Peierls phase in a synthetic, spin-orbit coupled Rydberg system. *Phys. Rev. X* **10**, 021031 (2020).

22. 22.

Edmonds, M. J., Valiente, M., Juzeliūnas, G., Santos, L. & Öhberg, P. Simulating an interacting gauge theory with ultracold Bose gases. *Phys. Rev. Lett.* **110**, 085301 (2013).

23. 23.

Schweizer, C. et al. Floquet approach to $\langle\{\{\mathbb{Z}\}\}_2\rangle$ lattice gauge theories with ultracold atoms in optical lattices. *Nat. Phys.* **15**, 1168–1173 (2019).

24. 24.

Yang, B. et al. Observation of gauge invariance in a 71-site Bose–Hubbard quantum simulator. *Nature* **587**, 392–396 (2020).

25. 25.

Mil, A. et al. A scalable realization of local U(1) gauge invariance in cold atomic mixtures. *Science* **367**, 1128–1130 (2020).

26. 26.

Liu, T., Clark, L. W. & Chin, C. Exotic domain walls in Bose-Einstein condensates with double-well dispersion. *Phys. Rev. A* **94**, 063646 (2016).

27. 27.

Eckardt, A. Colloquium: atomic quantum gases in periodically driven optical lattices. *Rev. Mod. Phys.* **89**, 011004 (2017).

28. 28.

Struck, J. et al. Tunable gauge potential for neutral and spinless particles in driven optical lattices. *Phys. Rev. Lett.* **108**, 225304 (2012).

29. 29.

Shvarchuck, I. et al. Bose-Einstein condensation into nonequilibrium states studied by condensate focusing. *Phys. Rev. Lett.* **89**, 270404 (2002).

30. 30.

Clark, L. W., Feng, L. & Chin, C. Universal space-time scaling symmetry in the dynamics of bosons across a quantum phase transition. *Science* **354**, 606–610 (2016).

31. 31.

Parker, C. V., Ha, L.-C. & Chin, C. Direct observation of effective ferromagnetic domains of cold atoms in a shaken optical lattice. *Nat. Phys.* **9**, 769–774 (2013).

32. 32.

Mermin, N. D. The topological theory of defects in ordered media. *Rev. Mod. Phys.* **51**, 591–648 (1979).

33. 33.

Gani, V. A., Kudryavtsev, A. E. & Lizunova, M. A. Kink interactions in the (1+1)-dimensional φ^6 model. *Phys. Rev. D* **89**, 125009 (2014).

34. 34.

Vilenkin, A. Cosmic strings and domain walls. *Phys. Rep.* **121**, 263–315 (1985).

35. 35.

Barbiero, L. et al. Coupling ultracold matter to dynamical gauge fields in optical lattices: from flux attachment to $\langle\{\{\mathbb{Z}\}\}_2\rangle$ lattice gauge theories. *Sci. Adv.* **5**, eaav7444 (2019).

Acknowledgements

We thank E. Mueller for helpful discussions and K. Patel for carefully reading the manuscript. This work is supported by the National Science Foundation (NSF) grant no. PHY-2103542, NSF QLCI-HQAN no. 2016136, and the Army Research Office STIR grant W911NF2110108. Z.Z. is supported by the Grainger Graduate Fellowship.

Author information

Affiliations

1. James Franck Institute, University of Chicago, Chicago, IL, USA

Kai-Xuan Yao, Zhendong Zhang & Cheng Chin

2. Enrico Fermi Institute, University of Chicago, Chicago, IL, USA

Kai-Xuan Yao, Zhendong Zhang & Cheng Chin

3. Department of Physics, University of Chicago, Chicago, IL, USA

Kai-Xuan Yao, Zhendong Zhang & Cheng Chin

Contributions

K.-X.Y. designed and performed the experiments and analysed the data. All authors contributed to discussions on the experiment and preparation of the

manuscript. C.C. supervised the project.

Corresponding author

Correspondence to [Cheng Chin](#).

Ethics declarations

Competing interests

The authors declare no competing financial interests.

Peer review information

Nature thanks Pietro Massignan and the other, anonymous, reviewer 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 Estimation of the zero-crossing position.](#)

The population imbalance between the $\pm k^*$ states in Fig. 2h is fitted to extract the zero-crossing position.

[Extended Data Fig. 2 Extraction of \$\epsilon_{\text{exp}}\$ from magnetization M.](#)

Experiment data in Fig. 3e are fitted to extract the value of ϵ_{exp} .

Supplementary information

[Peer Review File](#)

Source data

[Source Data Fig. 2](#)

[Source Data Fig. 3](#)

[Source Data Fig. 4](#)

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Yao, KX., Zhang, Z. & Chin, C. Domain-wall dynamics in Bose–Einstein condensates with synthetic gauge fields. *Nature* **602**, 68–72 (2022).
<https://doi.org/10.1038/s41586-021-04250-3>

- Received: 11 June 2021
- Accepted: 16 November 2021
- Published: 02 February 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04250-3>

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/s41586-021-04250-3>

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

- Article
- [Published: 02 February 2022](#)

Polarized phonons carry angular momentum in ultrafast demagnetization

- [S. R. Tauchert^{1,2}](#),
- [M. Volkov](#) [ORCID: orcid.org/0000-0003-1594-8542^{1,2}](#),
- [D. Ehberger²](#),
- [D. Kazenwadel](#) [ORCID: orcid.org/0000-0003-0225-5317¹](#),
- [M. Evers¹](#),
- [H. Lange](#) [ORCID: orcid.org/0000-0002-0051-2087¹](#),
- [A. Donges¹](#),
- [A. Book](#) [ORCID: orcid.org/0000-0003-0127-6052³](#),
- [W. Kreuzpaintner^{3,4,5}](#),
- [U. Nowak](#) [ORCID: orcid.org/0000-0003-2925-6774¹](#) &
- [P. Baum](#) [ORCID: orcid.org/0000-0002-1521-8729^{1,2}](#)

[Nature](#) volume **602**, pages 73–77 (2022)

- 1449 Accesses
- 65 Altmetric
- [Metrics details](#)

Subjects

- [Magnetic properties and materials](#)
- [Ultrafast photonics](#)

Abstract

Magnetic phenomena are ubiquitous in nature and indispensable for modern science and technology, but it is notoriously difficult to change the magnetic order of a material in a rapid way. However, if a thin nickel film is subjected to ultrashort laser pulses, it loses its magnetic order almost completely within femtosecond timescales¹. This phenomenon is widespread^{2,3,4,5,6,7} and offers opportunities for rapid information processing^{8,9,10,11} or ultrafast spintronics at frequencies approaching those of light^{8,9,12}. Consequently, the physics of ultrafast demagnetization is central to modern materials research^{1,2,3,4,5,6,7,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28}, but a crucial question has remained elusive: if a material loses its magnetization within mere femtoseconds, where is the missing angular momentum in such a short time? Here we use ultrafast electron diffraction to reveal in nickel an almost instantaneous, long-lasting, non-equilibrium population of anisotropic high-frequency phonons that appear within 150–750 fs. The anisotropy plane is perpendicular to the direction of the initial magnetization and the atomic oscillation amplitude is 2 pm. We explain these observations by means of circularly polarized phonons that quickly absorb the angular momentum of the spin system before macroscopic sample rotation. The time that is needed for demagnetization is related to the time it takes to accelerate the atoms. These results provide an atomistic picture of the Einstein–de Haas effect and signify the general importance of polarized phonons for non-equilibrium dynamics and phase transitions.

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: Pump–probe electron diffraction measurements of atomic motions during ultrafast demagnetization.

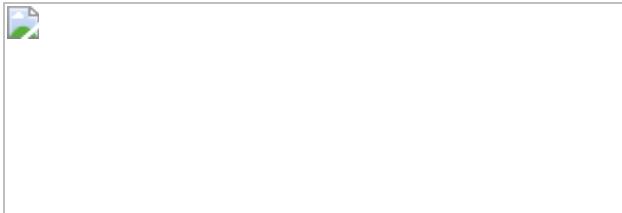


Fig. 2: Time-resolved diffraction results.

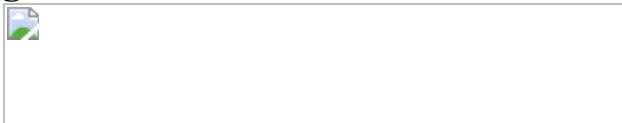


Fig. 3: Molecular dynamics simulations.



Fig. 4: Sequence of events.

Data availability

The data supporting the findings of this study are available from the corresponding author upon request.

References

1. 1.

Beaurepaire, E., Merle, J. C., Daunois, A. & Bigot, J. Y. Ultrafast spin dynamics in ferromagnetic nickel. *Phys. Rev. Lett.* **76**, 4250–4253 (1996).

2. 2.

Koopmans, B. et al. Explaining the paradoxical diversity of ultrafast laser-induced demagnetization. *Nat. Mater.* **9**, 259–265 (2010).

3. 3.

Wietstruk, M. et al. Hot-electron-driven enhancement of spin-lattice coupling in Gd and Tb 4f ferromagnets observed by femtosecond x-ray magnetic circular dichroism. *Phys. Rev. Lett.* **106**, 127401 (2011).

4. 4.

Graves, C. E. et al. Nanoscale spin reversal by non-local angular momentum transfer following ultrafast laser excitation in ferrimagnetic GdFeCo. *Nat. Mater.* **12**, 293–298 (2013).

5. 5.

von Korff Schmising, C. et al. Imaging ultrafast demagnetization dynamics after a spatially localized optical excitation. *Phys. Rev. Lett.* **112**, 217203 (2014).

6. 6.

Frietsch, B. et al. Disparate ultrafast dynamics of itinerant and localized magnetic moments in gadolinium metal. *Nat. Commun.* **6**, 8262 (2015).

7. 7.

Frietsch, B. et al. The role of ultrafast magnon generation in the magnetization dynamics of rare-earth metals. *Sci. Adv.* **6**, eabb1601 (2020).

8. 8.

Stanciu, C. D. et al. All-optical magnetic recording with circularly polarized light. *Phys. Rev. Lett.* **99**, 047601 (2007).

9. 9.

Radu, I. et al. Transient ferromagnetic-like state mediating ultrafast reversal of antiferromagnetically coupled spins. *Nature* **472**, 205–208 (2011).

10. 10.

Ostler, T. A. et al. Ultrafast heating as a sufficient stimulus for magnetization reversal in a ferrimagnet. *Nat. Commun.* **3**, 666 (2012).

11. 11.

Wienholdt, S., Hinzke, D., Carva, K., Oppeneer, P. M. & Nowak, U. Orbital-resolved spin model for thermal magnetization switching in rare-earth-based ferrimagnets. *Phys. Rev. B* **88**, 020406(R) (2013).

12. 12.

Siegrist, F. et al. Light-wave dynamic control of magnetism. *Nature* **571**, 240–244 (2019).

13. 13.

Malinowski, G. et al. Control of speed and efficiency of ultrafast demagnetization by direct transfer of spin angular momentum. *Nat. Phys.* **4**, 855–858 (2008).

14. 14.

Battiato, M., Carva, K. & Oppeneer, P. M. Superdiffusive spin transport as a mechanism of ultrafast demagnetization. *Phys. Rev. Lett.* **105**, 027203 (2010).

15. 15.

Melnikov, A. et al. Ultrafast transport of laser-excited spin-polarized carriers Au/Fe/MgO(001). *Phys. Rev. Lett.* **107**, 076601 (2011).

16. 16.

Rudolf, D. et al. Ultrafast magnetization enhancement in metallic multilayers driven by superdiffusive spin current. *Nat. Commun.* **3**, 1037 (2012).

17. 17.

Eschenlohr, A. et al. Ultrafast spin transport as key to femtosecond demagnetization. *Nat. Mater.* **12**, 332–336 (2013).

18. 18.

Koopmans, B., Ruigrok, J. J. M., Dalla Longa, F. & De Jonge, W. J. M. Unifying ultrafast magnetization dynamics. *Phys. Rev. Lett.* **95**, 267207 (2005).

19. 19.

Carva, K., Battiato, M. & Oppeneer, P. M. Ab initio investigation of the Elliott-Yafet electron-phonon mechanism in laser-induced ultrafast demagnetization. *Phys. Rev. Lett.* **107**, 207201 (2011).

20. 20.

La-O-Vorakiat, C. et al. Ultrafast demagnetization measurements using extreme ultraviolet light: Comparison of electronic and magnetic contributions. *Phys. Rev.* **2**, 011005 (2012).

21. 21.

Hinzke, D. et al. Multiscale modeling of ultrafast element-specific magnetization dynamics of ferromagnetic alloys. *Phys. Rev. B* **92**, 054412 (2015).

22. 22.

Dornes, C. et al. The ultrafast Einstein-de Haas effect. *Nature* **565**, 209–212 (2019).

23. 23.

Roth, T. et al. Temperature dependence of laser-induced demagnetization in Ni: a key for identifying the underlying mechanism. *Phys. Rev.* **2**, 021006 (2012).

24. 24.

Schellekens, A. J., Verhoeven, W., Vader, T. N., Koopmans, B. Investigating the contribution of superdiffusive transport to ultrafast demagnetization of ferromagnetic thin films. *Appl. Phys. Lett.* **102**, 252408 (2013).

25. 25.

Stamm, C. et al. Femtosecond modification of electron localization and transfer of angular momentum in nickel. *Nat. Mater.* **6**, 740–743 (2007).

26. 26.

Maldonado, P. et al. Tracking the ultrafast nonequilibrium energy flow between electronic and lattice degrees of freedom in crystalline nickel. *Phys. Rev. B* **101**, 100302 (2020).

27. 27.

Chen, Z. & Wang, L.-W. Role of initial magnetic disorder: a time-dependent ab initio study of ultrafast demagnetization mechanisms. *Sci. Adv.* **5**, eaau8000 (2019).

28. 28.

Kealhofer, C. et al. All-optical control and metrology of electron pulses. *Science* **352**, 429–433 (2016).

29. 29.

Wang, X. et al. Temperature dependence of electron-phonon thermalization and its correlation to ultrafast magnetism. *Phys. Rev. B* **81**, 220301 (2010).

30. 30.

Zhang, L. F. & Niu, Q. Angular momentum of phonons and the Einstein-de Haas effect. *Phys. Rev. Lett.* **112**, 085503 (2014).

31. 31.

Garanin, D. A. & Chudnovsky, E. M. Angular momentum in spin-phonon processes. *Phys. Rev. B* **92**, 024421 (2015).

32. 32.

Zhu, H. et al. Observation of chiral phonons. *Science* **359**, 579–582 (2018).

33. 33.

Birgeneau, R. J., Cordes, J., Dolling, G. & Woods, A. D. B. Normal modes of vibration in nickel. *Phys. Rev. A* **136**, 1359–1365 (1964).

34. 34.

Zahn, D. et al. Lattice dynamics and ultrafast energy flow between electrons, spins, and phonons in a 3D ferromagnet. *Phys. Rev. Res.* **3**, 023032 (2021).

35. 35.

Tengdin, P. et al. Critical behavior within 20 fs drives the out-of-equilibrium laser-induced magnetic phase transition in nickel. *Sci. Adv.* <https://doi.org/10.1126/science.aaw9486> (2018).

36. 36.

Hofherr, M. et al. Induced versus intrinsic magnetic moments in ultrafast magnetization dynamics. *Phys. Rev. B* **98**, 174419 (2018).

37. 37.

Fechner, M. et al. Magnetophononics: ultrafast spin control through the lattice. *Phys. Rev. Mater.* **2**, 064401 (2018).

38. 38.

Disa, A. S. et al. Polarizing an antiferromagnet by optical engineering of the crystal field. *Nat. Phys.* **16**, 937–941 (2020).

39. 39.

Gao, M. N., Zhang, W. & Zhang, L. F. Nondegenerate chiral phonons in graphene/hexagonal boron nitride heterostructure from first-principles calculations. *Nano Lett.* **18**, 4424–4430 (2018).

40. 40.

Grissonnanche, G. et al. Chiral phonons in the pseudogap phase of cuprates. *Nat. Phys.* **16**, 1108–1111 (2020).

41. 41.

Hirashita, N., Kinoshita, M., Aikawa, I. & Ajioka, T. Effects of surface hydrogen on the air oxidation at room temperature of HF treated Si (100) surfaces. *Appl. Phys. Lett.* **56**, 451–453 (1990).

42. 42.

Mazzara, C. et al. Hydrogen-terminated Si(111) and Si(100) by wet chemical treatment: linear and non-linear infrared spectroscopy. *Surf. Sci.* **427–428**, 208–213 (1999).

43. 43.

Ji, J.-Y., Shen, T.-C. Low-temperature silicon epitaxy on hydrogen-terminated Si(001) surfaces. *Phys. Rev. B* **70**, 115309 (2004).

44. 44.

Kreuzpaintner, W., Störmer, M., Lott, D., Solina, D. & Schreyer, A. Epitaxial growth of nickel on Si(100) by dc magnetron sputtering. *J. Appl. Phys.* **104**, 114302 (2008).

45. 45.

Kreuzpaintner, W., Störmer, M., Lott, D., Solina, D. & Schreyer, A. Epitaxial growth of nickel on Si(100) by dc magnetron sputtering. *J. Appl. Phys.* **104**, 114302 (2008).

46. 46.

Schmehl, A. et al. Design and realization of a sputter deposition system for the in situ- and in operando-use in polarized neutron reflectometry experiments. *Nucl. Instrum. Methods Phys. Res. A* **883**, 170–182 (2018).

47. 47.

Jiang, H., Klemmer, T. J., Barnard, J. A. & Payzant, E. A. Epitaxial growth of Cu on Si by magnetron sputtering. *J. Vac. Sci. Technol. A* **16**, 3376–3383 (1998).

48. 48.

Chang, C.-A. Reversed magnetic anisotropy in deformed (100) Cu/Ni/Cu structures. *J. Appl. Phys.* **68**, 4873–4875 (1990).

49. 49.

Chang, C.-A. Reversal in magnetic anisotropy of (100)Cu-Ni superlattices. *J. Magn. Magn. Mater.* **97**, 102–106 (1991).

50. 50.

Ye, J. et al. Design and realization of a sputter deposition system for the in situ and in operando use in polarized neutron reflectometry experiments: novel capabilities. *Nucl. Instrum. Methods Phys. Res. A* **964**, 163710 (2020).

51. 51.

Hull, C. M., Switzer, J. A. Electrodeposited epitaxial cu(100) on si(100) and lift-off of single crystal-like Cu(100) foils. *ACS Appl. Mater. Interfaces* **10**, 38596–38602 (2018).

52. 52.

Warren, B. E. *X-Ray Diffraction* (Dover, 1990).

53. 53.

Björck, M. & Andersson, G. *GenX*: an extensible X-ray reflectivity refinement program utilizing differential evolution. *J. Appl. Crystallogr.* **40**, 1174–1178 (2007).

54. 54.

Cemin, F. et al. Epitaxial growth of Cu(001) thin films onto Si(001) using a single-step HiPIMS process. *Sci. Rep.* **7**, 1655 (2017).

55. 55.

Chen, L., Andrea, L., Timalsina, Y. P., Wang, G.-C. & Lu, T.-M. Engineering epitaxial-nanospiral metal films using dynamic oblique angle deposition. *Cryst. Growth Des.* **13**, 2075–2080 (2013).

56. 56.

Seidel, M. et al. Efficient high-power ultrashort pulse compression in self-defocusing bulk media. *Sci. Rep.* **7**, 1410 (2017).

57. 57.

Srinivasan, R., Lobastov, V. A., Ruan, C.-Y. & Zewail, A. H. Ultrafast electron diffraction (UED). *Helv. Chim. Acta* **86**, 1761–1799 (2003).

58. 58.

Miller, R. J. D. Femtosecond crystallography with ultrabright electrons and x-rays: capturing chemistry in action. *Science* **343**, 1108–1116 (2014).

59. 59.

Kasmi, L., Kreier, D., Bradler, M., Riedle, E. & Baum, P. Femtosecond single-electron pulses generated by two-photon photoemission close to the work function. *New J. Phys.* **17**, 033008 (2015).

60. 60.

Ehberger, D. et al. Terahertz compression of electron pulses at a planar mirror membrane. *Phys. Rev. Appl.* **11**, 024034 (2019).

61. 61.

Simerska, M. The temperature dependence of the characteristic Debye temperature of nickel. *Czech. J. Phys. B* **12**, 858–859 (1962).

62. 62.

Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).

63. 63.

Foiles, S. M., Baskes, M. I. & Daw, M. S. Embedded-atom-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys. *Phys. Rev. B* **33**, 7983–7991 (1986).

64. 64.

Sandia National Laboratories *LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator)*
<https://lammps.sandia.gov/doc/Intro.html> (2019).

65. 65.

Coleman, S. P., Spearot, D. E. & Capolungo, L. Virtual diffraction analysis of Ni [010] symmetric tilt grain boundaries. *Model. Simul. Mater. Sci. Eng.* **21**, 055020 (2013).

66. 66.

Danan, H., Herr, A. & Meyer, A. J. New determinations of the saturation magnetization of nickel and iron. *J. Appl. Phys.* **39**, 669–670 (1968).

67. 67.

Scott, G. G. The gyromagnetic ratios of the ferromagnetic elements. *Phys. Rev.* **87**, 697–699 (1952).

68. 68.

You, W. et al. Revealing the nature of the ultrafast magnetic phase transition in Ni by correlating extreme ultraviolet magneto-optic and photoemission spectroscopies. *Phys. Rev. Lett.* **121**, 077204 (2018).

69. 69.

Volkov, M. et al. Attosecond screening dynamics mediated by electron localization in transition metals. *Nat. Phys.* **15**, 1145–1149 (2019).

70. 70.

Lee, E. W. Magnetostriction and magnetomechanical effects. *Rep. Prog. Phys.* **18**, 184–229 (1955).

71. 71.

Guo, G. Y. Orientation dependence of the magnetoelastic coupling constants in strained FCC Co and Ni: an ab initio study. *J. Magn. Magn. Mater.* **209**, 33–36 (2000).

72. 72.

Grossinger, R., Turtelli, R. S. & Mehmood, N. Materials with high magnetostriction. In *13th International Symposium on Advanced Materials (ISAM 2013)* **60**, 012002 (2014).

73. 73.

Pateras, A. et al. Room temperature giant magnetostriction in single-crystal nickel nanowires. *NPG Asia Mater.* **11**, 59 (2019).

74. 74.

Farle, M., Mirwald-Schulz, B., Anisimov, A. N., Platow, W. & Baberschke, K. Higher-order magnetic anisotropies and the nature of the spin-reorientation transition in face-centered-tetragonal Ni(001)/Cu(001). *Phys. Rev. B* **55**, 3708–3715 (1997).

75. 75.

Kittel, C. On the gyromagnetic ratio and spectroscopic splitting factor of ferromagnetic substances. *Phys. Rev.* **76**, 743–748 (1949).

76. 76.

Van Vleck, J. H. Concerning the theory of ferromagnetic resonance absorption. *Phys. Rev.* **78**, 266–274 (1950).

77. 77.

Scott, G. G. Review of gyromagnetic ratio experiments. *Rev. Mod. Phys.* **34**, 102–109 (1962).

Acknowledgements

We thank I. Wimmer for magnetic hysteresis data, B.-H. Chen for help with the optics, S. Geprägs for access to his X-ray diffractometer and F. Krausz for laboratory infrastructure. This research was supported by the European Union's Horizon 2020 research and innovation program via CoG 647771 and by the German Research Foundation (DFG) via SFB 1432.

Author information

Affiliations

1. Universität Konstanz, Fachbereich Physik, Konstanz, Germany
S. R. Tauchert, M. Volkov, D. Kazenwadel, M. Evers, H. Lange, A. Donges, U. Nowak & P. Baum
2. Ludwig-Maximilians-Universität München, Garching, Germany
S. R. Tauchert, M. Volkov, D. Ehberger & P. Baum
3. Technische Universität München, Physik-Department E21, Garching, Germany
A. Book & W. Kreuzpaintner
4. Institute of High Energy Physics, Chinese Academy of Sciences (CAS), Beijing, China
W. Kreuzpaintner
5. Spallation Neutron Source Science Center, Dongguan, China
W. Kreuzpaintner

Contributions

P.B. and U.N. conceived the experiment. S.T., M.V. and D.E. performed the diffraction experiments and analyzed the data. A.B. and S.T. produced the specimen under supervision of W.K. A.B. and W.K. characterized the epitaxial growth. D.K. performed the ultrafast optical measurements and thermal simulations. U.N. conceived the theory and M.E., H.L. and A.D. performed the simulations. P.B., U.N. and S.T. wrote the manuscript with help of all co-authors.

Corresponding author

Correspondence to [P. Baum](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Georg Woltersdorf 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 X-ray characterization of the nickel thin-film structure.](#)

a, X-ray reflectivity data and fit of the sample using a four-layer model with the scattering length density profile shown in the inset. The dashed lines in the inset indicate the slab model of the corresponding layers. The best fit parameters obtained by fitting the XRR intensities are shown in the table. The errors are estimated by a 5% increase over the optimum logarithmic figure of merit. **b**, Out-of-plane XRD scan in the angular regime of $40^\circ \leq 2\theta \leq 60^\circ$. The observed intensities at $2\theta \approx 50.53^\circ$ and $2\theta \approx 52.13^\circ$ correspond to Cu(002) and Ni(002). The lack of any Cu(111) and Ni(111) intensities shows the epitaxial growth. The inset graph shows the rocking-

scans over the Cu(002) and Ni(002) peak positions. **c**, In-plane XRD scan at an inclination angle $\Delta\chi = 54.51^\circ$. The intensities at $2\theta \approx 43.41^\circ$ and $2\theta \approx 44.47^\circ$ correspond to the Cu(111) and Ni(111) reflections, respectively. **d**, ϕ scans for the Ni(111), Cu(111) and Si(111) ip peaks, obtained at an inclination angle of $\Delta\chi = 54.74^\circ$. A clear fourfold symmetry of the Cu(111) and Ni(111) ip reflections is observed with an offset angle of 45° to the Si(111) substrate reflections. For reasons of clarity, the scans are shifted in intensity by a factor of two each. **e**, ϕ scans for the Cu(111) and Ni(111) reflections, obtained at inclination angles of $\Delta\chi = 15.80^\circ$, $\Delta\chi = 54.74^\circ$ and $\Delta\chi = 79.00^\circ$. For clarity, the scans are shifted in intensity by 0.1 each. Cu(111) intensities are shown in the angular regime of $0^\circ \leq \phi \leq 180^\circ$, while the Ni(111) intensities are shown for $180^\circ \leq \phi \leq 360^\circ$.

Extended Data Fig. 2 Rocking curve, magnetic hysteresis and optical penetration depth.

a, Rocking scan data obtained with the femtosecond electron beam. Shown is the Ni(200) peak when rotating the specimen around the [010] axis. **b**, Magnetic hysteresis curve of our nickel specimen, obtained by an in-plane SQUID measurement. **c**, Simulated optical energy disposition as a function of penetration depth. Upper panel: solid line, normalized electric field amplitude; dotted line, real part of the refractive index; dashed line, imaginary part of the refractive index. The laser comes from the left. Lower panel: absorption as a function of depth. The green, red, blue and grey areas denote nickel, copper, silicon and NiO_x , respectively.

Extended Data Fig. 3 Second-harmonic-generation FROG measurements of the optical pulses after compression.

a, Measured FROG trace. **b**, Retrieved FROG trace at 0.3% FROG error. **c**, Evaluated spectrum (blue) and spectral phase (green). **d**, Retrieved pulse shape (blue) with temporal phase (green). The pulse duration is 93 fs.

Extended Data Fig. 4 Numerical simulation of heat flow.

a, Temperature profile at 20 ps after laser excitation. Drawing is not to scale. **b**, Radial profile of the temperature increase ΔT due to quasi-static heat accumulation. **c**, Cooling dynamics of the front surface at $r = 0$.

Extended Data Fig. 5 Magneto-optical Faraday effect and fluence dependency of the electron diffraction results.

a, Magnetic hysteresis curves for a negative (black) and slightly positive pump–probe delay (blue). **b**, Magnetization as a function of delay time. **c**, Debye–Waller effect as a function of the applied laser excitation fluence. **d**, Bragg spot anisotropy as a function of the applied laser excitation fluence. **e**, Simulated anisotropy as function of the degree of demagnetization.

Extended Data Fig. 6 Changes of Bragg spots angles as a function of pump–probe delay.

Dots, changes $\Delta\alpha_x$ along the x axis; squares, changes $\Delta\alpha_y$ along the y axis (see Fig. 1d).

Extended Data Fig. 7 Absence of beam deflection effects.

a, Two time-delayed electron beams on the screen. **b**, Intensity changes in the reference pulse (black) and probe pulse (blue), showing a Debye–Waller effect in the probe beam only. **c**, Differences of the beam positions before and after laser excitation as a function of the pump–probe delay, converted to angle changes at the specimen. All changes remain below 5 μ rad.

Extended Data Fig. 8 Control experiment.

Analysis of the anisotropy of the silicon and copper spots as function of the pump–probe delay. **a**, isotropic Debye–Waller effect of Ni. **b**, Anisotropy of Si and Cu as a function of time.

Extended Data Fig. 9 Monte Carlo analysis of the time constants.

a, Distribution of the fitted response times for the Bragg spot asymmetry (blue) and the Debye–Waller effect (black). **b**, Correlation plot of the asymmetry fit parameters.

Extended Data Fig. 10 Additional molecular dynamics simulations results.

a, Finite-size effects of the anisotropy of crystallographically equivalent peaks comparing open boundary conditions (OBC) with periodic ones (PBC). For OBC a finite-size effect is observed: the relaxation time of the contrast increases with system size. PBC do not show this effect. **b**, Long-time evolution for $N=50$ testing the three cases (OBC, PBC, and global rotation according to the Einstein–de Hass effect (EdH)) (blue, solid lines). Also shown is the anisotropy of the mean-squared velocities $\langle 2\{v\}_y^2/\{v\}_x^2+\{v\}_z^2 \rangle$ (green, dotted lines). **c**, Temperature dependence; anisotropy of crystallographically equivalent peaks for the same angular momentum L_0 (same demagnetization) but different energy transfers to the lattice, leading to a temperature increase of $\Delta T=15$ K and 60 K, respectively.

Supplementary information

Peer Review File

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Tauchert, S.R., Volkov, M., Ehberger, D. *et al.* Polarized phonons carry angular momentum in ultrafast demagnetization. *Nature* **602**, 73–77 (2022). <https://doi.org/10.1038/s41586-021-04306-4>

- Received: 08 June 2021
- Accepted: 01 December 2021
- Published: 02 February 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04306-4>

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

[Lost magnetism pinned on atomic rotations](#)

- Georg Woltersdorf

News & Views 02 Feb 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04306-4>

- Article
- [Published: 16 December 2021](#)

Selective sulfidation of metal compounds

- [Caspar Stinn ORCID: orcid.org/0000-0002-0084-249X¹](#) &
- [Antoine Allanore ORCID: orcid.org/0000-0002-2594-0264¹](#)

[Nature](#) volume **602**, pages 78–83 (2022)

- 5734 Accesses
- 3 Citations
- 122 Altmetric
- [Metrics details](#)

Subjects

- [Chemical engineering](#)
- [Metals and alloys](#)

Abstract

There is urgent, unprecedented demand for critical by-product and co-product metallic elements for the infrastructure (magnets, batteries, catalysts and electronics) needed to power society with renewable electricity^{1,2,3}. However, the extraction of *d*-block and *f*-block metals from mineral and recycled streams is thermodynamically difficult, typically requiring complete dissolution of the materials, followed by liquid–liquid

separation using metal-ion complexing or chelating behaviour^{4,5}. The similar electronic structure of these metals results in poor separation factors, necessitating immense energy, water and chemicals consumption^{6,7,8}. Here a metal-processing approach based on selective anion exchange is proposed. Several simple process levers (gas partial pressure, gas flowrate and carbon addition) are demonstrated to selectively sulfidize a target metal from a mixed metal-oxide feed. The physical and chemical differences between the sulfide and oxide compounds (for example, density, magnetic susceptibility and surface chemistry) can then be exploited for vastly improved separation compared with liquid–liquid methods⁹. The process conditions of sulfidation are provided for 56 elements and demonstrated for 15 of them. An assessment of the environmental and economic impacts suggests a path towards 60–90% reductions in greenhouse gas emissions while offering substantial capital cost savings compared with liquid–liquid hydrometallurgy.

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: The S₂/SO₂ ratio for selective sulfidation of metals and the corresponding gaseous space time, space velocity and CDSR levers.

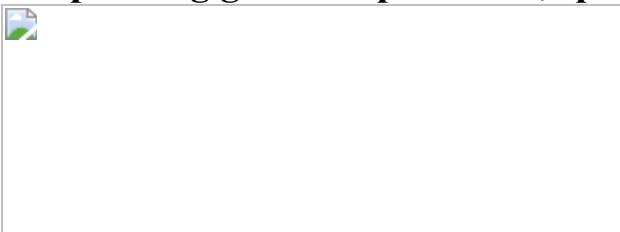


Fig. 2: Application of selective sulfidation for LIB recycling, rare-earth magnet recycling and rare-earth mineral processing.

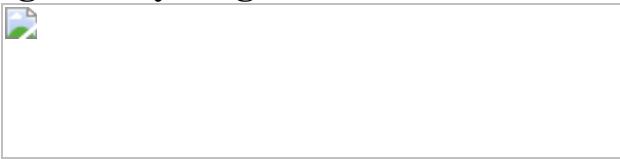
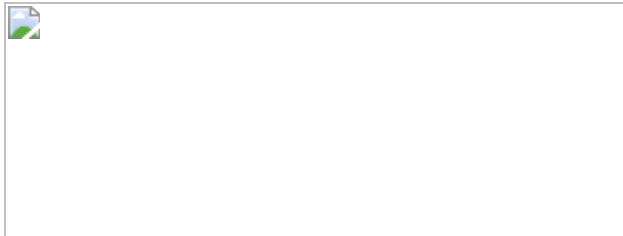


Fig. 3: Capital costs and environmental impact estimates for selective sulfidation coupled with physical separation, compared with conventional hydrometallurgical processing.



Data availability

All data relevant to the results presented are included herein, within the [Supplementary Information](#) or available from the corresponding author upon request.

Code availability

Spreadsheets and code supporting technoeconomic analysis and life cycle assessment are available on the Harvard Dataverse repository at <https://doi.org/10.7910/DVN/193PW2>.

References

1. 1.

Cheisson, T. & Schelter, E. J. Rare earth elements: Mendeleev's bane, modern marvels. *Science* **363**, 489–493 (2019).

2. 2.

Harper, G. et al. Recycling lithium-ion batteries from electric vehicles. *Nature* **575**, 75–86 (2019).

3. 3.

Enriquez, M. A. et al. Mineral supply for sustainable development requires resource governance. *Nature* **543**, 367–372 (2017).

4. 4.

Reck, B. K. & Graedel, T. E. Challenges in metal recycling. *Science* **337**, 690–696 (2012).

5. 5.

Olivetti, E. A. & Cullen, J. M. Toward a sustainable materials system. *Science* **360**, 1396–1398 (2018).

6. 6.

Ciez, R. E. & Whitacre, J. F. Examining different recycling processes for lithium-ion batteries. *Nat. Sustain.* **2**, 148–156 (2019).

7. 7.

K Lee, J. C. & Wen, Z. Pathways for greening the supply of rare earth elements in China. *Nat. Sustain.* **1**, 598–605 (2018).

8. 8.

Zhao, B., Zhang, J. & Schreiner, B. *Separation Hydrometallurgy of Rare Earth Elements* (Springer, 2016).

9. 9.

Bailey, G. et al. Review and new life cycle assessment for rare earth production from bastnäsite, ion adsorption clays and lateritic monazite. *Resour. Conserv. Recycl.* **155**, 104675 (2020).

10. 10.

Norgate, T. & Jahanshahi, S. Low grade ores—smelt, leach or concentrate? *Miner. Eng.* **23**, 65–73 (2010).

11. 11.

Yin, X. et al. Rare earth separations by selective borate crystallization. *Nat. Commun.* **8**, 14438 (2017).

12. 12.

Flytzani-Stephanopoulos, M., Sakbodin, M. & Wang, Z. Regenerative adsorption and removal of H₂S from hot fuel gas streams by rare earth oxides. *Science* **312**, 1508–1510 (2006).

13. 13.

Valsamakis, I. & Flytzani-Stephanopoulos, M. Sulfur-tolerant lanthanide oxysulfide catalysts for the high-temperature water-gas shift reaction. *Appl. Catal. B* **106**, 255–263 (2011).

14. 14.

Pease, J. D., Curry, D. C. & Young, M. F. Designing flotation circuits for high fines recovery. *Miner. Eng.* **19**, 831–840 (2006).

15. 15.

Pease, J. D., Young, M. F., Curry, D. & Johnson, N. W. Improving fines recovery by grinding finer. *Trans. Inst. Min. Metall. Sect. C* **119**, 216–222 (2010).

16. 16.

Han, J. et al. Effects of sodium salts on the sulfidation of lead smelting slag. *Miner. Eng.* **108**, 1–11 (2017).

17. 17.

Zhang, W., Zhou, Y., Zhu, J. & Pan, Y. New clean process for barium sulfide preparation by barite reduction with elemental sulfur. *Ind. Eng. Chem. Res.* **53**, 5646–5651 (2014).

18. 18.

Zhang, W. et al. Reaction mechanism study of new scheme using elemental sulfur for conversion of barite to barium sulfide. *Powder Technol.* **360**, 1348–1354 (2020).

19. 19.

Kaneko, T., Yashima, Y., Ahmadi, E., Natsui, S. & Suzuki, R. O. Synthesis of Sc sulfides by CS₂ sulfurization. *J. Solid State Chem.* **285**, 121268 (2020).

20. 20.

Ahmadi, E. & Suzuki, R. O. An innovative process for production of Ti metal powder via TiS_x from TiN. *Metall. Mater. Trans. B* **51B**, 140–148 (2020).

21. 21.

Afanasiev, P. et al. Preparation of the mixed sulfide Nb₂Mo₃S₁₀ catalyst from the mixed oxide precursor. *Catal. Lett.* **64**, 59–63 (2000).

22. 22.

Ahmad, S., Rhamdhani, M. A., Pownceby, M. I. & Bruckard, W. J. Thermodynamic assessment and experimental study of sulphidation of ilmenite and chromite. *Trans. Inst. Min. Metall. Sect. C* **123**, 165–177 (2014).

23. 23.

Harris, C. T., Peacey, J. G. & Pickles, C. A. Selective sulphidation and flotation of nickel from a nickeliferous laterite ore. *Miner. Eng.* **54**, 21–31 (2013).

24. 24.

Liu, W., Zhu, L., Han, J., Jiao, F. & Qin, W. Sulfidation mechanism of ZnO roasted with pyrite. *Sci. Rep.* **8**, 9516 (2018).

25. 25.

Sohn, H. Y. & Fan, D.-Q. On the initial rate of fluid–solid reactions. *Met. Mater. Trans. B* **48B**, 1827–1832 (2017).

26. 26.

Zagorac, D., Doll, K., Zagorac, J., Jordanov, D. & Matovic, B. Barium sulfide under pressure: discovery of metastable polymorphs and investigation of electronic properties on ab initio level. *Inorg. Chem.* **56**, 10644–10654 (2017).

27. 27.

Sohn, H. Y. & Kim, B.-S. A novel cyclic process using CaSO₄/CaS pellets for converting sulfur dioxide to elemental sulfur without generating secondary pollutants: Part I. feasibility and kinetics of the reduction of sulfur dioxide with calcium-sulfide pellets. *Metall. Mater. Trans. B* **33B**, 711–716 (2002).

28. 28.

Sahu, S. K., Chmielowiec, B. & Allanore, A. Electrolytic extraction of copper, molybdenum and rhenium from molten sulfide electrolyte. *Electrochim. Acta* **243**, 382–389 (2017).

29. 29.

Brown, A. M. & Ashby, M. F. Correlations for diffusion constants. *Acta Metall.* **28**, 1085–1101 (1980).

30. 30.

Nassar, N. T., Graedel, T. E. & Harper, E. M. By-product metals are technologically essential but have problematic supply. *Sci. Adv.* **1**, e1400180 (2015).

31. 31.

Olivetti, E. A., Ceder, G., Gaustad, G. G. & Fu, X. Lithium-ion battery supply chain considerations: analysis of potential bottlenecks in critical metals. *Joule* **1**, 229–243 (2017).

32. 32.

Dunn, J. B., Gaines, L., Sullivan, J. & Wang, M. Q. Impact of recycling on cradle-to-gate energy consumption and greenhouse gas emissions of automotive lithium-ion batteries. *Environ. Sci. Technol.* **46**, 12704–12710 (2012).

33. 33.

Shi, J. et al. Sulfation roasting mechanism for spent lithium-ion battery metal oxides under $\text{SO}_2\text{--O}_2\text{--Ar}$ atmosphere. *JOM* **71**, 4473–4481 (2019).

34. 34.

Stinn, C. & Allanore, A. In *Ni-Co 2021: The 5th International Symposium on Nickel and Cobalt* (eds Anderson, C. et al.) 99–110 (Springer Nature, 2021).

35. 35.

Wagner, M.-E. & Allanore, A. Chemical thermodynamic insights on rare-earth magnet sludge recycling. *ISIJ Int.* **60**, 2339–2349 (2020).

36. 36.

Narayanan, R. P., Kazantzis, N. K. & Emmert, M. H. Selective process steps for the recovery of scandium from Jamaican bauxite residue (red mud). *ACS Sustain. Chem. Eng.* **6**, 1478–1488 (2018).

37. 37.

Jowitt, Si. M., Werner, T. T., Weng, Z. & Mudd, G. M. Recycling of the rare earth elements. *Curr. Opin. Green Sustain. Chem.* **13**, 1–7

(2018).

38. 38.

Wang, J. & Hu, H. Selective extraction of rare earths and lithium from rare earth fluoride molten-salt electrolytic slag by sulfation. *Miner. Eng.* **160**, 106711 (2021).

39. 39.

Binnemans, K., Jones, P. T., Müller, T. & Yurramendi, L. Rare earths and the balance problem: how to deal with changing markets? *J. Sustain. Metall.* **4**, 126–146 (2018).

40. 40.

Firdaus, M., Rhamdhani, M. A., Durandet, Y., Rankin, W. J. & McGregor, K. Review of high-temperature recovery of rare earth (Nd/Dy) from magnet waste. *J. Sustain. Metall.* **2**, 276–295 (2016).

41. 41.

Lin, X. et al. A novel application of hematite precipitation for high effective separation of Fe from Nd–Fe–B scrap. *Sci. Rep.* **9**, 18362 (2019).

42. 42.

Jönsson, C. et al. The extraction of NdFeB magnets from automotive scrap rotors using hydrogen. *J. Clean. Prod.* **277**, 124058 (2020).

43. 43.

Rasheed, M. Z. et al. Review of the liquid metal extraction process for the recovery of Nd and Dy from permanent magnets. *Metall. Mater. Trans. B* **52**, 1213–1227 (2021).

44. 44.

Li, X. Z. et al. A supramolecular lanthanide separation approach based on multivalent cooperative enhancement of metal ion selectivity. *Nat. Commun.* **9**, 547 (2018).

45. 45.

Jordens, A., Cheng, Y. P. & Waters, K. E. A review of the beneficiation of rare earth element bearing minerals. *Miner. Eng.* **41**, 97–114 (2013).

46. 46.

Chi, R., Li, Z., Peng, C., Gao, H. & Xu, Z. Preparation of enriched cerium oxide from bastnasite with hydrochloric acid by two-step leaching. *Metall. Mater. Trans. B* **37**, 155–160 (2006).

47. 47.

Merritt, R. R. High temperature methods for processing monazite: II. Reaction with sodium carbonate. *J. Less Common Met.* **166**, 211–219 (1990).

48. 48.

Woods, D. R. in *Rules of Thumb in Engineering Practice* 376–436 (Wiley-VCH, 2007).

49. 49.

Nuss, P. & Eckelman, M. J. Life cycle assessment of metals: a scientific synthesis. *PLoS ONE* **9**, e101298 (2014).

50. 50.

Skinner, B. J. Earth resources (minerals/metals/ores/geochemistry/mining). *Proc. Natl Acad. Sci. USA* **76**, 4212–4217 (1979).

51. 51.

Jacob, K. T. & Iyengar, G. N. K. Thermodynamic study of Fe_2O_3 – $\text{Fe}_2(\text{SO}_4)_3$ equilibrium using an oxyanionic electrolyte (Na_2SO_4 –I). *Metall. Trans. B* **17**, 323–329 (1986).

52. 52.

Hsieh, K. C. & Chang, Y. A. A solid-state EMF study of ternary Ni–S–O, Fe–S–O, and quaternary Fe–Ni–S–O. *Metall. Trans. B* **17**, 133–146 (1986).

53. 53.

Dwivedi, R. K. & Kay, D. A. R. Thermodynamics of the oxidation of rare earth oxysulfides at high temperatures. *Metall. Trans. B* **15**, 523–528 (1984).

54. 54.

Akila, R., Jacob, K. T. & Shukla, A. K. Gibbs energies of formation of rare earth oxysulfides. *Metall. Trans. B* **18B**, 163–168 (1987).

55. 55.

Dwivedi, R. K. *Determination of the Thermodynamic Properties of Rare Earth-Oxygen-Sulfur Systems at High Temperatures*. PhD Thesis, McMaster Univ. (1982).

56. 56.

Suponitskii, Y. L., Kuz'micheva, G. M. & Eliseev, A. A. Lanthanide oxide sulphides. *Russ. Chem. Rev.* **57**, 209–220 (1988).

57. 57.

Wang, M. Enthalpy of formation of LiNiO_2 , LiCoO_2 and their solid solution, $\text{LiNi}_{1-x}\text{Co}_x\text{O}_2$. *Solid State Ionics* **166**, 167–173 (2004).

58. 58.

Chang, K., Hallstedt, B. & Music, D. Thermodynamic and electrochemical properties of the Li–Co–O and Li–Ni–O systems. *Chem. Mater.* **24**, 97–105 (2011).

59. 59.

Konings, R. J. M. et al. The thermodynamic properties of the *f*-elements and their compounds. Part 2. The lanthanide and actinide oxides. *J. Phys. Chem. Ref. Data* **43**, 013101 (2014).

60. 60.

Kriklya, A. I., Bolgar, A. S. & Pribyl'skii, N. Y. Heat capacity and enthalpy of γ - Dy_2S_3 over a wide range of temperature. *Sov. Powder Metall. Met. Ceram.* **31**, 697–700 (1992).

61. 61.

Chakraborti, N. Modified predominance area diagrams for the Fe–S–O system. *Can. J. Chem. Eng.* **61**, 763–765 (1983).

62. 62.

Madon, N. & Strickland-constable, R. F. Production of carbon disulfide. *Ind. Eng. Chem.* **50**, 1189–1192 (1958).

63. 63.

Fogler, H. S. *Elements of Chemical Reaction Engineering* (Prentice Hall, 2016).

64. 64.

Sohn, H. Y. Review of fluid-solid reaction analysis—Part 2: single porous reactant solid. *Can. J. Chem. Eng.* **97**, 2068–2076 (2019).

65. 65.

Sohn, H. Y. & Szekely, J. A structural model for gas–solid reactions with a moving boundary—III. *Chem. Eng. Sci.* **27**, 763–778 (1972).

66. 66.

Sohn, H. Y. & Szekely, J. The effect of intragrain diffusion on the reaction between a porous solid and a gas. *Chem. Eng. Sci.* **29**, 630–634 (1974).

67. 67.

Ishida, M. & Wen, C. Y. Comparison of kinetic and diffusional models for solid–gas reactions. *AICHE J.* **14**, 311–317 (1968).

68. 68.

Sohn, H. Y. & Perez-Fontes, S. E. Application of the law of additive reaction times to fluid–solid reactions in porous pellets with changing effective diffusivity. *Met. Mater. Trans. B* **41B**, 1261–1267 (2010).

69. 69.

Hirschfelder, J. O., Curtiss, C. F. & Bird, R. B. *Molecular Theory of Gases and Liquids* (John Wiley, 1954).

70. 70.

Berard, M. F., Wirkus, C. D. & Wilder, D. R. Diffusion of oxygen in selected monocrystalline rare earth oxides. *J. Am. Ceram. Soc.* **51**, 643–647 (1968).

71. 71.

Cussler, E. L. *Diffusion: Mass Transfer in Fluid Systems* (Cambridge Univ. Press, 2009).

72. 72.

Deen, W. M. *Introduction to Chemical Engineering Fluid Dynamics* (Cambridge Univ. Press, 2016).

73. 73.

Freer, R. Self-diffusion and impurity diffusion in oxides. *J. Mater. Sci.* **15**, 803–824 (1980).

74. 74.

Liao, B. Q., Wan, C. R. & Wang, J. A concept for the estimation of HETS for rare earth separations in extraction columns. *Sep. Sci. Technol.* **39**, 2597–2607 (2004).

75. 75.

Flett, D. S. Solvent extraction in hydrometallurgy: the role of organophosphorus extractants. *J. Organomet. Chem.* **690**, 2426–2438 (2005).

76. 76.

Towler, G. & Sinnott, R. *Chemical Engineering Design* (Elsevier, 2013).

77. 77.

Cheng, C. Y. & Zhu, Z. Solvent extraction technology for the separation and purification of niobium and tantalum: A review. *Hydrometallurgy* **107**, 1–12 (2011).

78. 78.

Dincer, I. & Bicer, Y. *Mitacs Accelerate Project Final Report* (2015).

79. 79.

Green, D. W. & Perry, R. H. *Perry's Chemical Engineer's Handbook* (McGraw-Hill, 2008).

80. 80.
- Christensen, P. & Dysert, L. *Cost Estimate Classification System as Applied in Engineering, Procurement, and Construction for the Process Industries. AACE International Recommended Practice No. 18R-97 COST, TCM Framework: 7.3 - Cost Estimating and Budgeting* (2005).
81. 81.
- Mineral Commodity Summaries 2021* (USGS, 2021).
82. 82.
- Annual Coal Report 2020* (USEIA & USDOE, 2021).
83. 83.
- Henry Hub Natural Gas Spot Price* (USEIA, 2021).
84. 84.
- Misaka, T. & Mochizuki, Y. In *Electrosatic Precipitation* (ed. Yan, K.) 518–522 (Springer, 2009).
85. 85.
- Bleiwas, D. I. *Estimated Water Requirements for the Conventional Flotation of Copper Ores* USGS Open-File Report 2012-1089 (2012).
86. 86.
- Bleiwas, D. I. *Estimates of Electricity Requirements for the Recovery of Mineral Commodities, with Examples Applied to Sub-Saharan Africa* USGS Open-File Report 2011-1253 (2011).
87. 87.

Bezuidenhout, G. A., Davis, J., van Beek, B. & Eksteen, J. J. Operation of a concentrated mode dual-alkali scrubber plant at the Lonmin smelter. *J. South. African Inst. Min. Metall.* **112**, 657–665 (2012).

88. 88.

King, M. J., Davenport, W. G. & Moats, M. S. *Sulfuric Acid Manufacture—Analysis, Control and Optimization* (2013).

89. 89.

ISO 14044 (ISO, 2006).

90. 90.

Ecoinvent-Association *ecoinvent 3.6* (Ecoinvent Center, 2019);
www.ecoinvent.org

91. 91.

Klett, C., Reeb, B., Missalla, M. & Schmidt, H.-W. in *Light Metals 2011* (ed. Lindsay, S. J.) 125–130 (John Wiley, 2011).

92. 92.

Fu, C. & Gundersen, T. Using exergy analysis to reduce power consumption in air separation units for oxy-combustion processes. *Energy* **44**, 60–68 (2012).

93. 93.

de Bakker, J. Energy use of fine grinding in mineral processing. *Metall. Mater. Trans. E* **1**, 8–19 (2014).

94. 94.

TRACI 2.1: Tool for Reduction and Assessment of Chemicals and Other Environmental Impacts (USEPA, 2014);

<https://www.epa.gov/chemical-research/tool-reduction-and-assessment-chemicals-and-other-environmental-impacts-traci>

95. 95.

Carbon Dioxide Emissions Coefficients (USEIA, 2021).

96. 96.

Stinn, C. & Allanore, A. Selective sulfidation of metal compounds—supporting computing files. *Harvard Dataverse* <https://doi.org/10.7910/DVN/193PW2> (2021).

Acknowledgements

We thank the US Department of Energy and the US National Science Foundation for their financial support; H. Higuchi and Sumitomo Metal Mining for providing samples of scandium oxide; and K. Daehn, A. Culbertson, T. Close, L. Rush, A. Caldwell and M. E. Wagner for their insight.

Author information

Affiliations

1. Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Casper Stinn & Antoine Allanore

Contributions

C.S. and A.A. designed the project, wrote the manuscript and prepared the figures. C.S. carried out the experiments, modelling and analysis.

Corresponding author

Correspondence to [Antoine Allanore](#).

Ethics declarations

Competing interests

The authors have filed for intellectual property protection pertaining to work herein.

Peer review

Peer review information

Nature thanks Junwei Han, Fanran Meng, Fiseha Tesfaye 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 Gas species partial pressures as thermodynamic levers to control compound stability.](#)

a, Sc–O–S Kellogg diagram ($T = 1,000\text{ }^{\circ}\text{C}$, 1 atm), illustrating the role of $\langle\langle P \rangle\rangle_{\{\{\rm{S}\}\}\{2\}}\rangle\langle\langle P \rangle\rangle_{\{\{\rm{O}\}\}\{2\}}\rangle\langle\langle P \rangle\rangle_{\{\{\rm{SO}\}\}\{2\}}\rangle$ on scandium compound stability. **b**, Fe–O–S–C predominance diagram, illustrating the role of $P_{\rm{CO}}$, $\langle\langle P \rangle\rangle_{\{\{\rm{CO}\}\}\{2\}}\rangle\langle\langle P \rangle\rangle_{\{\{\rm{S}\}\}\{2\}}\rangle\langle\langle P \rangle\rangle_{\{\{\rm{SO}\}\}\{2\}}\rangle$ on iron compound stability ($\langle\langle P \rangle\rangle_{\{\{\rm{S}\}\}\{2\}}\rangle$ fixed at 0.05 atm). The sulfide becomes the only

stable compound of Fe in the presence of carbon at increasing $\langle\{P\}_{\{\{\rm S\}\}}_{\{2\}}\rangle$ values, as for many critical metals including Co, In, Mn, Ni, Sn, W and Zn. In **a**, **b**, solid lines correspond to phase domains, dotted lines correspond to equilibrium gas compositions at 1 atm total pressure and dashed lines correspond to $\langle\{P\}_{\{\{\rm SO\}\}}_{\{2\}}\rangle$.

Extended Data Fig. 2 Reaction and solution contributions to sulfidation selectivity.

a, b, (See Supplementary equations (1), (2), (3), (4), (5), (6) and (7) in Supplementary Information). When $\langle\{\psi\}_{\{rxn\}}\gg\langle\psi\}_{\{sol\}}\rangle$, the sulfidation thermodynamics are reaction-dominated, solution effects are minimal and the sulfidation thermodynamics are well described by those of the pure compound. When $\langle\{\psi\}_{\{sol\}}\gg\langle\psi\}_{\{rxn\}}\rangle$, the sulfidation thermodynamics are solution-dominated and reaction effects are minimal, the sulfidation thermodynamics are not well described by those of the pure compound, and knowledge of the solution behaviour is essential to determine sulfidation spontaneity. **c, d**, Equilibrium S_2/SO_2 ratio as a function of sulfur partial pressure and temperature at 1 atm for a gas consisting exclusively of sulfur–oxygen-containing species at equilibrium. As derived in the [Supplementary Information](#), the equilibrium $\langle\{P\}_{\{\{\rm S\}\}}_{\{2\}}\rangle/\langle\{P\}_{\{\{\rm SO\}\}}_{\{2\}}\rangle$ ratio that satisfies $\langle\psi\rangle$ corresponds to $\langle\{P\}_{\{\{\rm S\}\}}_{\{2\}}\rangle/\langle\{P\}_{\{\{\rm SO\}\}}_{\{2\}}\rangle_{\text{crit}}$.

Extended Data Fig. 3 Experimental devices used for sulfidation and recovery.

a, A packed-bed, flow-through reactor (12), is placed in a furnace (1) equipped with an alumina tube (2) to conduct selective sulfidation using elemental sulfur evaporated from the bottom of the tube (11). **b**, A Hallimond cell is used for the recovery by flotation of the sulfide phases formed in **a**.

Extended Data Fig. 4 Lanthanum oxide sulfidation reaction kinetics as measured in a graphite packed bed reactor.

a, Conversion versus time as a function of temperature shows reaction rate increasing with temperature, consistent with the notion of thermal activation of the reaction. **b**, Sulfur partial pressure corresponding to conversion rate in **a**. **c**, Modified Sherwood number $(Sh\prime)$ for lanthanum oxide sulfidation kinetics experiments. Following Sohn's criteria⁶⁵, for $(Sh\prime) > 30$ external mass transfer limitations to the observed reaction rate are negligible. **d**, Fluid–solid reaction modulus $(\hat{\sigma})$ for lanthanum oxide sulfidation kinetic experiments. For $(\hat{\sigma})^2 < 0.01$, intergrain diffusion limitations to the observed rate of reaction are negligible⁶⁵. Intragrain diffusion limitations are addressed in the [Supplementary Information](#) and Supplementary Table 3. **e**, Comparison of the kinetically limited ([Supplementary Information](#)) rate of oxygen liberation for sulfidation versus sulfur gas concentration. The slope of natural log of the oxygen liberation rate vs the natural log of the sulfur concentration is the reaction order, observed to be approximately first order with respect to sulfur. **f**, Arrhenius plot of the natural log of the reaction rate constant vs inverse temperature. The activation energy is found to be 114 kJ mol⁻¹, with a pre-exponential factor of 1.08×10^4 s⁻¹.

Extended Data Fig. 5 Selective sulfidation of LiNi_{1/3}Mn_{1/3}Co_{1/3}O₂ (abbr. NMC111 or NMC333) separated into Ni-rich sulfide (1), Co-rich sulfide (2) and Mn oxysulfide (3) phases.

a, Optical dark field image showing Ni, Co and Mn-rich phases coalesced to approximately 100–500 μm in size that support physical separation following sulfidation at 1,000 °C. **b**, SEM–EDS analysis reveals distinct Ni-rich sulfide (1, $Ni_{0.75}Co_{0.25}S$), Co-rich sulfide (2, $Ni_{0.33}Co_{0.67}S$) and Mn oxysulfide (3, $MnO_{0.2}S_{0.8}$) phases. **c**, SEM–EDS maps illustrate minimal Mn inclusion in Ni–Co phases and vice versa post-sulfidation.

Extended Data Fig. 6 Selective sulfidation of calcined rare-earth, iron, boron ((Nd,Pr,Dy)-Fe-B) magnet separated into an iron-rich sulfide phase (1) with neodymium-rich oxide (2) inclusions.

a, Upon sulfidation, calcined (Nd,Pr,Dy)–Fe–B particles 90–212 µm in size sintered to approximately 1–2 mm in size, with Nd-rich oxide (1) regions approximately 20–100 µm in size that are large enough to support physical separation from the bulk Fe-rich sulfide (2) phases (SEM/BEC image). **b**, SEM–EDS analysis reveals minimal inclusion of Fe,Dy into the Nd,Pr-rich inclusions and vice versa post-sulfidation.

Extended Data Fig. 7 Sulfidative sintering and selective sulfidation of synthetic defluorinated, dethoriated, light rare-earth element bastnaesite (Ln_2O_3) separates into neodymium-rich and lanthanum-rich phases.

a, Ln_2O_3 particles (25–45 µm) sulfidized with carbon in an alumina flow-through packed-bed reactor at 1,400 °C sintered to approximately 100–300 µm in size (darkfield optical image). **b**, $\text{Ln}_{10}\text{OS}_{14}$ (1) and $\text{Ln}_2\text{O}_2\text{S}$ (2) approximately 20–100 µm in size are observed to form upon sulfidation, large enough to support physical separation (polarized optical image, 90°). **c**, EPMA/WDS elemental analysis reveals sulfidation is selective, with Nd enriched in the oxygen-rich $\text{Ln}_2\text{O}_2\text{S}$ phase (2) and La enriched in the sulfur-rich $\text{Ln}_{10}\text{OS}_{14}$ phase (1).

Extended Data Fig. 8 Flowsheet of a generic selective sulfidation process.

The process consists of selective sulfidation in a multihearth fluidized bed reactor, product comminution and physical separation via froth flotation, and downstream gas handling and treatment via a cyclone separator for solid particle removal and acid plant for SO_2 recovery, and assumes an equimolar, mixed, binary oxide feed. The system boundary for life cycle assessment is depicted, over the impact categories of global warming potential (GWP), terrestrial acidification (TA) and water resource depletion (WRD), for a functional unit of 1 kg of selective sulfidation feed. The impacts of flows originating within the system boundary are evaluated from the cradle to usage in the process, while the impacts of flows originating outside the system boundary are evaluated from the system gate to usage in

the process. The impacts of flows exiting within the system boundary are evaluated from outlet of the process to the grave, while flows exiting outside the system boundary are evaluated from production in the process to the system gate.

Extended Data Fig. 9 Capital cost (CAPEX) and operating cost (OPEX) distributions for the generic selective sulfidation process (Extended Data Fig. 8, Supplementary Figs. 1–3).

Distributions for selective sulfidation with and without feed pretreatments and with and without CDSR are determined via Monte Carlo simulation, with probability distributions for CAPEX, OPEX and operating condition parameters described in Supplementary Tables 14, 15. Pretreatment steps for feed drying, sintering and roasting/calcination each marginally increase the CAPEX and OPEX of selective sulfidation by 10% to 20%. CDSR generally decreases CAPEX at the expense of increases in OPEX and environmental impacts (Extended Data Fig. 10).

Extended Data Fig. 10 Global warming potential (GWP), water resource depletion (WRD) and terrestrial acidification (TA) distributions for the generic selective sulfidation process (Extended Data Fig. 8, Supplementary Figs. 1–3), with and without feed pretreatments, with and without CDSR, for a functional unit of 1 kg of selective sulfidation feed.

Distributions are determined via Monte Carlo simulation, with probability distribution for operating condition parameters described in Supplementary Tables 14, 15. The inclusion of feed pretreatments increases GWP by about 50%, WRD by 30% and TA by double, while CDSR increases GWP by a factor of 3–5 times and WRD by a factor of 3–4 times. The bimodal nature of the GWP distribution is due to differences in oxygen content of the three model chemistries considered in the Monte Carlo simulation, highlighting the role of system chemistry in determining environmental impact.

Supplementary information

Supplementary Information

This file contains Supplementary Methods.

Supplementary Tables

This file contains Supplementary Tables 1–33; see Summary tab for table descriptions.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Stinn, C., Allanore, A. Selective sulfidation of metal compounds. *Nature* **602**, 78–83 (2022). <https://doi.org/10.1038/s41586-021-04321-5>

- Received: 17 June 2020
- Accepted: 09 December 2021
- Published: 16 December 2021
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04321-5>

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/s41586-021-04321-5>

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

- Article
- [Published: 02 February 2022](#)

Fluctuation-induced quantum friction in nanoscale water flows

- [Nikita Kavokine](#) [ORCID: orcid.org/0000-0002-8037-7996^{1,2}](#),
- [Marie-Laure Bocquet³](#) &
- [Lydéric Bocquet](#) [ORCID: orcid.org/0000-0003-3577-5335¹](#)

[Nature](#) volume **602**, pages 84–90 (2022)

- 1883 Accesses
- 139 Altmetric
- [Metrics details](#)

Subjects

- [Carbon nanotubes and fullerenes](#)
- [Fluid dynamics](#)
- [Surfaces, interfaces and thin films](#)
- [Two-dimensional materials](#)

Abstract

The flow of water in carbon nanochannels has defied understanding thus far¹, with accumulating experimental evidence for ultra-low friction, exceptionally high water flow rates and curvature-dependent hydrodynamic slippage^{2,3,4,5}. In particular, the mechanism of water–carbon friction remains unknown⁶, with neither current theories⁷ nor classical^{8,9} or ab initio

molecular dynamics simulations¹⁰ providing satisfactory rationalization for its singular behaviour. Here we develop a quantum theory of the solid–liquid interface, which reveals a new contribution to friction, due to the coupling of charge fluctuations in the liquid to electronic excitations in the solid. We expect that this quantum friction, which is absent in Born–Oppenheimer molecular dynamics, is the dominant friction mechanism for water on carbon-based materials. As a key result, we demonstrate a marked difference in quantum friction between the water–graphene and water–graphite interface, due to the coupling of water Debye collective modes with a thermally excited plasmon specific to graphite. This suggests an explanation for the radius-dependent slippage of water in carbon nanotubes⁴, in terms of the electronic excitations of the nanotubes. Our findings open the way for quantum engineering of hydrodynamic flows through the electronic properties of the confining wall.

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: Theory of solid–liquid quantum friction.

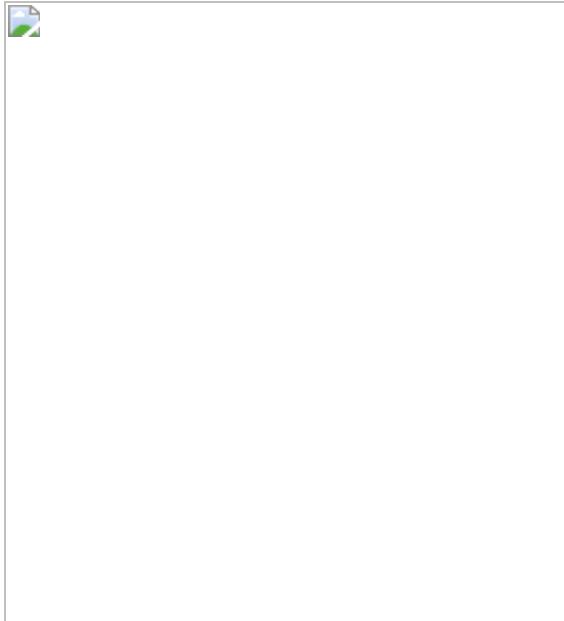


Fig. 2: Surface dielectric response of water.

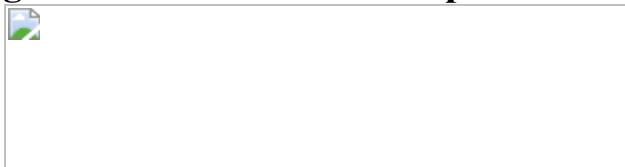
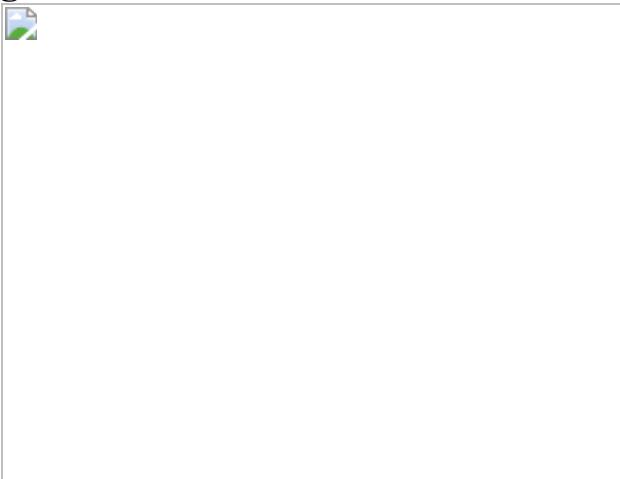


Fig. 3: Quantum friction of water on a jellium surface.



Fig. 4: Quantum friction at the water–carbon interface.



Data availability

The MD simulation data (Fig. 2 and [Supplementary Figs. 1 and 2](#)) are available on Zenodo (<https://doi.org/10.5281/zenodo.5242930>). The rest of the data are included with the paper.

References

1. 1.

Bocquet, L. Nanofluidics coming of age. *Nat. Mater.* **19**, 254–256 (2020).

2. 2.

Holt, J. K. et al. Fast mass transport through sub-2-nanometer carbon nanotubes. *Science* **312**, 1034–1037 (2006).

3. 3.

Whitby, M., Cagnon, L., Thanou, M. & Quirke, N. Enhanced fluid flow through nanoscale carbon pipes. *Nano Lett.* **8**, 2632–2637 (2008).

4. 4.

Secchi, E. et al. Massive radius-dependent flow slippage in carbon nanotubes. *Nature* **537**, 210–213 (2016).

5. 5.

Xie, Q. et al. Fast water transport in graphene nanofluidic channels. *Nat. Nanotech.* **13**, 238–245 (2018).

6. 6.

Faucher, S. et al. Critical knowledge gaps in mass transport through single-digit nanopores: a review and perspective. *J. Phys. Chem. C* **123**, 21309–21326 (2019).

7. 7.

Bocquet, L. & Barrat, J. L. Flow boundary conditions from nano- to micro-scales. *Soft Matter* **3**, 685–693 (2007).

8. 8.

Thomas, J. A. & McGaughey, A. J. Reassessing fast water transport through carbon nanotubes. *Nano Lett.* **8**, 2788–2793 (2008).

9. 9.

Falk, K., Sedlmeier, F., Joly, L., Netz, R. R. & Bocquet, L. Molecular origin of fast water transport in carbon nanotube membranes: superlubricity versus curvature dependent friction. *Nano Lett.* **10**, 4067–4073 (2010).

10. 10.

Tocci, G., Joly, L. & Michaelides, A. Friction of water on graphene and hexagonal boron nitride from ab initio methods: very different slippage despite very similar interface structures. *Nano Lett.* **14**, 6872–6877 (2014).

11. 11.

Kavokine, N., Netz, R. R. & Bocquet, L. Fluids at the nanoscale: from continuum to subcontinuum transport. *Annu. Rev. Fluid Mech.* **53**, 377–410 (2021).

12. 12.

Sam, A. et al. Fast transport of water in carbon nanotubes: a review of current accomplishments and challenges. *Mol. Simul.* **47**, 905–924 (2021).

13. 13.

Maali, A., Cohen-Bouhacina, T. & Kellay, H. Measurement of the slip length of water flow on graphite surface. *Appl. Phys. Lett.* **92**, 2007–2009 (2008).

14. 14.

Misra, R. P. & Blankschtein, D. Insights on the role of many-body polarization effects in the wetting of graphitic surfaces by water. *J. Phys. Chem. C* **121**, 28166–28179 (2017).

15. 15.

Wodtke, A. M., Tully, J. C. & Auerbach, D. J. Electronically non-adiabatic interactions of molecules at metal surfaces: can we trust the Born-Oppenheimer approximation for surface chemistry? *Int. Rev. Phys. Chem.* **23**, 513–539 (2004).

16. 16.

Dou, W. & Subotnik, J. E. Perspective: how to understand electronic friction. *J. Chem. Phys.* **148** (2018).

17. 17.

Sokoloff, J. B. Enhancement of the water flow velocity through carbon nanotubes resulting from the radius dependence of the friction due to electron excitations. *Phys. Rev. E* **97**, 33107 (2018).

18. 18.

Volokitin, A. I. & Persson, B. N. Near-field radiative heat transfer and noncontact friction. *Rev. Mod. Phys.* **79**, 1291–1329 (2007).

19. 19.

Song, X., Chandler, D. & Marcus, R. A. Gaussian field model of dielectric solvation dynamics. *J. Phys. Chem.* **100**, 11954–11959 (1996).

20. 20.

Rammer, J. & Smith, H. Quantum field-theoretical methods in transport theory of metals. *Rev. Mod. Phys.* **58**, 323–359 (1986).

21. 21.

Bocquet, L. & Charlaix, E. Nanofluidics, from bulk to interfaces. *Chem. Soc. Rev.* **39**, 1073–1095 (2010).

22. 22.

Pendry, J. B. Shearing the vacuum – Quantum friction. *J. Phys. Condens. Matter* **9**, 10301–10320 (1997).

23. 23.

Carlson, S., Brünig, F. N., Loche, P., Bonthuis, D. J. & Netz, R. R. Exploring the absorption spectrum of simulated water from MHz to

infrared. *J. Phys. Chem. A* **124**, 5599–5605 (2020).

24. 24.

Sato, T. & Buchner, R. Dielectric relaxation processes in ethanol/water mixtures. *J. Phys. Chem. A* **108**, 5007–5015 (2004).

25. 25.

Koeberg, M., Wu, C. C., Kim, D. & Bonn, M. THz dielectric relaxation of ionic liquid:water mixtures. *Chem. Phys. Lett.* **439**, 60–64 (2007).

26. 26.

Lang, N. D. & Kohn, W. Theory of metal surfaces: charge density and surface energy. *Phys. Rev. B* **1**, 4555–4568 (1970).

27. 27.

Paniagua-Guerra, L. E., Gonzalez-Valle, C. U. & Ramos-Alvarado, B. Effects of the interfacial modeling approach on equilibrium calculations of slip length for nanoconfined water in carbon slits. *Langmuir* **36**, 14772–14781 (2020).

28. 28.

Radha, B. et al. Molecular transport through capillaries made with atomic-scale precision. *Nature* **538**, 222–225 (2016).

29. 29.

Portail, M., Carrere, M. & Layet, J. M. Dynamical properties of graphite and peculiar behaviour of the low-energy plasmon. *Surf. Sci.* **433**, 863–867 (1999).

30. 30.

Laitenberger, P. & Palmer, R. E. Plasmon dispersion and damping at the surface of a semimetal. *Phys. Rev. Lett.* **76**, 1952–1955 (1996).

31. 31.

Pitarke, J. M., Silkin, V. M., Chulkov, E. V. & Echenique, P. M. Theory of surface plasmons and surface-plasmon polaritons. *Rep. Prog. Phys.* **70**, 1–87 (2007).

32. 32.

Lavor, I. R., Cavalcante, L. S. R., Chaves, A., Peeters, F. M. & Van Duppen, B. Probing the structure and composition of van der Waals heterostructures using the nonlocality of Dirac plasmons in the terahertz regime. *2D Mater.* **8**, 015014 (2020).

33. 33.

Tománek, D. et al. Theory and observation of highly asymmetric atomic structure in scanning-tunneling-microscopy images of graphite. *Phys. Rev. B* **35**, 7790–7793 (1987).

34. 34.

Endo, M. et al. Stacking nature of graphene layers in carbon nanotubes and nanofibres. *J. Phys. Chem. Solids* **58**, 1707–1712 (1997).

35. 35.

Franklin, R. E. The structure of graphitic carbons. *Acta Crystallogr.* **4**, 253–261 (1951).

36. 36.

Speck, J. S., Endo, M. & Dresselhaus, M. S. Structure and intercalation of thin benzene derived carbon fibers. *J. Cryst. Growth* **94**, 834–848 (1989).

37. 37.

Charlier, J. C., Blase, X. & Roche, S. Electronic and transport properties of nanotubes. *Rev. Mod. Phys.* **79**, 677–732 (2007).

38. 38.

Xie, Y., Fu, L., Niehaus, T. & Joly, L. Liquid-solid slip on charged walls: the dramatic impact of charge distribution. *Phys. Rev. Lett.* **125**, 014501 (2020).

39. 39.

Pham, T. A., Ping, Y. & Galli, G. Modelling heterogeneous interfaces for solar water splitting. *Nat. Mater.* **16**, 401–408 (2017).

Acknowledgements

We thank A. Robert for help with MD simulations and acknowledge fruitful discussions with A. Robert, B. Douçot, R. Netz, B. Coasne, N. Lorente and B. Rotenberg. L.B. acknowledges funding from the EU H2020 Framework Programme/ERC Advanced Grant agreement number 785911-Shadoks and ANR project Neptune. This work has received the support of ‘Institut Pierre-Gilles de Gennes’, programmes ANR-10-IDEX-0001-02 PSL and ANR-10-LABX-31. We acknowledge the French HPC resources of GENCI for grant number A9-A0070807364. The Flatiron Institute is a division of the Simons Foundation. We acknowledge the inspiration and contributions to science of late Jorge Iribas Cerdá.

Author information

Affiliations

1. Laboratoire de Physique de l’École Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université Paris-Diderot, Sorbonne Paris Cité, Paris, France

Nikita Kavokine & Lydéric Bocquet

2. Center for Computational Quantum Physics, Flatiron Institute, New York, NY, USA

Nikita Kavokine

3. PASTEUR, Département de Chimie, École Normale Supérieure, PSL University, Sorbonne Universités, CNRS, Paris, France

Marie-Laure Bocquet

Contributions

L.B., M.-L.B. and N.K. conceived the project. N.K. developed the theoretical framework. N.K. and L.B. co-wrote the paper, with input from M.-L.B. All authors discussed the results and commented on the manuscript.

Corresponding authors

Correspondence to [Nikita Kavokine](#) or [Lydéric Bocquet](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Mischa Bonn 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.

Supplementary information

Supplementary Information

Supplementary Sections 1–7 including Figs. 1–5. See contents page for details.

Peer Review File

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Kavokine, N., Bocquet, ML. & Bocquet, L. Fluctuation-induced quantum friction in nanoscale water flows. *Nature* **602**, 84–90 (2022).
<https://doi.org/10.1038/s41586-021-04284-7>

- Received: 07 May 2021
- Accepted: 25 November 2021
- Published: 02 February 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04284-7>

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

[Weirdly flowing water finally has an explanation: 'quantum friction'](#)

- Benjamin Thompson
- Shamini Bundell

Nature Podcast 02 Feb 2022

[Water and carbon make a quantum couple](#)

Research Briefing 02 Feb 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04284-7>

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

- Article
- [Published: 02 February 2022](#)

Irreversible synthesis of an ultrastrong two-dimensional polymeric material

- [Yuwen Zeng](#) [ORCID: orcid.org/0000-0003-2619-108X¹](#),
- [Pavlo Gordiichuk¹](#),
- [Takeo Ichihara¹](#),
- [Ge Zhang¹](#),
- [Emil Sandoz-Rosado²](#),
- [Eric D. Wetzel²](#),
- [Jason Tresback³](#),
- [Jing Yang¹](#),
- [Daichi Kozawa](#) [ORCID: orcid.org/0000-0002-0629-5589¹](#),
- [Zhongyue Yang¹](#),
- [Matthias Kuehne](#) [ORCID: orcid.org/0000-0002-5096-7522¹](#),
- [Michelle Quien](#) [ORCID: orcid.org/0000-0003-4990-2577¹](#),
- [Zhe Yuan¹](#),
- [Xun Gong¹](#),
- [Guangwei He¹](#),
- [Daniel James Lundberg](#) [ORCID: orcid.org/0000-0003-3578-6122¹](#),
- [Pingwei Liu¹](#),
- [Albert Tianxiang Liu¹](#),
- [Jing Fan Yang](#) [ORCID: orcid.org/0000-0001-8180-7452¹](#),
- [Heather J. Kulik](#) [ORCID: orcid.org/0000-0001-9342-0191¹](#) &
- [Michael S. Strano](#) [ORCID: orcid.org/0000-0003-2944-808X¹](#)

- 4835 Accesses
- 432 Altmetric
- [Metrics details](#)

Subjects

- [Mechanical properties](#)
- [Two-dimensional materials](#)

Abstract

Polymers that extend covalently in two dimensions have attracted recent attention^{1,2} as a means of combining the mechanical strength and in-plane energy conduction of conventional two-dimensional (2D) materials^{3,4} with the low densities, synthetic processability and organic composition of their one-dimensional counterparts. Efforts so far have proven successful in forms that do not allow full realization of these properties, such as polymerization at flat interfaces^{5,6} or fixation of monomers in immobilized lattices^{7,8,9}. Another frequently employed synthetic approach is to introduce microscopic reversibility, at the cost of bond stability, to achieve 2D crystals after extensive error correction^{10,11}. Here we demonstrate a homogenous 2D irreversible polycondensation that results in a covalently bonded 2D polymeric material that is chemically stable and highly processable. Further processing yields highly oriented, free-standing films that have a 2D elastic modulus and yield strength of 12.7 ± 3.8 gigapascals and 488 ± 57 megapascals, respectively. This synthetic route provides opportunities for 2D materials in applications ranging from composite structures to barrier coating materials.

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: Synthesis and characterization of a 2D polymeric material.



Fig. 2: Characterization of 2DPA-1 nanofilms.

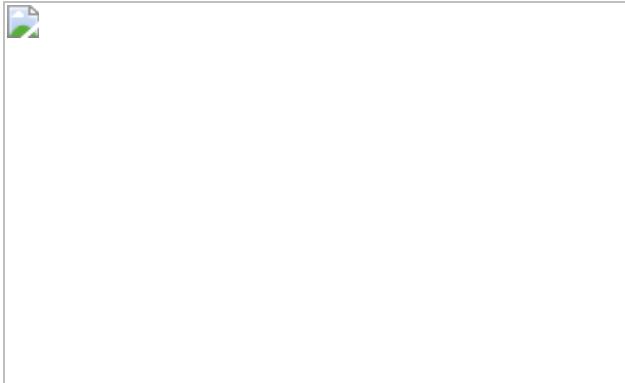


Fig. 3: Molecular alignment in 2DPA-1 nanofilms.

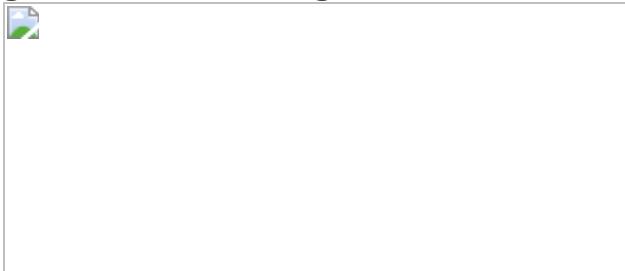


Fig. 4: Mechanical properties of 2DPA-1 nanofilms.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. 1.

Sakamoto, J., van Heijst, J., Lukin, O. & Schlüter, A. D. Two-dimensional polymers: just a dream of synthetic chemists? *Angew. Chem. Int. Ed.* **48**, 1030–1069 (2009).

2. 2.

Schlüter, A. D., Payamyar, P. & Öttinger, H. C. How the world changes by going from one- to two-dimensional polymers in solution. *Macromol. Rapid Commun.* **37**, 1638–1650 (2016).

3. 3.

Lee, C., Wei, X., Kysar, J. W. & Hone, J. Measurement of the elastic properties and intrinsic strength of monolayer graphene. *Science* **321**, 385–388 (2008).

4. 4.

Pop, E., Varshney, V. & Roy, A. K. Thermal properties of graphene: fundamentals and applications. *MRS Bull.* **37**, 1273–1281 (2012).

5. 5.

Stupp, S. I., Son, S., Lin, H. C. & Li, L. S. Synthesis of two-dimensional polymers. *Science* **259**, 59–63 (1993).

6. 6.

Grill, L. et al. Nano-architectures by covalent assembly of molecular building blocks. *Nat. Nanotechnol.* **2**, 687–691 (2007).

7. 7.

Kissel, P. et al. A two-dimensional polymer prepared by organic synthesis. *Nat. Chem.* **4**, 287–291 (2012).

8. 8.

Kissel, P., Murray, D. J., Wulf Lange, W. J., Catalano, V. J. & King, B. T. A nanoporous two-dimensional polymer by single-crystal-to-single-crystal photopolymerization. *Nat. Chem.* **6**, 774–778 (2014).

9. 9.

Kory, M. J. et al. Gram-scale synthesis of two-dimensional polymer crystals and their structure analysis by X-ray diffraction. *Nat. Chem.* **6**, 779–784 (2014).

10. 10.

Kandambeth, S., Dey, K. & Banerjee, R. Covalent organic frameworks: chemistry beyond the structure. *J. Am. Chem. Soc.* **141**, 1807–1822 (2019).

11. 11.

Diercks, C. S. & Yaghi, O. M. The atom, the molecule, and the covalent organic framework. *Science* **355**, eaal1585 (2017).

12. 12.

Baek, K. et al. Free-standing, single-monomer-thick two-dimensional polymers through covalent self-assembly in solution. *J. Am. Chem. Soc.* **135**, 6523–6528 (2013).

13. 13.

Gee, G. & Rideal, E. K. Reaction in monolayers of drying oils I—the oxidation of the maleic anhydride compound of β -elaeostearin. *Proc. R. Soc. Lond. A* **153**, 116–128 (1935).

14. 14.

Ozaki, H. et al. Formation of atomic cloth observed by Penning ionization electron spectroscopy. *J. Chem. Phys.* **103**, 1226–1228 (1995).

15. 15.

Zhong, Y. et al. Wafer-scale synthesis of monolayer two-dimensional porphyrin polymers for hybrid superlattices. *Science* **366**, 1379–1384 (2019).

16. 16.

Côté, A. P. et al. Porous, crystalline, covalent organic frameworks. *Science* **310**, 1166–1170 (2005).

17. 17.

Berlanga, I. et al. Delamination of layered covalent organic frameworks. *Small* **7**, 1207–1211 (2011).

18. 18.

Zhang, B. et al. Crystalline dioxin-linked covalent organic frameworks from irreversible reactions. *J. Am. Chem. Soc.* **140**, 12715–12719 (2018).

19. 19.

Guan, X. et al. Chemically stable polyarylether-based covalent organic frameworks. *Nat. Chem.* **11**, 587–594 (2019).

20. 20.

Cai, Z., Liu, B., Zou, X. & Cheng, H. M. Chemical vapor deposition growth and applications of two-dimensional materials and their heterostructures. *Chem. Rev.* **118**, 6091–6133 (2018).

21. 21.

Zhang, G., Zeng, Y., Gordiichuk, P. & Strano, M. S. Chemical kinetic mechanisms and scaling of two-dimensional polymers via irreversible solution-phase reactions. *J. Chem. Phys.* <https://doi.org/10.1063/5.0044050> (2021).

22. 22.

Payamyar, P., King, B. T., Öttinger, H. C. & Schlüter, A. D. Two-dimensional polymers: concepts and perspectives. *Chem. Commun.* **52**, 18–34 (2016).

23. 23.

Varoon, K. et al. Dispersible exfoliated zeolite nanosheets and their application as a selective membrane. *Science* **334**, 72–75 (2011).

24. 24.

Yeh, T.-M., Wang, Z., Mahajan, D., Hsiao, B. S. & Chu, B. High flux ethanol dehydration using nanofibrous membranes containing graphene oxide barrier layers. *J. Mater. Chem. A* **1**, 12998–13003 (2013).

25. 25.

Li, P. et al. In situ nanomechanical characterization of multi-layer MoS₂ membranes: from intraplanar to interplanar fracture. *Nanoscale* **9**, 9119–9128 (2017).

26. 26.

Bunch, J. S. et al. Impermeable atomic membranes from graphene sheets. *Nano Lett.* **8**, 2458–2462 (2008).

27. 27.

Sun, P. Z. et al. Limits on gas impermeability of graphene. *Nature* **579**, 229–232 (2020).

28. 28.

Leterrier, Y. Durability of nanosized oxygen-barrier coatings on polymers. *Prog. Mater Sci.* **48**, 1–55 (2003).

29. 29.

Fang, Q. et al. Strong and flaw-insensitive two-dimensional covalent organic frameworks. *Matter* **4**, 1017–1028 (2021).

30. 30.

Sandoz-Rosado, E., Beaudet, T. D., Andzelm, J. W. & Wetzel, E. D. High strength films from oriented, hydrogen-bonded “graphamid” 2D polymer molecular ensembles. *Sci. Rep.* **8**, 3708 (2018).

31. 31.

Liu, P. et al. Layered and scrolled nanocomposites with aligned semi-infinite graphene inclusions at the platelet limit. *Science* **353**, 364–367 (2016).

Acknowledgements

Membraackackne fabrication, permeability testing and transport analysis of 2DPA-1 was supported by the Center for Enhanced Nanofluidic Transport (CENT), an Energy Frontier Research Center sponsored by the US Department of Energy (DOE), Office of Science, Basic Energy Sciences under award #DE-SC0019112. The synthetic chemistry and mechanical testing aspects of this work were funded by the Army Research Laboratory under cooperative agreement W911NF-18-2-0055. H.J.K. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund, which supported the molecular modelling aspects of this work. We acknowledge fabrication support from the Center for Nanoscale Systems at Harvard, a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), which is supported by the National Science Foundation under NSF award no. 1541959. This research used beamline 11-

BM Complex Materials Scattering (CMS) of the National Synchrotron Light Source II (NSLS-II) and the Center for Functional Nanomaterials (CFN), both of which are US DOE Office of Science User Facilities operated for the DOE Office of Science by Brookhaven National Laboratory under contract no. DE-SC0012704. We thank E. Tsai for her assistance in performing experiments at the beamline, R. Verduzco for beamline access, A. Penn and E. Brignole for MIT.Nano assistance for STEM and Cryo EM, and S. Xin Li for discussions on 2D membrane properties. We acknowledge MIT.Nano facilities and the Cypher VRS DURIP award (N000142012203) for the support on AFM characterizations. M.K. acknowledges support by the German Research Foundation (DFG) Research Fellowship KU 3952/1-1.

Author information

Affiliations

1. Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Yuchen Zeng, Pavlo Gordiichuk, Takeo Ichihara, Ge Zhang, Jing Yang, Daichi Kozawa, Zhongyue Yang, Matthias Kuehne, Michelle Quien, Zhe Yuan, Xun Gong, Guangwei He, Daniel James Lundberg, Pingwei Liu, Albert Tianxiang Liu, Jing Fan Yang, Heather J. Kulik & Michael S. Strano

2. U.S. Army Combat Capabilities Development Command, Army Research Laboratory, Aberdeen Proving Ground, MD, USA

Emil Sandoz-Rosado & Eric D. Wetzel

3. Center for Nanoscale Systems, Harvard University, Cambridge, MA, USA

Jason Tresback

Contributions

Y.Z. and M.S.S. conceived and designed the reaction system, with initial laboratory synthesis and characterization of 2DPA-1 performed by Y.Z. Characterization using HR-AFM measurement and other imaging tools, and subsequent data analysis, was performed by Y.Z., P.G., X.G. and J.T. P.L., A.T.L., and J.F.Y. offered substrates for AFM characterization. Post-synthesis modification for molecular imaging and intermediate characterization was performed by Y.Z.; M.Q. and D.J.L. contributed to data analysis. Y.Z. and M.S.S. designed and developed the NMR characterization technique for defect quantification. Y.Z., M.Q. and G.Z. performed budge tests. Z. Yuan, G.Z., G.H. and M.Q. calculated permeabilities for 2DPA-1 from bulge test results. Y.Z. performed nanoindentation tests at MIT with complementary measurements from E.D.W. and E.S.-R. at ARL. Substrates for nanoindentation tests are designed and fabricated by M.K. Y.Z. designed and T.I. conducted polarized photoluminescence measurements. Y.Z. measured mechanical properties of scrolled fibres. G.Z., D.K., J. Yang, M.S.S. and H.J.K. contributed theory and material simulation. Z. Yang and H.J.K. performed molecular structure calculations for 2DPA-1. Y.Z. and M.S.S. co-wrote the manuscript with input from all other authors.

Corresponding author

Correspondence to [Michael S. Strano](#).

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 Schematic illustration of rotation suppression and auto-catalysis.

a, Linkage-core conjugations inhibit out-of-plane rotation. **b**, Auto-catalytic self-templating. Computational method: gas-phase geometry optimizations with Q-Chem v4.2 to compute 298 K conjugation enthalpies using the ωB97-XD/6-311 + G(d, p) DFT functional and basis set combination.

Extended Data Fig. 2 Characterization of 2DPA-1.

a, Fourier-transform infrared (FTIR) spectrum of as-synthesized 2DPA-1 powder. **b**, Atomic force microscopy (AFM) image of bilayer nanoclusters and its height histogram along the white line (inset). **c**, AFM image of stacked nanosheets; inset shows its height histogram along the white line. **d**, AFM surface topology of a spin-coated film. **e**, AFM image from amplitude channel at its second eigenmode, obtained from a film surface. **f**, The size distribution obtained from **e**.

Extended Data Fig. 3 Silylation of 2DPA-1.

a, Synthetic scheme of the silylation reaction. **b**, FTIR spectra of reaction mixture and its starting material. rt, room temperature; TMSOTf, trimethylsilyl trifluoromethanesulfonate; TEA, triethylamine.

Extended Data Fig. 4 Bulge test of 2DPA-1 films for air permeability.

a, Cross-sectional view of a clean holey substrate. **b**, Bubble height versus time. Film thickness is 12.8 nm.

Extended Data Fig. 5 Scrolled fibre tensile test of 2DPA-1 composites.

a, Schematic illustration of an Archimedean scroll fibre. **b**, Optical micrograph of a hair (left) and a scrolled fibre (right). Scale bar, 100 mm. **c**, Representative true stress–strain curves from a 2D composite scrolled fibre, its polycarbonate (PC) control fibre, and a graphene/PC composite fibre (data reproduced from ref. ³¹). The volume fraction for 2DPA-1/PC is 6.9% and for graphene/PC it is 0.19%. **d**, Plot of modulus enhancement ($(E - E_{PC})/E_{PC}$) versus different volume fractions of 2DPA-1.

Extended Data Fig. 6 True stress–strain plots of composite scrolled fibres and their PC controls.

a, Volume fraction (V_{2DP}) = 0.9%. **b**, V_{2DP} = 2.3%. **c**, V_{2DP} = 6.9%. **d**, V_{2DP} = 7.7%. **e**, V_{2DP} = 13.3%.

Extended Data Fig. 7 Optical set-up for photoluminescence measurements.

Excitation wavelength 532 nm, excitation power 500 μ W for photoluminescence measurements, and 2 μ W for excitation polarization. EMCCD, electron-multiplying charge-coupled device.

Supplementary information

Supplementary Information

This file contains Supplementary Information, including Supplementary Figures 1–45, Supplementary Tables 1–3, and additional references.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Zeng, Y., Gordiichuk, P., Ichihara, T. *et al.* Irreversible synthesis of an ultrastrong two-dimensional polymeric material. *Nature* **602**, 91–95 (2022). <https://doi.org/10.1038/s41586-021-04296-3>

- Received: 13 January 2021
- Accepted: 01 December 2021
- Published: 02 February 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04296-3>

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

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

- Article
- Open Access
- [Published: 19 January 2022](#)

Novel flight style and light wings boost flight performance of tiny beetles

- [Sergey E. Farisenkov](#) [ORCID: orcid.org/0000-0003-2634-2204](#)^{1,na1},
- [Dmitry Kolomenskiy](#) [ORCID: orcid.org/0000-0003-0107-6894](#)^{2,3 na1},
- [Pyotr N. Petrov](#)¹,
- [Thomas Engels](#)⁴,
- [Nadezhda A. Lapina](#)¹,
- [Fritz-Olaf Lehmann](#) [ORCID: orcid.org/0000-0002-9511-3651](#)⁴,
- [Ryo Onishi](#) [ORCID: orcid.org/0000-0001-9250-0712](#)²,
- [Hao Liu](#)⁵ &
- [Alexey A. Polilov](#) [ORCID: orcid.org/0000-0002-6214-3627](#)^{1,6}

Nature volume 602, pages 96–100 (2022)

- 28k Accesses
- 1475 Altmetric
- [Metrics details](#)

Subjects

- [Biological physics](#)
- [Biophysics](#)
- [Entomology](#)
- [Evolution](#)
- [Mechanical engineering](#)

Abstract

Flight speed is positively correlated with body size in animals¹. However, miniature featherwing beetles can fly at speeds and accelerations of insects three times their

size². Here we show that this performance results from a reduced wing mass and a previously unknown type of wing-motion cycle. Our experiment combines three-dimensional reconstructions of morphology and kinematics in one of the smallest insects, the beetle *Paratuposa placentis* (body length 395 µm). The flapping bristled wings follow a pronounced figure-of-eight loop that consists of subperpendicular up and down strokes followed by claps at stroke reversals above and below the body. The elytra act as inertial brakes that prevent excessive body oscillation. Computational analyses suggest functional decomposition of the wingbeat cycle into two power half strokes, which produce a large upward force, and two down-dragging recovery half strokes. In contrast to heavier membranous wings, the motion of bristled wings of the same size requires little inertial power. Muscle mechanical power requirements thus remain positive throughout the wingbeat cycle, making elastic energy storage obsolete. These adaptations help to explain how extremely small insects have preserved good aerial performance during miniaturization, one of the factors of their evolutionary success.

[Download PDF](#)

Main

Driven by curiosity about the smallest objects, scientific exploration of the microscopic world has facilitated the miniaturization of various industrial products. But miniaturization is not just a human-made artifice: success stories of miniaturization are abundant in the living world. For more than 300 million years, ecological pressures have forced insects to develop extremely small bodies down to 200 µm long³ without losing their ability to fly. As the physical properties of flight depend on size, constraints that are insignificant at the macro scale become significant at the micro scale, and vice versa⁴. Compared with larger sizes, flight at small sizes is dominated by viscous air friction rather than inertial forces resulting from the acceleration of the surrounding air. This competition between friction and inertia is key for flight at all size scales and thus applies to all animals that move through air.

Large insects generally fly faster than smaller ones¹. Nevertheless, some of the smallest insects fly surprisingly well. For example, it was recently revealed that minute featherwing beetles (Coleoptera: Staphylinoidea: Ptiliidae) typically fly with similar speeds to their larger relatives (Staphylinidae), despite a threefold difference in body length². Moreover, ptiliids can accelerate twice as fast as carrion beetles (Staphylinoidea: Silphidae), although the latter are an order of magnitude larger. As the size-specific flight-muscle volume is smaller in Ptiliidae than in larger beetles⁵, their excellent flight performance must result from the peculiar structure of their wings and flight style. Ptiliids have feather-like bristled wings—a condition known as

ptiloptery (Fig. 1b)—instead of the membranous wings possessed by most insects. This visually striking modification of the flight apparatus evolved convergently in extremely small representatives of several insect orders. The functional benefits of ptiloptery, however, have remained largely unknown.

Fig. 1: External morphology of *P. placentis*.

 **figure 1**

a–c, Scanning electron microscopy images showing relative size of *P. placentis* (left) and *A. proteus* (right) (**a**), wing of *P. placentis* (**b**) and part of a seta (**c**). Every image was obtained from one randomly selected specimen; for detailed morphometry, see Supplementary Information.

Although many studies have focused on the secrets of flight in minute insects^{6,7}, most experimental data that elucidate wing motion and aerodynamics have been obtained from larger insect species^{8,9,10,11}. Thus, unsteady aerodynamics of millimetre-size insects such as fruit flies^{12,13} and mosquitoes¹⁴ have received considerable attention in recent decades, whereas studies focusing on tiny insects remained scarce. Two-dimensional numerical studies on the aerodynamics of insect wings have previously shown that the flow past evenly spaced cylinder lattices reduces aerodynamic force production in bristled wings^{15,16}. By contrast, experiments with mechanical comb-like models have suggested slightly larger lift-to-drag ratios during the clap-and-fling phase in bristled wings compared with membranous wings^{17,18,19}, but did not cover the full wingbeat cycle. Meanwhile, using state-of-the-art high-speed videography, it has become clear that small insects use a wingbeat cycle that is different from that of the larger ones^{10,11}, but, to our knowledge, the role of ptiloptery in this cycle has not been considered.

In this study, we analysed the flight of the miniature featherwing beetle *Paratuposa placensis*. We constructed a morphological model based on data gained from light, confocal and electron microscopy measurements, a kinematical model using synchronized high-speed videography, and a dynamical model using computational methods of solid and fluid mechanics. The combination of these methods offers a comprehensive view of how bristled wings work and explains why common sub-millimetre flying insects have bristled rather than membranous wings.

Structural features of *P. placensis*

P. placensis is one of the smallest non-parasitic insect species, with a body length of about $395 \pm 21 \mu\text{m}$ (all measurements are given as mean \pm s.d.). This size is similar to the size of some unicellular protists such as *Amoeba proteus* (Fig. 1a). The body mass of *P. placensis* is $2.43 \pm 0.19 \mu\text{g}$ (Supplementary Information). The bristled wing consists of a petiole, a narrow wing blade and a fringe of setae (bristles) covered with secondary outgrowths (Fig. 1b,c). The wing length is $493 \pm 18 \mu\text{m}$ and the setae occupy $95.1 \pm 0.3\%$ of the aerodynamically effective wing area (interior of the green contour in Fig. 2b).

Fig. 2: Kinematics of *P. placensis*.

 **figure 2**

a, Frame sequence of a single stroke in two projections. **b**, Measurement scheme for Euler angles. **c**, Trajectory of the wing tip: recovery strokes (magenta line) and power strokes (green line) and measurement scheme for angle of body pitch (χ) and pitch of stroke plane to the horizon (β). **d**, Wing Euler angles as functions of dimensionless time t/T , $T = 1/f$: stroke deviation (θ), positional (φ) and pitch (ψ). **e**, Elytron Euler angles (θ , φ and ψ); body pitch angle (χ). **f**, AoA and wing speed at radius of gyration (R_g) versus t/T .

Wing kinematics

The wingbeat cycle of *P. placentis* consists of two power strokes, during which most of the total flight force is generated²⁰, and two recovery strokes with wings clapping above and below the body (Fig. [2a, c](#), Supplementary Videos [1–6](#)). Dorsal and ventral

recovery strokes are unique to the Ptiliidae and replace the conventional clap-and-fling kinematics described in other insects, including miniature thrips⁹ and parasitoid wasps⁸. Despite the large stroke amplitude, the wings do not always clap tightly at the end of the ventral recovery stroke, depending on flight conditions (Supplementary Information). The setal fringes of the left and right wings may intersect during the fling phases of the recovery strokes. The morphological downstroke and upstroke are remarkably similar: the angle of attack (AoA) reaches 73° during the downstroke and 85° in magnitude during the upstroke (Fig. 2f). The cycle-averaged Reynolds number (Re) based on the mean speed of the radius of gyration is 9 and reaches 20 during power strokes when wing velocity is highest. The increased AoA during power strokes and the presence of recovery strokes are similar to the kinematics of swimming in miniature aquatic crustaceans (Supplementary Information), which move at similar flow regimes—for example, larvae of *Artemia* sp.²⁰, with a Reynolds number of 10.

Vertical force generation

The wide rounded self-intersecting paths of the wing tips and dynamically changing orientation of the wings (Figs. 2c, 3a) maximize the aerodynamic asymmetry between power and recovery strokes. Upon each power stroke, geometrical AoA and wing velocity simultaneously reach their maxima (Fig. 2f). While forces and velocities are anti-aligned (Fig. 3a), their peaks are synchronized (Figs. 3d, 2f). The wing thus first produces an increased upward force as it quickly moves flat-on with net downward displacement and, subsequently, a small downward force while slowly moving edge-on upwards. The near-clap motion reduces the parasite downward force upon recovery²¹. Decomposition of the vertical force exerted on the wing into drag and lift (Methods) is shown in Fig. 3b, d. The vertical force due to drag exhibits greater positive peaks than that due to lift. This is accompanied by extended times of slightly negative drag-based vertical force. In association with these peaks, airflow simulation reveals a pair of strong vortex rings that are typical for drag-producing bodies (Fig. 3c, Supplementary Videos 5, 6). Approximately 32% of the cycle-averaged vertical force results from drag and 68% results from lift, indicating that the beetle benefits from both components. On average, the aerodynamic mechanisms produce bodyweight-supporting lift of 2.7 μg (Fig. 3g) (the beetle's estimated body mass is 2.4 μg) and a vertical acceleration of 1.0 m s^{-2} . The net contribution of body and elytra to the vertical force is negligible (Fig. 3g).

Fig. 3: Aerodynamic forces acting on the wings of *P. placensis*.

 figure 3

a, Wing tip trajectories and direction of total vertical force: downward force (recovery stroke) is shown in green, upward force (power stroke) is shown in red. Posture at t/T of 0.6 is shown in red, and posture at t/T of 0.82 is in green. Cyan arrows show aerodynamic force; magenta arrows show wing-tip velocity; yellow discs and arrows

show dorsal surface orientation of the wing at nine labelled time instants. Opaque and transparent lines and arrows correspond to right and left wing, respectively. **b**, Vector scheme of forces acting on wing. **c**, Airflow simulation visualized using iso-surfaces of vorticity magnitude (see also Supplementary Video 5). **d**, Vertical aerodynamic force (v.f.) exerted on one wing versus time. Yellow highlighted zones denote the time span of power strokes. Tot. mem., vertical force of membranous wing model. **e**, Body mass-specific aerodynamic (aero.) and inertial (inert.) power, and their sum as the total power. **f**, Pitching torque about centre of mass. The positive direction is nose down. **g**, Contribution of different parts to total aerodynamic force acting on the beetle in the vertical direction, averaged over the wingbeat cycle. **h**, Mean and peak body mass-specific aerodynamic power in computations for bristled and membranous (mem.) wings.

Stabilizing role of the elytra

The unusually large horizontal and vertical excursion of the wings during flapping poses a peculiar flight dynamics problem. The forces are small during the recovery strokes but the moment arm relative to the centre of mass is large. This results in a pitching moment large enough to overturn the body around its pitching axis (Extended Data Fig. 7, Supplementary Information). To compensate for these moments in synchrony with wing flapping, the insect opens and closes the elytra with large amplitudes ($\psi_{\max} - \psi_{\min} = 52^\circ$) compared to other flying beetles^{22,23,24}. Figures 2e, 3f show that the elytra act as an inertial brake. At times between t/T of 0 and 0.3, the wings are raised in a dorsal position and produce nose-up torque. As soon as the wings start their downstroke, the elytra close, causing a nose-down recoil torque on the body. During ventral clapping, the wings produce nose-down torque and the elytra decelerate and reopen. We found that the elytra movements decrease the amplitude of body-pitching oscillation by approximately 50% compared with flight without elytra (Supplementary Information). It is thus likely that the inertial brake observed in *P. placensis* is a feature of ptiliid beetles flying at high wingbeat amplitudes and unique to their peculiar flying style.

Bristled versus membranous wings

Numerical modelling suggests a wing mass in *P. placensis* of approximately 0.024 µg, which is about 1% of the body mass. By contrast, estimates of the mass of a membranous wing with the same outline amount to 0.13, 0.14 or 0.19 µg, depending on wing thickness (Supplementary Information). These estimates were based on some of the smallest membranous-winged insects, namely the wasp *Trichogramma telengai* (0.73 µm), the beetle *Orthoperus atomus* (0.85 µm), and the beetle *Limnebius atomus* (1.12 µm). *L. atomus* is closely related to the Ptiliidae. The maximum entry in the

inertia matrix (I_{zz}) of the bristled wing is $1,600 \mu\text{g } \mu\text{m}^2$, and for the membranous wings it is $13,800$, $16,000$ and $20,800 \mu\text{g } \mu\text{m}^2$, respectively. Secondary outgrowths of the bristles are unique to Ptiliidae wings and reduce the wing mass by 44% compared with the bristled wing model with smooth cylindrical bristles at the same drag²⁵. The bristled wing architecture with secondary outgrowths thus considerably reduces wing mass compared to a membranous architecture, while maintaining the needed aerodynamic properties. This conclusion is also supported by an allometric analysis of wing mass in differently sized insects (Extended Data Fig. 1, Supplementary Information).

Whereas instantaneous vertical forces generated by the simulated membranous wing outscore the bristled wing of *P. placensis* (Fig. 3d), the latter produces as much as 68% of the mean vertical force of the membranous wing. The vertical force peaks during power strokes produce similar-sized peaks in the mechanical power required for wing actuation. Cycle-averaged power consumption in *P. placensis* is relatively low and amounts to only $28 \text{ W per kg body mass}$, but instantaneous power may reach up to 110 W kg^{-1} at $t/T = 0.82$ (Fig. 3e,h) owing to aerodynamic power. The total mechanical power of the bristled wing model (Fig. 3e) remains positive during the entire wingbeat cycle, because low inertia of the wing and high viscous damping of the surrounding air enable continuous energy transfer from the flight apparatus to the wake. No elastic energy storage is required. By contrast, for a membranous wing, the inertial power is similar in peak magnitude to the aerodynamic power. Such a wing requires perfect elastic energy storage to achieve its minimum mean mechanical power of 37 W kg^{-1} and powerful flight muscles to satisfy $180\text{--}210 \text{ W kg}^{-1}$ peak power requirements (Fig. 3h). The latter computational estimates include aerodynamic added mass effects during wing motion and are detailed in the Supplementary Information.

At low Reynolds numbers, impermeable membranous wings barely outperform leaky bristled wings in generating aerodynamic force. Thus, the small advantage of using a membranous wing is outweighed by the advantage gained in reducing inertial torques and power by minimizing wing mass. This trade-off of energy savings for a small penalty in aerodynamic force generation is available only at Reynolds numbers of about 10 or lower, where sufficiently low leakiness can be achieved with a small number of slender bristles.

Conclusions and outlook

The findings reported here expand our understanding of the flight mechanics at low Reynolds numbers. In flight, small insects need to produce forces to support their body weight in conditions of high viscous drag on the body and wings. *P. placensis* uses kinematic strategies that maximize wing flapping amplitude but at the potential cost of

an increase in inertial power requirements. This is resolved by ptiloptery, an effective structural architecture that serves to reduce inertial costs of wing flapping, making elastic energy storage obsolete and reducing peak mechanical power requirements of the flight muscles. The wingbeat cycle of *P. placentis* is highly functionally divided into power and slow-recovery strokes. The wings thereby produce pronounced high torques that cause the high-amplitude body pitch oscillation. Inertial braking provided by moving elytra represents an ingenious solution to this problem, enhancing posture stability without providing additional forces for flight. In *P. placentis*, these mechanisms improve the temporal distribution of muscle mechanical power requirements and help to maintain aerial performance at an extremely small body size. If this flight style is common for miniature beetles, it may largely explain their worldwide abundance. Further studies of other microinsects with bristled wings will help to reveal the causes of the convergent evolution of ptiloptery during miniaturization in many groups of insects.

Methods

Data reporting

No statistical methods were used to predetermine sample sizes. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Material

Adult featherwing beetles (*P. placentis* (Coleoptera: Ptiliidae)) were collected in Cát Tiên National Park, Vietnam, in November 2017. The beetles were collected and delivered to the laboratory together with the substrate for their safety. High-speed video recordings were made on the same day during a few hours after collecting.

Morphology and morphometry

The material for morphological studies was fixed in alcoholic Bouin solution or in 70% ethanol. Wing structure was studied using a scanning electron microscope (SEM Jeol JSM-6380 and FEI Inspect F50), after dehydration of the samples and critical point drying, followed by gold sputtering. A confocal microscope (CLSM Olympus FV10i-O) and a transmitted light microscope (Olympus BX43) were also used, for which the samples were clarified and microscopic slides were made²⁶ (Supplementary Information). Measurements were taken from digital photographs in Autodesk AutoCAD software in ten replications (unless otherwise noted). Body weights and

weights of particular body parts were calculated on the basis of three-dimensional reconstructions (Supplementary Information).

Wing mass and moments of inertia

The volumes of the petiole and membranous part (the blade) of the wing were measured using CLSM image-based geometrical models. Uniform cuticle density $1,200 \text{ kg m}^{-3}$ was assumed²⁷. The wing mass was obtained by summing up the contributions from the petiole, blade and setae. To calculate the mass of the setae, we first estimated their linear density ($0.96 \mu\text{g m}^{-1}$) using a three-dimensional model²⁵ and multiplied it by the length. The petiole and the blade of the wing model have constant thickness without veins. A possible range of the membrane thickness was hypothesized on the basis of measurements in *T. telengai* (Hymenoptera: Trichogrammatidae, body length 0.45 mm), *O. atomus* (Coleoptera: Corylophidae, body length 0.8 mm) and *L. atomus* (Coleoptera: Hydraenidae, body length 1.1 mm), on $0.5 \mu\text{m}$ thick histological sections obtained by diamond knife cutting using a Leica microtome, after fixation and embedding in araldite. These values are the minimal thicknesses measured in each species. The measurements were performed using an Olympus BX43 microscope. The measurement error of linear dimensions is of the order of magnitude of 1% in the spanwise and chordwise directions and 10% for the thickness. The s.d. of wing cuticle density²⁵ is approximately 100 kg m^{-3} . This suggests that the overall root sum square error of the wing mass calculation is of about 13%. To evaluate the moments of inertia, surface density of the membranous parts and linear density of the bristles were calculated. The moments of inertia of the individual setae were calculated using the formula for a thin rod at an angle and the parallel axis theorem. The moments of inertia of the membranous parts were calculated using a two-dimensional quadrature rule with the discretization step of $50 \mu\text{m}$.

High-speed recording

Flight of the beetles was recorded in closed $20 \times 20 \times 20 \text{ mm}$ chambers, custom made of 1.0 mm thick microscopic slides and 0.15 mm cover-glass at a natural level of illumination in visible light. There were 20–30 insects in the flight chamber during the recording. For temperature stabilization the flight chamber was chilled by an air fan from the outside. The ambient temperature measured by a digital thermocouple was 22–24 °C; the temperature of the flight chamber was 22–26 °C.

High-speed video recordings were made using two synchronized Evercam 4000 cameras (Evercam) with a frequency of 3,845 FPS and a shutter speed of $20 \mu\text{s}$ in infrared light (850 nm LED). The high-speed cameras were mounted on optical rails precisely orthogonal to each other and both situated at 0° from the horizon. Two IR LED lights were placed opposite to the cameras and one light above the flight

chamber. A graphical representation of the experimental setup can be found in the previous study².

Measurement of kinematics

For analysis, 13 recordings were selected. For four of them (PP2, PP4, PP5 and PP12) we reconstructed the kinematics of body parts in four kinematic cycles for each and performed CFD calculations because the flight of these specimens was especially similar to conventional hovering: relatively slow normal flight with horizontal velocity $0.057 \pm 0.014 \text{ m s}^{-1}$ (hereafter mean \pm s.d.) and $0.039 \pm 0.031 \text{ m s}^{-1}$ vertical velocity (PP2, PP4, PP5 and PP12). In CFD analysis with the membranous wing model, we selected kinematics of PP2, which does not cross the wings while clapping. This case is convenient for comparing the performance of bristled wings with substitute membranous wings, because it guarantees that the latter do not intersect. The perimeter of the membrane is formed by lines connecting the tips of the bristles (see the previous study²⁵ for more information). The descriptions of kinematics and aerodynamics, as well as the illustrations, refer to results obtained for individual PP2. For the results obtained for other specimens, see Supplementary Information and Extended Data Figs. [2](#), [4–6](#).

Average wingbeat frequency was calculated as the mean of the wingbeat frequency in all recordings. In each recording, the number of frames was counted in several complete kinematic cycles, 104 cycles in total.

For the mathematical description of the kinematics of the wings and elytra, we used the Euler angles system^{28,29} (Fig. [2b](#)) based on frame-by-frame reconstruction of the location of the insect's body parts (wings, elytra and body itself) performed in Autodesk 3Ds Max. Three-dimensional models of the body and elytra were obtained by confocal microscope image stacking, and the flat wing model was based on light microscopy photos of dissected wings. We used the rigid flat wing model for reconstruction of the kinematics because the deformations of the wings are minor (Supplementary Information). First, we prepared frame sequences with four full kinematical cycles in each. The frames were then centred and cropped by point between the bases of the wings and then placed as orthogonal projections. Virtual models of body parts were placed into a coordinate system with two image planes. Then we manually changed the position and rotated body parts until their orthogonal projections were superimposed on the image planes. For calculating the Euler angles, a coordinate system was created (Fig. [2a](#)). The XOY plane is a plane parallel to the stroke plane, and intersecting with the base of wing or elytron, which is positioned in the zero point. To determine the position of the stroke plane, we calculated the major axis trend line of the wingtip coordinates instead of the linear trend line²⁹, because the wingtip trajectory of *P. placensis* forms a wide scatter plot. Stroke deviation angle (θ) and

positional angle (φ) were calculated from the coordinates of the base and apex. Pitch angle (ψ) is the angle between the stroke plane and the chord perpendicular to the line between the base and apex. The body pitch angle (χ) is the angle between the stroke plane and longitudinal axis of the body, calculated as the line between the tip of the abdomen and the midpoint between the apical antennomeres. Pitch angle (β) of the stroke plane relative to the horizon was also measured.

For flight speed analysis we performed tracking of the centre of the body (middle point between the extreme edges of the head and abdomen) in Tracker (Open Source Physics) in both projections and calculated the instantaneous velocity and its vertical and horizontal components in each frame. The obtained speed values were filtered by loess fitting in R (stats package). The minimum distance between the wingblade tips during bottom claps was also calculated.

Computational fluid dynamics

Time intervals of low-speed flight with duration longer than four wing beats were selected. The angles φ , θ and ψ of the left wing, right wing and elytra and the body angle χ were interpolated on a uniform grid with time step size $\Delta t = 2.6 \times 10^{-6}$ s. By solving numerically $\varphi(t) = 0$ with respect to t , we identified four subsequent wingbeat cycles and calculated the average cycle period T and the wingbeat frequency $f = 1/T$. We then spline-interpolated the data for each of the four cycles on a grid subdividing the time interval $[0, T]$ with step Δt , calculated phase averages, then calculated the average between the left and right wing. This yielded the plots shown in Fig. 2c, d. Constant forward and upward/downward flight velocity was prescribed using the time average values of the loess-filtered time series.

The computational fluid dynamics analysis was performed using the open-source Navier–Stokes solver WABBIT³⁰, which is based on the artificial compressibility method to enforce velocity-pressure coupling, volume penalization method to model the no-slip condition at the solid surfaces, and dynamic grid adaptation using the wavelet coefficients as refinement indicators. The flying insect was represented as an assembly of five rigid solid moving parts: the two elytra and the two wings move relative to the body, and the body oscillates about its lateral axis (Supplementary Information). The kinematic protocol is described in Supplementary Information and Extended Data Fig. 2c. The computational domain is a $12R \times 12R \times 12R$ cube, where R is the wing length, with volume penalization used in combination with periodic external boundary conditions to enforce the desired far-field velocity³⁰. The computational domain was decomposed in nested Cartesian blocks, each containing $25 \times 25 \times 25$ grid points. The blocks were created, removed and redistributed among parallel computation processes so as to ensure maximum refinement level near the solid boundaries and constant wavelet coefficient thresholding otherwise during the

simulations. The numerical simulations started from the quiescent air condition, continued for a time period of two wingbeat cycles with a coarse spatial grid resolution of $\Delta x_{\min} = 0.00781R$ to let the flow develop to its ultimate periodic state, then the spatial discretization size was allowed to reduce to $\Delta x_{\min} = 0.00098R$ if the wing was bristled or to $\Delta x_{\min} = 0.00049R$ if it was membranous, and the simulation continued for one more wingbeat period to obtain high-resolution results. The air temperature was 25 °C in all cases; its density was $\rho = 1.197 \text{ kg m}^{-3}$ and its kinematic viscosity was $\nu = 1.54 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}$; the artificial speed of sound was prescribed as $c_0 = 30.38fR$, based on an earlier experimental validation²⁵. The volume penalization and other case-specific parameter values are provided in Supplementary Information. The CFD simulation accuracy is discussed in Supplementary Information and Extended Data Fig. 8.

Decomposition of the aerodynamic force of a wing into lift and drag components

The drag component of the total instantaneous aerodynamic force acting on the wing is defined as its projection on the direction of the wing velocity at the radius of gyration. The lift component is defined as a vector subtraction of the total force and the drag component. The total lift and drag force vectors are projected on the vertical (z) direction to obtain the time courses shown in Fig. 3d.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Extended data sets and raw data are available in an Open Science Framework repository (<https://osf.io/v3wrk/>).

Code availability

Computational fluid dynamics simulations were performed using an open-source code WABBIT, which can be downloaded from GitHub (<https://github.com/adaptive-cfd/WABBIT>) and has been described in detail elsewhere³⁰.

References

1. 1.

Tennekes, H. *The Simple Science of Flight: From Insects to Jumbo Jets* (MIT Press, 2009).

2. 2.

Farisenkov, S. E., Lapina, N. A., Petrov, P. N. & Polilov, A. A. Extraordinary flight performance of the smallest beetles. *Proc. Natl Acad. Sci. USA* **117**, 24643–24645 (2020).

3. 3.

Polilov, A. A. Small is beautiful: features of the smallest insects and limits to miniaturization. *Annu. Rev. Entomol.* **60**, 103–121 (2015).

4. 4.

Fukuda, T., Nakajima, M., Takeuchi, M. & Hasegawa, Y. in *Living Machines: A Handbook of Research in Biomimetic and Biohybrid Systems* (eds Prescott, T. J. et al.) (Oxford Univ. Press, 2018).

5. 5.

Yavorskaya, M. I., Beutel, R. G., Farisenkov, S. E. & Polilov, A. A. The locomotor apparatus of one of the smallest beetles—the thoracic skeletomuscular system of *Nephanes titan* (Coleoptera, Ptiliidae). *Arthropod Struct. Dev.* **48**, 71–82 (2019).

6. 6.

Horridge, G. A. The flight of very small insects. *Nature* **178**, 1334–1335 (1956).

7. 7.

Sane, S. P. Neurobiology and biomechanics of flight in miniature insects. *Curr. Opin. Neurobiol.* **41**, 158–166 (2016).

8. 8.

Weis-Fogh, T. Quick estimates of flight fitness in hovering animals, including novel mechanisms for lift production. *J. Exp. Biol.* **59**, 169–230 (1973).

9. 9.

Santhanakrishnan, A. et al. Clap and fling mechanism with interacting porous wings in tiny insect flight. *J. Exp. Biol.* **217**, 3898–3909 (2014).

10. 10.

Cheng, X. & Sun, M. Very small insects use novel wing flapping and drag principle to generate the weight-supporting vertical force. *J. Fluid Mech.* **855**, 646–670 (2018).

11. 11.

Lyu, Y. Z., Zhu, H. J. & Sun, M. Flapping-mode changes and aerodynamic mechanisms in miniature insects. *Phys. Rev. E* **99**, 012419 (2019).

12. 12.

Dickinson, M. H., Lehmann, F. O. & Sane, S. P. Wing rotation and the aerodynamic basis of insect right. *Science* **284**, 1954–1960 (1999).

13. 13.

Dickinson, M. H. & Muijres, F. T. The aerodynamics and control of free flight manoeuvres in *Drosophila*. *Philos. Trans. R. Soc. B* **371**, 20150388 (2016).

14. 14.

Bomphrey, R. J., Nakata, T., Phillips, N. & Walker, S. M. Smart wing rotation and trailing-edge vortices enable high frequency mosquito flight. *Nature* **544**, 92–95 (2017).

15. 15.

Lee, S. H. et al. Aerodynamic characteristics of unsteady gap flow in a bristled wing. *Phys. Fluids* **30**, 071901 (2018).

16. 16.

Lee, S. H., Lee, M. & Kim, D. Optimal configuration of a two-dimensional bristled wing. *J. Fluid Mech.* **888**, A23 (2020).

17. 17.

Kasoju, V. T., Terrill, C. L., Ford, M. P. & Santhanakrishnan, A. Leaky flow through simplified physical models of bristled wings of tiny insects during clap

and fling. *Fluids* **3**, 44 (2018).

18. 18.

Kasoju, V. T. & Santhanakrishnan, A. Aerodynamic interaction of bristled wing pairs in fling. *Phys. Fluids* **33**, 031901 (2021).

19. 19.

Ford, M. P., Kasoju, V. T., Gaddam, M. G. & Santhanakrishnan, A. Aerodynamic effects of varying solid surface area of bristled wings performing clap and fling. *Bioinspir. Biomim.* **14**, 046003 (2019).

20. 20.

Walker, J. A. Functional morphology and virtual models: physical constraints on the design of oscillating wings, fins, legs, and feet at intermediate Reynolds numbers. *Integr. Comp. Biol.* **42**, 232–242 (2002).

21. 21.

Cheng, X. & Sun, M. Revisiting the clap-and-fling mechanism in small wasp *Encarsia formosa* using quantitative measurements of the wing motion. *Phys. Fluids* **31**, (2019).

22. 22.

Sitorus, P. E., Park, H. C., Byun, D., Goo, N. S. & Han, C. H. The role of elytra in beetle flight: I. generation of quasi-static aerodynamic forces. *J. Bionic Eng.* **7**, 354–363 (2010).

23. 23.

Oh, S., Lee, B., Park, H., Choi, H. & Kim, S. T. A numerical and theoretical study of the aerodynamic performance of a hovering rhinoceros beetle (*Trypoxylus dichotomus*). *J. Fluid Mech.* **885**, A18 (2019).

24. 24.

Atkins, D. M. observations on the flight, wing movements and wing structure of male *Priacma serrata* (Lee.) (Coleoptera: Cupedidae). *Can. Entomol.* **90**, 339–347 (1958).

25. 25.

Kolomenskiy, D. et al. Aerodynamic performance of a bristled wing of a very small insect: dynamically scaled model experiments and computational fluid dynamics simulations using a revolving wing model. *Exp. Fluids* **61**, 194 (2020).

26. 26.

Polilov, A. A., Reshetnikova, N. I., Petrov, P. N. & Farisenkov, S. E. Wing morphology in featherwing beetles (Coleoptera: Ptiliidae): features associated with miniaturization and functional scaling analysis. *Arthropod Struct. Dev.* **48**, 56–70 (2019).

27. 27.

Vincent, J. F. V. & Wegst, U. G. K. Design and mechanical properties of insect cuticle. *Arthropod Struct. Dev.* **33**, 187–199 (2004).

28. 28.

Ellington, C. P. The aerodynamics of hovering insect flight. III. Kinematics. *Philos. Trans. R. Soc. B* **305**, 41–78 (1984).

29. 29.

Cheng, X. & Sun, M. Wing-kinematics measurement and aerodynamics in a small insect in hovering flight. *Sci. Rep.* **6**, 25706 (2016).

30. 30.

Engels, T., Schneider, K., Reiss, J. & Farge, M. A wavelet-adaptive method for multiscale simulation of turbulent flows in flying insects. *Commun. Comput. Phys.* **30**, 1118–1149 (2021).

Acknowledgements

A.A.P. thanks R. D. Zhantiev for mentorship and for inspiring him to study miniature insects. We thank A. K. Tsaturyan for helpful discussions. The work of S.E.F., A.A.P., P.N.P. and N.A.L. was supported by the Russian Science Foundation (project number 19-14-00045, study of morphology, high-speed recording and reconstruction of kinematics). This study was performed using equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University (A.A.P. and D.K., project no. 2183 ‘Computational fluid dynamics of the smallest insects’), TSUBAME3.0 supercomputer at the Tokyo Institute of Technology (D.K.) and HPC resources of IDRIS (T.E., allocation number A0102A01664 attributed by the

Grand Équipement National de Calcul Intensif (GENCI)). The work of D.K. was supported by the JSPS KAKENHI (grant number 18K13693). H.L. was partially supported by the JSPS KAKENHI (grant number 19H02060). The contributions of F.-O.L. and T.E. were supported by grants from the Deutsche Forschungsgemeinschaft to F.-O.L. (LE905/16-1 and LE905/18-1). SEM studies were performed using the Shared Research Facility Electron microscopy in life sciences at Lomonosov Moscow State University (unique equipment ‘Three-dimensional electron microscopy and spectroscopy’).

Author information

Author notes

1. These authors contributed equally: Sergey E. Farisenkov, Dmitry Kolomenskiy

Affiliations

1. Department of Entomology, Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia
Sergey E. Farisenkov, Pyotr N. Petrov, Nadezhda A. Lapina & Alexey A. Polilov
2. Global Scientific Information and Computing Center, Tokyo Institute of Technology, Tokyo, Japan
Dmitry Kolomenskiy & Ryo Onishi
3. Skoltech Center for Design, Manufacturing and Materials, Skolkovo Institute of Science and Technology, Moscow, Russia
Dmitry Kolomenskiy
4. Department of Animal Physiology, Institute of Biological Sciences, University of Rostock, Rostock, Germany
Thomas Engels & Fritz-Olaf Lehmann
5. Graduate School of Engineering, Chiba University, Chiba, Japan
Hao Liu
6. Joint Russian-Vietnamese Tropical Research and Technological Center, Southern Branch, Ho Chi Minh City, Vietnam

Alexey A. Polilov

Contributions

A.A.P. conceptualized and designed this study; S.E.F., N.A.L and A.A.P. designed the experiment and collected the data; H.L., F.-O.L and R.O. conceptualized the computational analysis; T.E. and D.K. performed the CFD simulations; S.E.F. and D.K. analysed the data; S.E.F., P.N.P., D.K. and A.A.P. wrote the manuscript. All authors edited the manuscript and approved the final version.

Corresponding authors

Correspondence to [Sergey E. Farisenkov](#) or [Alexey A. Polilov](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Michael Dickinson, Adrian Thomas 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 Relationship between wing mass \$m_w\$ and wing area \$S\$.](#)

Filled circles show data points for living mass of membranous insect wings from published sources (see Supplementary Information). Punctured circles show dried mass of membranous wings of Staphylinoidea beetles obtained as described in

Supplementary Information. The cross indicates calculated weight of the bristled wing of the featherwing beetle *Paratuposa placentis*. Plus signs delimit the estimated mass range of equivalent membranous wings. Blue dashed line is an allometric trend based on all data for membranous wings ($m_w = 2.17 \cdot S^{1.42}$, $R^2 = 0.946$). Orange dashed line is an allometric trend based on data for dried wings of Staphylinidae ($m_w = 1.33 \cdot S^{1.45}$, $R^2 = 0.992$). The data correspond to mass and area of one wing in Coleoptera and Diptera and of one forewing-hindwing pair in Hymenoptera.

Extended Data Fig. 2 Kinematics descriptions of *Paratuposa placentis* individuals PP2, PP4, PP5, PP12.

a, Euler angles of wings and elytra and body pitch angle. **b**, Vertical and horizontal components of body acceleration. **c**, Euler angles of wings and elytra and body pitch-up angle between horizontal plane and longitudinal axis of body, prepared for CFD.

Extended data Fig. 3 CFD simulation of the flight of *Paratuposa placentis*: visualization of the flow using two-dimensional slices in the 3rd simulation cycle.

Time instances are $t/T = 0.0$ (dorsal recovery stroke) $t/T = 0.37$ (first power stroke), $t/T = 0.60$ (ventral recovery stroke) and $t/T = 0.82$ (second power stroke), from top to bottom. **a**, Planes perpendicular to wings. **b**, Velocity vectors (arrows) and velocity magnitude (background colouring), showing two-component in-plane velocity with the third vector component omitted for clarity. **c**, Pressure. **d**, Magnitude of three-dimensional vorticity with colouring in logarithmic scale.

Extended Data Fig. 4 Calculated components of aerodynamic force of *Paratuposa placentis* individuals PP2, PP4, PP5, PP12.

a, Vertical and horizontal components of the aerodynamic force exerted on the insect. **b**, Total vertical aerodynamic force acting on the insect and its breakdown into the vertical component of the aerodynamic force on the pair of wings, the pair of elytra and the body. **c**, Lift-drag decomposition of the vertical force acting on one wing.

Extended Data Fig. 5 Three-dimensional reconstruction of wing-tip trajectories (continuous lines), aerodynamic force vectors (cyan arrows), velocity vectors (magenta arrows) and wing orientation (yellow circles and arrows) during flight in *Paratuposa placentis* individuals.

a, PP2. **b**, PP4. **c**, PP5. **d**, PP12. For interactive 3D pdf version, see Supplementary Fig. 9.

Extended Data Fig. 6 Body-mass-specific mechanical power components of *Paratuposa placentis* individuals PP2, PP4, PP5 and PP12.

The mechanical power essentially remains positive through the entire wingbeat cycle period, although occasionally it takes small negative values. The aerodynamic contribution is dominant.

Extended Data Fig. 7 Components of pitching moment acting on *Paratuposa placentis* individuals PP2, PP4, PP5, PP12.

Positive direction is nose down.

Extended Data Fig. 8 Aerodynamic performance of *Paratuposa placentis* individual PP2 obtained from three different simulations using three different values of maximum refinement level j_{\max} .

a, Aerodynamic force. **b**, Mass-specific aerodynamic power.

Extended Data Fig. 9 Comparison of *Paratuposa placentis* and other insects that have bristled or membranous wings.

a, Time variation of the vertical force coefficient in three different species. Data for the large rhinoceros beetle *Trypoxylus dichotomus* and for the tiny chalcid wasp *Encarsia formosa* are adapted so that the cycle begins with the downstroke. The force coefficient is defined as $C_V = 2F_V/\rho(2\Phi R_g f)^2 S$, where F_V is vertical force, ρ is air density, Φ is flapping amplitude, R_g is wing geometric radius of gyration, f is flapping frequency, and S is wing area. Note that the bristled wings of *Encarsia formosa* were modelled as impermeable solid plates. For additional information, see Supplementary Information. **b**, Wing length relative to body length.

Extended Data Fig. 10 Comparison of *Paratuposa placentis* and *Encarsia formosa*.

a, Time variation of body-mass-specific mechanical power. **b**, Mean (averaged over wingbeat cycle) and peak (maximum over the cycle) body-mass-specific mechanical power.

Supplementary information

Supplementary Information

This file contains Supplementary Sections 1–19, which include Supplementary Figs. 1–9 and Supplementary Tables 1–4.

Reporting Summary

Peer Review File

Supplementary video 1

Flight of *P. placensis* PP2. Synchronized videos in two orthogonal projections.

Supplementary video 2

Flight of *P. placensis* PP4. Synchronized videos in two orthogonal projections.

Supplementary video 3

Flight of *P. placensis* PP5. Synchronized videos in two orthogonal projections.

Supplementary video 4

Flight of *P. placensis* PP12. Synchronized videos in two orthogonal projections.

Supplementary video 5

Computer rendering of the *P. placensis* PP2 model and airflow visualization. Top left: rendering of the kinematic model used for CFD in lateral projection. Top right: the same model rendering in top projection. Bottom left and right: airflow visualization using iso-surfaces of the vorticity magnitude $|\omega|/f = 10$ and $|\omega|/f = 30$ in lateral and top projections. The laboratory reference-frame axes x , y and z are shown as red, yellow and green arrows, respectively.

Supplementary video 6

Computer rendering of the *P. placensis* PP5 model and airflow visualization. Top left: rendering of the kinematic model used for CFD in lateral projection. Top right: the same model rendering in top projection. Bottom left and right: airflow visualization

using iso-surfaces of the vorticity magnitude $|\omega|/f = 10$ and $|\omega|/f = 30$ in lateral and top projections. The laboratory reference-frame axes x , y and z are shown as red, yellow and green arrows, respectively.

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

Farisenkov, S.E., Kolomenskiy, D., Petrov, P.N. *et al.* Novel flight style and light wings boost flight performance of tiny beetles. *Nature* **602**, 96–100 (2022).
<https://doi.org/10.1038/s41586-021-04303-7>

- Received: 13 May 2021
- Accepted: 02 December 2021
- Published: 19 January 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04303-7>

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/s41586-021-04303-7>

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

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

Mutation bias reflects natural selection in *Arabidopsis thaliana*

- [J. Grey Monroe](#) ORCID: [orcid.org/0000-0002-4025-5572](#)^{1,2},
- [Thanvi Srikant](#)¹,
- [Pablo Carbonell-Bejerano](#) ORCID: [orcid.org/0000-0002-7266-9665](#)¹,
- [Claude Becker](#)¹ nAff10,
- [Marielle Lensink](#)²,
- [Moises Exposito-Alonso](#) ORCID: [orcid.org/0000-0001-5711-0700](#)^{3,4},
- [Marie Klein](#)^{1,2},
- [Julia Hildebrandt](#)¹,
- [Manuela Neumann](#) ORCID: [orcid.org/0000-0003-2778-3028](#)¹,
- [Daniel Kliebenstein](#) ORCID: [orcid.org/0000-0001-5759-3175](#)²,
- [Mao-Lun Weng](#)⁵,
- [Eric Imbert](#) ORCID: [orcid.org/0000-0001-9158-0925](#)⁶,
- [Jon Ågren](#) ORCID: [orcid.org/0000-0001-9573-2463](#)⁷,
- [Matthew T. Rutter](#)⁸,
- [Charles B. Fenster](#)⁹ &
- [Detlef Weigel](#) ORCID: [orcid.org/0000-0002-2114-7963](#)¹

[Nature](#) volume 602, pages 101–105 (2022)

- 90k Accesses
- 2 Citations
- 1125 Altmetric
- [Metrics details](#)

Subjects

- [Epigenomics](#)

- [Genetic variation](#)
- [Molecular evolution](#)
- [Mutation](#)

Abstract

Since the first half of the twentieth century, evolutionary theory has been dominated by the idea that mutations occur randomly with respect to their consequences¹. Here we test this assumption with large surveys of de novo mutations in the plant *Arabidopsis thaliana*. In contrast to expectations, we find that mutations occur less often in functionally constrained regions of the genome—mutation frequency is reduced by half inside gene bodies and by two-thirds in essential genes. With independent genomic mutation datasets, including from the largest *Arabidopsis* mutation accumulation experiment conducted to date, we demonstrate that epigenomic and physical features explain over 90% of variance in the genome-wide pattern of mutation bias surrounding genes. Observed mutation frequencies around genes in turn accurately predict patterns of genetic polymorphisms in natural *Arabidopsis* accessions ($r = 0.96$). That mutation bias is the primary force behind patterns of sequence evolution around genes in natural accessions is supported by analyses of allele frequencies. Finally, we find that genes subject to stronger purifying selection have a lower mutation rate. We conclude that epigenome-associated mutation bias² reduces the occurrence of deleterious mutations in *Arabidopsis*, challenging the prevailing paradigm that mutation is a directionless force in evolution.

[Download PDF](#)

Main

The random occurrence of mutations with respect to their consequences is an axiom upon which much of biology and evolutionary theory rests¹. This simple proposition has had profound effects on models of evolution developed since the modern synthesis, shaping how biologists have thought about and studied genetic diversity over the past century. From this view, for example, the common observation that genetic variants are found less often in functionally constrained regions of the genome is believed to be due solely to selection after random mutation. This paradigm has been defended with both theoretical and practical arguments: that selection on gene-level mutation rates cannot overcome genetic drift; that previous evidence of non-random mutational patterns relied on analyses in natural populations that were confounded by the effects of natural selection; and that past proposals of adaptive mutation bias have not been framed in the context of potential mechanisms that could underpin such non-random mutations^{3,4,5,6}.

Yet, emerging discoveries in genome biology inspire a reconsideration of classical views. It is now known that nucleotide composition, epigenomic features and bias in DNA repair can influence the likelihood that mutations occur at different places across the genome^{7,8,9,10,11,12,13}. At the same time, we have learned that specific gene regions and broad classes of genes, including constitutively expressed and essential housekeeping genes, can exist in distinct epigenomic states¹⁴. This could in turn provide opportunities for adaptive mutation biases to evolve by coupling DNA repair with features enriched in constrained loci². Indeed, evidence that DNA repair is targeted to genic regions and active genes has been found^{15,16,17,18,19,20}. Here we synthesize these ideas by investigating the causes, consequences and adaptive value of mutation bias in the plant *Arabidopsis thaliana*.

De novo mutations in *Arabidopsis*

The greatest barrier to investigating gene-level mutation variability has been a lack of data characterizing new mutations before they experience natural selection. We addressed this limitation by compiling large sets of de novo mutations in *A. thaliana* (hereafter referred to as *Arabidopsis*), for which there is rich information on sequence and epigenomic features plausibly linked to mutation rates. We first reanalysed existing *Arabidopsis* mutation accumulation lines¹², combining putative germline and somatic mutations (Fig. 1a, Extended Data Figs. 1, 2, Supplementary Data 1; Methods). A filtering pipeline to eliminate false positives and based on mapping quality, depth and variant frequency retained less than 10% of called variants in a final high-confidence set of mutations. We found no evidence of selection on these mutations. The germline mutations had accumulated in randomly chosen single-seed descendants, so very few mutations, only those causing inviability or sterility, should have been removed by selection¹². Somatic mutations experience even less selection^{21,22}. Therefore, as expected, non-synonymous changes and premature stop codons accounted for a greater share of variants than in natural populations, and their frequencies were indistinguishable from a null model of random mutation. We also confirmed that there was no bias in detecting non-synonymous mutations when comparing genes predicted to be sensitive or insensitive to mutation (Fig. 1b).

Fig. 1: Identifying epigenomic and other features associated with mutations in *Arabidopsis*.

 **figure 1**

a, Experimental design for identifying germline and somatic mutations in the main dataset¹². **b**, Relaxed purifying selection in de novo mutation calls: rates of non-synonymous (non-syn) and stop codon variants (stop) as compared with polymorphisms detected in 1,135 natural accessions from the 1001 Genomes (1001G)

project³⁵ and to a null model based on mutation spectra and nucleotide composition of coding sequences. Comparison of de novo mutations between genes predicted to have or not have lethal effects when mutated is also shown³⁷. P values from χ^2 test; * $P < 0.05$. NS, not significant. **c**, Genome-wide distributions in gene body density, observed mutation rates and candidate predictive features in relation to transcription start sites (TSS) and transcription termination sites (TTS). Darker shading represents greater density. SNV, single-nucleotide variant; CHGm, CHHm, CGm, methylation in the CHG, CHH and CG contexts, respectively. **d**, Modelling approach to predict mutation probability from a range of features. ATAC-seq, assay for transposase-accessible chromatin using sequencing; AIC, Akaike information criterion. **e**, Predictive models and t -values of predictor variables from the generalized linear model.

Epigenome-mediated mutation bias

We tested whether the location of mutations in our dataset was associated with epigenomic features, focusing on biochemical properties previously linked to mutation: gene expression, GC content, cytosine methylation, histone modifications and chromatin accessibility (Fig. 1c). We built linear models of mutation frequencies in genic regions as a function of these features across the genome (Fig. 1d; Methods).

These models revealed features positively and negatively associated with mutations, with several having been already linked to mutagenesis or DNA repair (Fig. 1e). For example, the negative relationship between GC content and mutation²³ is consistent with GC-biased gene conversion²⁴ and reduced DNA denaturation in GC-rich regions²⁵. Likewise, previous work has linked H3K4me1 to genome stability, DNA repair and lower mutation rates^{26,27,28,29,30,31}. By contrast, methylated cytosines correlate with elevated mutation rate, consistent with the effects of cytosine deamination^{12,32}, while highly accessible chromatin regions (for example, transcription factor-binding sites) can impair nucleotide excision repair³³. In conclusion, we uncovered associations between mutation frequencies and biochemical features known to affect DNA repair and vulnerability to damage.

We note in advance here that all downstream analyses led to the same conclusions for single-nucleotide variants (SNVs) and insertions and deletions (indels), or for germline and somatic mutations. All were less frequent in gene bodies and essential genes, and we therefore report combined results. Our conclusions also did not change when we repeated the analyses after training our initial epigenome prediction model on non-coding regions only. Finally, we confirmed that observed mutation biases could not be explained by variation in read depth, mappability, the distribution of false positives or selection on mutations (Extended Data Fig. 3).

Lower mutation rate in gene bodies

We calculated predicted mutation probabilities (predicted mutations per base pair) as a function of epigenomic features around genes and found that mutation rates were lower within gene bodies (Fig. 2a). These predictions were confirmed by observed mutations in multiple independent datasets (Fig. 2b, Supplementary Data 1). We called mutations in new *Arabidopsis* mutation accumulation populations, the largest reported to date: germline and somatic mutations in 400 lines established from eight genetically diverse founder genotypes, four each from the extreme North and South of Europe. Observed distributions of germline and somatic mutations were very similar to epigenome-predicted mutation rates. These data also provided evidence for genetic variation in mutation bias, raising the possibility of mutation bias evolvability (Extended Data Fig. 4). Somatic variants identified from 10 rosettes and from reanalysing deep sequencing data of 64 leaves in two *Arabidopsis* plants²¹ further confirmed predicted patterns, as did previously discovered germline mutations in a bottlenecked *Arabidopsis* lineage³² (Extended Data Figs. 3, 4).

Fig. 2: Lower mutation rate in gene bodies.

 **figure 2**

a, Mutation probability score (predicted SNVs plus indels per base pair from models in Fig. 1e; mean \pm 2 s.e.m. in grey) based on epigenomic states and mutations observed in original mutation accumulation lines. **b**, Observed de novo mutations from all independent mutation accumulation datasets (mean \pm 2 s.e.m. in grey),

bootstrapped). **c**, Segregating polymorphisms (SNVs plus indels, S , mean \pm 2 s.e.m. in grey, bootstrapped) in 1,135 *Arabidopsis* accessions³⁵. **d**, Tajima's D calculated from polymorphisms in *Arabidopsis* accessions³⁵ around TSS and TTS (mean \pm 2 s.e.m. in grey). Note that these TSS and TTS plots do not consider gene length or intergenic distances and that, for example, not all sequences downstream of TSSs are genic sequences, and not all sequences upstream of TSSs are intergenic sequences. Specifically, we did not distinguish between intergenic regions (or genes) longer or shorter than 3,000 bp.

By combining mutation datasets, we found that the frequency of mutation was 58% lower in gene bodies than in nearby intergenic space. Epigenome-predicted mutation probabilities explained over 90% of the variance in the pattern of observed mutations around gene bodies (Fig. 2a, b, Extended Data Fig. 5). Since only 20–30% of gene body sites are estimated to be subject to selection, mutation bias in genic regions could affect sequence evolution around genes more than selection³⁴.

Genetic diversity in a global set of *Arabidopsis* accessions³⁵ supported these results (Fig. 2c, d). Over 90% of the variance in polymorphism levels found around gene bodies could be explained by our experimentally observed mutation rates (Extended Data Fig. 5). To determine whether low levels of polymorphism in gene bodies were indeed caused by reduced mutation rather than purifying selection, we analysed the site frequency spectrum. Theory shows that purifying selection causes an enrichment of rare alleles (reduced frequency of deleterious variants), whereas site frequency spectrum scales with mutation rate such that lower mutation rate causes a depletion of rare alleles (fewer young alleles)³⁶. Our analysis of the site frequency spectrum statistic Tajima's D around genes confirmed a depletion of rare alleles in gene bodies (less negative D), consistent with a reduced mutation rate. We validated this inference with extensive forward population genetic simulations (Extended Data Fig. 6). In conclusion, evolution around genes in *Arabidopsis* appears to be explained by mutation bias to a greater extent than by selection.

Gene structure and mutation

We further discovered emergent relationships between gene structure and mutation rate (Extended Data Fig. 7). Owing to the distribution of epigenomic features along gene bodies, mutation probabilities are highest in extreme 5' and 3' coding exons. Natural polymorphisms in *Arabidopsis* and *Populus trichocarpa* showed a similar pattern. Consistent with the effects of mutation bias, D was more negative in peripheral exons. The predicted mutation rate of coding regions was 28% and 39% higher in genes annotated as lacking 5' untranslated regions (UTRs) and 3' UTRs, respectively. The inferred effect size of 5' UTRs and 3' UTRs on coding-exon mutation probabilities and polymorphism was greatest in extreme 5' and 3' coding

exons. UTR lengths were negatively correlated with mutation probabilities and polymorphisms in peripheral coding exons. Mutation probabilities were also 90% greater in genes lacking introns and lower in genes with more ($r = -0.34$) and longer ($r = -0.24$) introns. These patterns were mirrored by patterns of polymorphism and Tajima's D . In conclusion, an unexpected emergent effect of UTRs and introns in *Arabidopsis* appears to be lower mutation rates in coding regions.

Fewer mutations in essential genes

We next investigated mutation rates in relation to gene functions, discovering that genes with the lowest epigenome-predicted mutation rates were enriched for conserved biological functions (for example, translation). By contrast, genes with the highest predicted mutation rates had specialized functions (for example, environmental response) (Fig. 3a). Comparing genes whose effects have been measured with knockout experiments³⁷ confirmed that essential genes are enriched for epigenomic features associated with low mutation, and, as predicted, observed mutation rates were significantly lower in the coding regions of essential genes. By contrast, genes with environmentally conditional functions had the highest mutation rates. Intron mutations showed the same pattern, confirming that these results are not due to selection on coding sequences biasing our mutation datasets (Fig. 3c). We found no evidence that reduced mutation rate in essential genes could be explained by the potential intrinsic mutational properties of CG methylation, expression level or GC content. Instead, the observed 37% reduction in mutation rates in essential genes is consistent with a reduction in mutation, plausibly explained by their enrichment for low-mutation-associated epigenomic features (for example, H3K4me1).

Fig. 3: Lower mutation probability in essential genes.

 **figure 3**

a, Variation in epigenome-derived mutation probability scores in coding sequence (CDS) among genes and gene ontology terms enriched in genes in the top ('high-mutation probability genes') and bottom ('low-mutation probability genes') deciles. Mm., macromolecular; N₂, nitrogen; nucleob., nucleobase-containing

compound; reg., regulation; resp., response. **b**, Enrichment of epigenomic and other features in coding sequences of 719 genes known to be essential from mutant analyses ($\text{mean} \pm 2 \text{ s.e.m.}$). **c**, Total observed mutation rate ($\pm 2 \text{ s.e.m.}$, bootstrapped) in genes ($n = 2,339$) with experimentally determined functions³⁸. The bars are coloured according to relative differences in mutation rates among genes classified by function (that is, orange refers to high mutation rate and blue represents low mutation rate). $P \approx 0$ for both CDS and intron mutations.

These results were further supported by our discovery of reduced mutation rate in genes with lethal knockout effects³⁸ and broadly expressed genes³⁹. Again, these results were consistent with epigenomic profiles (Extended Data Fig. 8). In conclusion, we find that genes with the most important functions experience reduced mutation rate, as predicted by their epigenomic features.

Reduction in mutation load

Comparing predicted mutation rates with signatures of evolutionary constraint revealed that genes subject to purifying selection are enriched for epigenomic features associated with low mutation rate (Fig. 4a, b). We confirmed these predictions with our dataset of empirical mutations—mutation rate was significantly correlated with measures of evolutionary constraint on coding and regulatory function (Fig. 4a, c). These patterns were replicated in analyses of mutations in introns, where selection is weaker than in exons, further indicating that results are not due to selection biasing our mutation datasets. These findings demonstrate that genes subject to stronger purifying selection are maintained in epigenomic states that underlie a significant reduction in their mutation rate (Extended Data Fig. 9). In conclusion, mutation bias acts to reduce levels of deleterious variation in *Arabidopsis* by decreasing mutation rate in constrained genes.

Fig. 4: Adaptive reduction in deleterious mutations.

 **figure 4**

a, Correlations between epigenomic and other features, predicted and observed mutation rates, and measures of evolutionary constraint and rates of sequence evolution. Synonymous (Ps) and non-synonymous polymorphism (Pn) in natural populations, synonymous (Ds) and non-synonymous divergence (Dn) from *Arabidopsis lyrata*, environmental variance of gene expression (Ve expr.) and genetic variance of gene expression (Vg expr.). ‘Pred. mut.’ is the predicted mutation rate as a function of epigenomic and other features. ‘Obs. mut.’ is the observed mutation rate in genes based on de novo mutations called across all mutation accumulation datasets. *** $P < 2 \times 10^{-16}$, ** $P < 0.05$. **b, c**, Relationship between H3K4me1 (**b**) and estimates of evolutionary constraint and rate of sequence evolution (**c**) across quantiles of observed mutation rates per gene. Pearson correlation reflects raw correlation across genes. Data are visualized by mean values \pm 2 s.e.m. in 50 quantiles (each quantile = 2% of genes). **d**, Conceptual diagrams summarizing our findings.

Evolution of mutation bias

Our findings reveal adaptive mutation bias that is mediated by a link between mutation rate and the epigenome. This is mechanistically plausible in light of evidence that DNA repair factors can be recruited by specific features of the epigenome⁸. Hypomutation targeted to features enriched in functionally constrained loci throughout the genome would reduce the relative frequency of deleterious mutations. The adaptive value of this bias can be conceptualized by the analogy of loaded dice with a reduced probability of rolling low numbers (that is, deleterious mutations), and thus a greater probability of rolling high numbers (that is, beneficial mutations) (Fig. 4d).

This intuitive model fits established theory showing that adaptive mutation bias could evolve despite drift when the length of sequence affected (L_{segment}) is large^{2,3,5,40}.

While this criterion can rarely be satisfied for single-gene modifiers, it can be if the mutation is suppressed in many constrained loci. For example, the total sequence length of the coding regions of essential genes enriched for H3K4me1 is three times the estimated minimum L_{segment} required for targeted hypomutation to evolve in *Arabidopsis*, assuming a 30% reduction in mutation rate (Extended Data Fig. 10). Thus, while perhaps initially surprising, our synthesis between epigenomics and population genetic theory predicts that the observed biases could readily arise via natural selection².

Conclusions

While it will be important to test the degree and extent of mutation bias beyond *Arabidopsis*, the adaptive mutation bias described here provides an alternative explanation for many previous observations in eukaryotes, including reduced genetic variation in constrained loci⁴¹ and the genomic distributions of widely used population genetic statistics⁴². Since mutational biases are a product of evolution, they could differ between organisms, potentially explaining differences in the distribution of fitness effects of new mutations among species^{43,44}. Finally, because epigenomic features are plastic, epigenome-associated mutation bias could even contribute to environmental effects on mutation⁴⁵. Our discovery yields a new account of the forces driving patterns of natural variation, challenging a long-standing paradigm regarding the randomness of mutation and inspiring future directions for theoretical and practical research on mutation in biology and evolution.

Methods

Identification of de novo mutations in *A. thaliana*

Col-0 mutation accumulation lines

Our training set of mutations was identified from 107 mutation accumulation lines of the *A. thaliana* Col-0 accession, which is the basis of the *A. thaliana* TAIR10 reference genome sequence¹². The lines had been previously grown for 24 generations of single-seed descent before sequencing with 150-bp paired-end reads on the Illumina HiSeq 3000 platform, of pools of approximately 40 seedlings of each line from the 25th generation (Fig. 1a). Seedlings were sampled at the four-leaf stage, at 2 weeks of age. Variants were identified with GATK HaplotypeCaller¹². In many organisms, germline mutations are primarily influenced by processes specific to reproductive organs¹⁰.

Because plants may lack a completely segregated germline⁴⁶, we hypothesized that mechanisms that influence local mutation rates in the germline may be reflected in the distribution of somatic mutations as well, or at least that the processes governing mutation rate variability across the genome may be similar in germline and somatic tissue. Therefore, in addition to the original variants called¹², we implemented a custom filtering pipeline to identify a high-confidence set of additional de novo mutations (Extended Data Fig. 1). This set included, in addition to somatic variants, germline variants that had not been called in the original analyses¹². Somatic mutations were previously excluded because they appear as heterozygous calls¹². Germline mutations were previously excluded if at least 1 out of the 107 lines also included a putative somatic mutation at the same position¹². On the basis of previously reported germline mutation rates (1–2 per genome and generation) and with the knowledge that these lines were self-fertilized each generation, we expected the seedlings that were sequenced to be segregating for 2–4 additional heterozygous germline variants, which would have been called as somatic mutations by our pipeline (approximately 2–5% of putatively somatic mutations). Because we combined putative somatic and germline mutations to characterize the mutational landscape of the *A. thaliana* genome, this did not have an obvious effect on our results.

Testing for mutation calling artefacts by resequencing ten siblings of a single-mutation accumulation line

To test for the possibility that our results were in part artefacts of the pooled-seedling sequencing approach¹², we resequenced entire rosettes of individual plants that were sibling from the same mutation accumulation line (#73) and asked whether the distribution of called variants (that is, putative somatic mutations around TSS and TTS) was similar to the patterns seen with the seedling pools of the 107 individual lines described in the preceding section (Extended Data Fig. 6). Specifically, we grew 10 siblings of line #73 and extracted DNA from 3-week-old whole rosettes. Barcoded PCR-free libraries for the 10 siblings were sequenced, with 150-bp paired-end reads, at approximately 60× depth each on a single lane of the Illumina HiSeq 3000 platform. Additionally, for one sibling, the same library was sequenced in an independent lane at approximately 600× depth. After adapter and quality trimming with cutadapt (version 2.3) and removing duplicates with samtools markdup (version 1.10), reads were aligned to the TAIR10 reference genome with bwa-mem (version 0.7.17) and variants were called independently for each sample with GATK HaplotypeCaller version 4.1.0.

Measuring the effects of mappability of reads

We wanted to ensure that variation in mappability could not explain the observed distribution of de novo variants. To evaluate the possibility that results were an artefact

of bias in mappability across gene regions, we calculated mappability for $k = 100$, $e = 1$, across the *A. thaliana* reference genome using GenMap⁴⁷. We then plotted and visualized mappability around TSSs and TTSs to confirm that differences in mappability were not the same as the signals of mutation bias detected in our numerous datasets of de novo mutation. While we did not see any evidence that mappability bias covaried with patterns of mutation bias, for building our predictive model of mutation rate as a function of epigenomic and other features, we still chose to filter out variants called in regions of poor mappability (± 100 bp of mappability < 1), as our analysis of resequenced siblings suggested that variants called in low-mappability regions are more likely to be false positives (since variants called in many independent lines had lower mappability).

Simulating reads and identifying true false positives

To further rule out artefacts, we calculated the expected distribution of false positives using simulated short reads. We simulated Illumina reads based on the TAIR10 reference genome using ART⁴⁸ with the following parameters: -l 150 -f 30 -m 500 -s 30. Reads were mapped to the TAIR10 genome with NextGenMap, the same caller as used in the original calling of mutation accumulation lines⁴⁹, and variants were called with GATK HaplotypeCaller. This was repeated for a total of 1,000 simulated genomes. Because these are simulated reads, all variants that are called must be false positives. To test the possibility that the main results found in this study, such as elevated mutation and polymorphism upstream of TSSs, are artefacts of bias resulting from Illumina sequencing (which is included in simulations) or from mapping error (which is captured by mapping the simulated reads), we plotted the distributions of false positives around these regions to confirm that the distribution of false positives was more similar to likely false positives (for example, called in many lines) and unlike the higher confidence variants called in real sequencing data.

Identification of de novo mutations in a new *A. thaliana* mutation accumulation experiment

To validate our predictive model of the mutation probability score, we used a second *A. thaliana* mutation accumulation experiment descended from eight founders collected in natural environments⁵⁰. The lines were grown for seven to ten generations of single-seed descent before 150-bp paired-end read Illumina sequencing of pools of 40 seedlings. The specifics of the populations were as follows: founder CN1A18: 56 lines for 10 generations; founder CN2A16: 51 lines for 10 generations; founder SJV12: 48 lines for 7 generations; founder SJV 15: 36 lines for 7 generations; founder RÖD4: 50 lines for 8 generations; founder RÖD6: 50 lines for 8 generations; founder SB4: 53 lines for 8 generations; and founder SB5: 56 lines for 8 generations.

Mutations were identified as described in ref. [11](#). Briefly, raw reads were mapped to the TAIR10 reference genome, variants were called using GATK HaplotypeCaller, merged with the GenotypeGVCFs tool and filtered by variant quality ($QD > 30$) and read depth ($DP > 3$). A germline mutation was called if a single mutation accumulation line per founder population had a homozygous alternative allele. Somatic mutations were called as heterozygous variants found in only one of the mutation accumulation lines derived from a single founder genotype. This should remove any true heterozygous calls, variants between cryptic duplications in the founder, and low confidence calls, as suggested by our preceding analyses by resequencing siblings from the original mutation accumulation experiment.

Identification of de novo somatic mutations in a resequencing dataset of *A. thaliana* leaves

To further test our power to predict the distribution of de novo mutations in an independent experiment, we used published data generated from Illumina sequencing of 64 samples of leaf tissue (rosettes and cauline leaves) of two Col-0 plants^{[21](#)}. Raw fastq files were downloaded from NCBI and mapped to the TAIR10 reference genome using bwa-mem, and duplicate reads (that is, PCR duplicates) were filtered using samtools markdup. Variants for every sample were called with GATK HaplotypeCaller. Variants were filtered to include only those found in a single sample (as our previous work had already shown that putative somatic variants called in many independent samples tend to be enriched for regions of low mappability and exhibit distributions more similar to the expected distribution of false positives).

De novo mutations in a natural mutation accumulation lineage

We analysed mutations that had accumulated in a single *A. thaliana* lineage that recently colonized North America^{[32](#)}. The 100 samples came both from modern populations as well as historical herbarium specimens and contained 8,891 new variants with at least 50% genotyping rate in the population. Phylogenetic coalescent analyses indicated that these 100 samples shared a common ancestor around 1519–1660, presumably the ancestor that colonized North America, and thus that these lines have recent mutations that accumulated after a population bottleneck (small N_e) and therefore under weak selection^{[32](#)}. We used these to study the level of polymorphisms around TSSs and TTSs in a wild population with a simple demographic history.

Constructing a model to predict mutation probability

Sequence and epigenomic features

We were interested in studying epigenomic features plausibly linked to mutation rate^{[16,17,18,19,28,51,52,53,54,55](#)}. To build a high-resolution predictive model of mutation rate variation, we extracted or generated data describing genome-wide sequence and epigenomic features. First, we calculated GC content (% of sequence), which can affect DNA denaturation^{[5,25,56,57,58](#)}, across regions^{[9,23,59,60,61,62,63,64](#)}. From the Plant Chromatin State Database, we also downloaded 62 BigWig formatted datasets characterizing the distribution of histone modifications^{[14](#)} H3K4me2, H3K4me1, H3K4me3, H3K27ac, H3K14ac, H3K27me1, H3K36ac, H3K36me3, H3K56ac, H3K9ac, H3K9me1, H3K9me2 and H3K23ac, many of which have been linked to mutational processes^{[8,9,11,12,19,33,65,66,67,68,69,70](#)}. For each specific histone modification, depths were scaled (0 to 1) and averaged across each region for downstream analyses.

Col-0 cytosine methylation

Because cytosine methylation is known to affect mutation rates via deamination of methylated cytosines^{[9,11,12,33,66](#)}, we wanted to include cytosine methylation as a predictor variable in our model. Methylated cytosine positions for Col-0 (6909) wild-type leaves were obtained from the 1001 Epigenomes dataset GSM1085222 (ref. ^{[71](#)}) under the file GSM1085222_mC_calls_Col_0.tsv.gz. Because the context of cytosines can vary and influence the functional effect of methylation, cytosines were further classified into three categories (CG/CHG/CHH) for all downstream analyses. For each region, we calculated the number of methylated cytosines in each category per bp.

Chromatin accessibility

ATAC-seq can measure chromatin accessibility, which also affects mutation rates^{[9,11,12,33,66,72](#)}. Col-0 seeds were stratified on MS-agar (with sucrose) plates at 4 °C for 4 days in the dark. Plates were transferred to 23 °C long-days and kept vertically for easier harvesting of seedlings. On the eleventh day of light exposure, 10–20 seedlings each from three MS-agar plates were fixed with formaldehyde by vacuum infiltration and stored at –80 °C.

Fixed tissue was chopped finely with 500 µl of general purpose buffer (GPB; 0.5 mM spermine•4HCl, 30 mM sodium citrate, 20 mM MOPS, 80 mM KCl, 20 mM NaCl, pH 7.0, sterile filtered with a 0.2-µm filter, followed by the addition of 0.5% of Triton-X-100 before usage). The slurry was filtered through one-layered Miracloth (pore size: 22–25 µm), followed by filtration through a cell strainer (pore size: 40 µm) to collect nuclei. Approximately 50,000 DAPI-stained nuclei were sorted using fluorescence-activated cell sorting (FACS) as two technical replicates. Sorted nuclei were heated to 60 °C for 5 min, followed by centrifugation at 4 °C (1,000g for 5 min). Supernatant was removed, and the nuclei were resuspended with a transposition mix (homemade

Tn5 transposase, a TAPS-DMF buffer and water) followed by a 37 °C treatment for 30 min. 200 µl SDS buffer and 8 µl 5 M NaCl were added to the reaction mixture, followed by 65 °C treatment overnight. Nuclear fragments were then cleaned up with Zymo DNA Clean & Concentrator columns. 2 µl of eluted DNA was subjected to 13 PCR cycles, incorporating Illumina barcodes, followed by a 1.8:1 ratio clean-up using SPRI beads. Genomic DNA libraries were prepared using the same library preparation protocol from the Tn5 enzymatic digestion step onwards.

Each technical replicate (derived from nuclei sorting) was sequenced with 3.5 million 150-bp paired-end reads on an Illumina HiSeq 3000 instrument. The reads were aligned as two single-end reads to the TAIR10 reference genome using bowtie2 (default options), filtered for the SAM flags 0 and 16 (only reads mapped uniquely to the forward and reverse strands), and converted separately to .bam files. The .bam files were merged, sorted, and PCR duplicates were removed using picardtools. The sorted .bam files were merged with the corresponding sorted bam file of a second technical replicate (samtools merge --default options) to obtain a final depth of approximately 6 million reads for each replicate.

Peaks were called for each biological replicate using MACS2 using the following parameters:

```
macs2 callpeak -t [ATACseqlibrary].bam -c [Control_library].bam -f BAM --nomodel  
--shift -50 --extsize 100 --keep-dup=1 -g 1.35e8 -n [Output_Peaks] -B -q 0.05
```

Peak files and .bam alignment files from three biological replicates were processed with the R package DiffBind to identify consensus peaks that overlapped in at least two replicates (FDR < 0.01). Library quality was estimated by measuring the frequency of reads in peak (FRIP) scores for all three replicates, which were 0.36, 0.36 and 0.39, above the standard quality threshold of 0.3.

Gene expression

Gene expression was calculated as the mean across 1,203 accessions⁷¹, from which we also extracted the genetic variance (Vg) and environmental variance (Ve) as well as the coefficient of variation (variance/mean) in expression for each gene. This dataset provided information for 17,247 genes with complete data.

Predictive model of mutation rates

We wanted to ask whether intragenomic mutation variability in the genome could be predicted by features of the genome that previous work had shown to have potential or demonstrated relationships with mutations. To model mutation rate genome-wide at

the level of individual genes, we created a generalized linear model. The response variable was the untransformed (that is, assuming normality, to avoid risk of increased false positives caused by transformation^{73,74}) observed mutation rate across every genic feature (upstream, UTR, coding, intron and downstream). The predictor variables were GC content, classes of cytosine methylation, histone modifications, chromatin accessibility and expression of each gene. From this full model, a limited predictive model was selected on the basis of forward and backward selection with the lowest AIC value by the stepAIC function in R. These models were created separately for indels (adjusted R -squared: 0.001791; F -statistic: 34.6 on 16 and 299635 d.f.; $P < 2.2 \times 10^{-16}$) and SNVs (adjusted R -squared: 0.0009687; F -statistic: 37.32 on 8 and 299643 d.f.; $P < 2.2 \times 10^{-16}$). For downstream analyses, we used the predicted mutation probability (the mutation probability score) based on these models (predicted SNVs + indels) for genes, exons and other regions of interest from the TAIR10 genome annotation. While the linear regression approach used here enables hypothesis testing to some extent (one can generate confidence intervals and P values describing the level of significance of individual effects), our primary goal was to create a predictive model of mutation bias as a function solely from genomic and epigenomic features; the causality of the associations uncovered in these analyses for individual predictors must be confirmed with future functional work.

Variance inflation factor

To test whether our results were skewed by overly correlated predictor variables (included in the model even after model reduction by minimizing AIC), we explored models where predictor variables were manually removed on the basis of their variance inflation factor score. Specifically, we used the vif function from the R package car to calculate variance inflation factor scores for each variable in our best AIC models for SNVs and indels. We then removed all variables with scores below 3. We recalculated mutation probability scores for every genomic feature. Because the resulting predicted mutation probability scores were very similar, with Pearson correlation $r = 0.95$ between gene-level mutation probability scores from the full model and the reduced model, we report only results based on the full model.

Analysis of natural polymorphism rates

Rates of polymorphism among genic exons

We calculated rates of natural polymorphism across exons in TAIR10 gene models from sequence variation among 1,135 natural *A. thaliana* accessions³⁵. These analyses revealed elevated polymorphism rates in peripheral (first and last) exons. To test whether this is an artefact unique to *A. thaliana*, we calculated rates of natural

polymorphism across exons from sequence variation among 544 *P. trichocarpa* accessions⁷⁵. Specifically, we downloaded VCF and annotation data from Phytozome (v3.0) and calculated rates of variation across exons grouped by order (from 5' to 3') and total exon number.

Signatures of selection and constraint from natural populations

We calculated gene-level summary statistics for signatures of selection and constraint in the following way. Synonymous and non-synonymous polymorphism among natural *A. thaliana* accessions and divergence from *A. lyrata* (Pn, Ps, Dn and Ds, respectively) were calculated using mkTest.rb (<https://github.com/kr-colab>). The alpha test statistic for evidence of selection, which is a derivative of the McDonald-Kreitman test^{76,77,78}, was calculated from these values for each gene where data were available (not all genes have orthologues assigned in *A. lyrata*) as $1 - (Ds \times Pn) / (Dn \times Ps)$. Positive values of alpha are conventionally interpreted as evidence of positive selection because non-synonymous variants in genes with such values tend to become fixed. For each decile of genes classified according to mutation probability, we calculated the proportion for which alpha is positive. Enrichment of non-synonymous variants compared to genome-wide average were confirmed by independent calculation of Waterson's diversity estimate (θ) of non-synonymous variation. The frequency of loss-of-function mutations was calculated as before^{79,80}, where loss of function was defined as premature stop codons and frameshifts disrupting at least 10% of the coding region of the canonical gene model. Genes experiencing purifying selection should exhibit lower levels of natural polymorphism than what would be predicted by mutation rate alone. To test this, we built a linear model of coding region polymorphisms as a function of predicted mutation rates. We calculated scaled residuals for each gene and tested whether they are more negative in genes expected to be under purifying selection. To estimate constraints on gene regulatory function, we looked at average expression across diverse genotypes. We also tested for relationships between predicted mutation rates and the coefficient of variation in gene expression, additive genetic variance for gene expression across diverse genotypes, and environmental variance in gene expression⁷¹.

Relationships between epigenomic and other features, mutation rates and gene function

The preceding analyses revealed significant associations between epigenomic and other features and signatures under selection indicating that genes that experience purifying selection are enriched for features associated with low mutation rate. To further dissect the mechanistic basis of this pattern, we wanted to directly test for relationships between epigenomic states, mutation rates and gene function. We

analysed gene ontology categories for genes in the top and bottom deciles ranked by predicted mutation rate⁸¹, reporting gene ontologies that were significantly enriched in these groups after Bonferroni adjustment of raw P values.

We also analysed a manually curated dataset of mutation-induced lethality obtained from phenotyping lines with loss-of-function mutations³⁷. Genes annotated as lethal effect when mutated (that is, required for viability) were compared with genes showing non-lethal phenotypic effects to assess differences in epigenomic and other features.

We analysed a dataset of phenotypes from 2,400 *A. thaliana* knockout lines³⁸. Genes had been classified as being essential (such as an RNA processing gene where loss of function results in lethality⁸²), causing morphological defects (for example, altered stomata and trichome size), cellular biochemical defects (for example, intracellular transport of small molecules) and conditional defects (for example, effects depending on the environment). We then compared epigenomic and other features in essential genes to other classes of genes. These analyses showed that genes with essential functions were enriched for features associated with reduced mutation, whereas genes annotated as having non-essential functions were depleted for these features.

Estimating selection on different types of de novo mutations

Synonymous, non-synonymous and stop-gained variants are expected to have different effects on gene function, although they are of the same mutational class (SNVs). They are all from coding regions, which have an overall mutation probability that is distinct from other regions of the genomes, such as introns, in our model of de novo mutations. For comparison, we calculated the rates of synonymous, non-synonymous and stop-gained SNVs in natural populations of *A. thaliana*, which have been subject to long-term natural selection. We also derived an expected null ratio of non-synonymous to synonymous mutations using knowledge on the relative base composition of all coding regions in the reference genome, the relative proportion of coding region mutations (for example, CG to TA mutations are most common), and the proportion of all possible codon transitions that lead to synonymous versus non-synonymous mutations. Ratios of non-synonymous to synonymous and stop-gained to synonymous mutations were compared between observed de novo mutations and those observed in natural populations or the null expectation by chi-squared tests.

Expected non-synonymous-to-synonymous substitution ratios in the absence of selection

To further validate that the observed de novo mutations we used to train our mutation probability model were not subject to appreciable selection, we simulated 10,000 de

novo mutations across the *Arabidopsis* genome with custom scripts in R. Mutations in coding regions were randomly assigned to non-synonymous or synonymous changes based on codon use and observed mutational spectra of coding regions. We then calculated the observed ratio of non-synonymous to synonymous mutations in the simulated data. We repeated this simulation 10,000 times to produce a distribution of expected non-synonymous-to-synonymous ratios. We then compared the non-synonymous-to-synonymous ratio in our observed de novo mutations to this distribution. Finally, we tested whether our observation fell within the 95% bootstrapped interval.

Expected number of synonymous mutations under random variation

Because we had found that observed mutations were less frequent in coding regions, we wanted to determine whether this difference was significantly higher than expected by chance. We therefore asked how the number of synonymous mutations observed compared with that expected under a random process, starting with a simulated set of random mutations across the genome. We calculated the number of these mutations in coding regions that are expected to lead to a synonymous nucleotide substitution based on codon use and observed mutational spectra of coding regions. We repeated this simulation 1,000 times to generate a distribution of expected synonymous mutations. Comparing our observed de novo synonymous mutations to the mean of this distribution, we calculated the reduction in the observed synonymous mutation rate.

Non-synonymous-to-synonymous ratios and mutation probabilities in more deleterious ('lethal effect versus non-lethal effect') genes

We wanted to test whether the rates of non-synonymous-to-synonymous variation were lower in genes that are predicted to experience stronger negative selection. We split genes with a high-essentiality and low-essentiality prediction score (see above) or empirically determined lethal versus non-lethal effects of loss-of-function alleles (see above)³⁷. We then calculated the differences in the observed mutation rate between these groups of genes and compared them with a *t*-test. We also calculated the number of observed non-synonymous and synonymous SNVs in these groups of genes and compared their ratios by a chi-squared test.

Non-synonymous-to-synonymous ratios in mutation probability deciles

We wanted to test whether mutation probability deciles predicted by our model differed in their rates of non-synonymous to synonymous mutations in our observed de novo mutations. If there was a strong gradient (for example, if genes predicted to have low mutation rate had lower rates of non-synonymous variation than genes predicted

to have high mutation rate), this could suggest an effect of purifying selection acting directly on the detected mutations. To improve the power to detect differences among genes differing by mutation probability scores, we also assigned mean expression values to genes for which expression could not be called in our expression dataset⁷¹ and calculated mutation probability score. We binned genes into mutation probability deciles and compared mutation deciles and their corresponding non-synonymous-to-synonymous ratio to confirm that there was no relationship suggestive of selection.

Minor allele frequencies in natural populations

Our results had indicated that mutation rates were high upstream and downstream of genes relative to the gene bodies, not only in observed and predicted de novo mutations but also in natural polymorphisms. If this pattern was driven by mutation bias, we would expect to see lower minor allele frequencies upstream and downstream of genes, because this would indicate the presence of newly derived alleles from recent mutation rather than lower minor allele frequency caused by greater negative selection since we expect a priori that gene bodies (particularly coding regions whose code makes them sensitive to mutation) are subject to greater constraint. Conversely, lower minor allele frequencies in gene bodies would be consistent with the action of purifying selection in gene bodies, because lower allele frequencies are expected when negative selection had an opportunity to reduce allele frequencies. We therefore calculated the minor allele frequency (vcftools --freq) and their mean for every polymorphic position in the genome of 1,135 natural *A. thaliana* accessions³⁵ in relation to TSSs and TTSs across the entire genome.

Tajima's *D* around gene bodies

Tajima showed that reduced mutation and purifying selection, while having the same effect to reduce the number of polymorphisms, have opposite effects on his statistic, *D*³⁶. That is, mutation rate has a scaling effect on *D* such that reduced mutation rates lead to less negative *D*, whereas purifying selection leads to more negative *D*. Therefore, analysis of *D* can be used to quantify the relative importance of these alternative, but not mutually exclusive, forces shaping rates of sequence evolution. *D* is, on average, negative across the *A. thaliana* genome, and *D* also scales with mutation rate. Thus, if *D* is more negative in regions with lower polymorphism, this could indicate that purifying selection is the dominant force underlying lower rates of variation. By contrast, if *D* is less negative in regions of low polymorphism, this would indicate that lower mutation rate is the primary force responsible for lower rates of variation. Therefore, to further investigate whether the observed rates of polymorphism around gene bodies in 1,135 natural *A. thaliana* accessions were driven at least in part by mutation biases or only by selection, we calculated Tajima's *D*

(vcftools --TajimaD) in 100-bp windows across the entire genome and averaged these values in relation to TSSs and TTSs for every gene. We used bootstrapping ($n = 100$) to calculate the confidence interval (± 2 s.e.m.) around this mean value.

Tajima's D in exons

We used Tajima's D to estimate the extent to which mutation bias rather than selection after random mutation could explain differences in rates of natural polymorphism in exons (elevated polymorphism in peripheral exons). We calculated Tajima's D in every exon and grouped genes according to their total number of exons and plotted the average Tajima's D in relation to exons ordered from 5' to 3' ends. Tajima's D was consistently more negative in peripheral exons, reflecting the effects of increased population mutation rate in these loci, so we further investigated the underlying causes by testing whether genes with and without (and longer or shorter) UTRs have differences in Tajima's D in peripheral exons. Finally, we asked whether genes with more and longer introns have less negative Tajima's D values, to test whether the lower rates of polymorphism observed in these genes was caused at least in part by reduced mutation rate, rather than selection after random mutation.

Simulations of mutation bias and selection using SLiM

Our observation that Tajima's D is less negative in regions of low polymorphism, such as gene bodies, suggested that the reduced polymorphism therein is caused by a lower mutation rate, consistent with the mutation biases that we discovered in the analysed mutation datasets. To verify this interpretation, we conducted simulations using the software SLiM (v3)⁸³. These simulations modelled genic and intergenic space, based explicitly on the first 100 genes on chromosome 1. For each simulation, we modelled a population of 1,000 individuals for 10,000 generations. The selfing rate was assigned to 0.98, a low estimate based on field observations^{84,85}. The baseline mutation rate (per base and per generation) was derived from the empirically measured population mutation rate¹³ (from $N_e = \sim 300,000$, $u = \sim 1 \times 10^{-9}$ and adjusted for $N_e = 1,000$).

Recombination rate (probability per genome per generation) was 1×10^{-4} . To investigate the effects of mutation bias and selection, we assigned a scaled mutation rate in gene bodies of 0.2, 0.5 or 1, reflecting an 80%, 50% or 0% reduction relative to the baseline mutation rate in intergenic spaces. We also assigned proportions of deleterious mutations to be 0, 0.1 and 0.3, reflecting a 0%, 10% and 30% frequency of deleterious mutations independently in gene bodies and intergenic regions. All possible combinations of the three parameters were then simulated 200 times. Tajima's D was calculated across the entirety of each genome in 100-bp windows using VCFtools. The position of each window was calculated in relation to the TSSs and TTSs of each gene. Counts of polymorphisms and Tajima's D were averaged across all

genomes in 10-bp windows for regions 3 kb upstream and downstream of the TSS and TTS of each gene. The variation in polymorphism level and Tajima's D values were compared with the empirical observations of natural polymorphisms in 1,135 natural *A. thaliana* accessions⁶⁶ using Pearson correlation.

Relationship between mutation probability, epigenomic and other features, and breadth of expression across tissues

Because we found that essential genes have higher levels of epigenomic and other features than lower predicted mutation rates, we wanted to further test the hypothesis that essential housekeeping genes were also enriched for such features and therefore experience a subsequently lower probability of mutation and lower de novo mutation calls. We used gene expression data from 54 tissues³⁹. We calculated the correlation between the number of tissues with expression of more than 0 and either the predicted mutation probability score or the observed mutations for each gene. Because these results confirmed that genes expressed in more tissues have lower predicted mutation probability scores, we examined epigenetic features H3K4me1, H3K36me3 and CG methylation, which are enriched in essential genes, finding that genes expressed in all tissues were also enriched for these features.

Determining the effect of strong purifying selection on coding sequences

Our results had revealed significant biases in mutation probability in relation to gene bodies. Because we had found that mutations were significantly higher upstream of genes and significantly lower within gene bodies in five independent datasets, we considered the possibility that this overwhelming bias was the result of extremely strong purifying selection on de novo mutations (that is, removal of lethal mutations before they could be detected by us). We therefore simulated 10,000 random mutations across the TAIR10 genome. If mutations fell within coding regions, we randomly assigned them to be removed by selection (that is, dominant lethal). For this, we explored three levels of selection: $s = 0.01$ where 1% of mutations were removed (that is, had lethal effects), $s = 0.1$ where 10% of mutations were removed, $s = 0.2$ where 20% of mutations were removed, or $s = 0.3$ where 30% of mutations were removed. While $s = 0.3$ represents an exceptionally and unexpectedly high level of selection, especially in soma, evidenced by empirical estimates of the extent of gene essentiality in *A*.

thaliana, this served as a positive control for observing the effects of extraordinarily strong selection on the expected distribution of mutations in a random mutation model.

Comparing expected and observed levels of synonymous mutation

Because we had observed a significant reduction in mutation rate in coding regions, we wanted to test whether this was driven only by functionally impactful mutation (for example, amino acid substitutions). To do so, we simulated 6,182 random SNVs. For each variant, we asked whether it was found within the coding region of any gene. We counted the total number of coding region variants and multiplied this number with the expected fraction, 0.28, of synonymous variants based on *A. thaliana* codon usage and mutation spectrum. We iterated this simulation 100 times to produce a confidence interval of expected synonymous variants in our training set of de novo mutations.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

A complete table of called mutations is available in Supplementary Data 1. Genic feature (that is, upstream, UTR, intron, CDS, and so on) level data (mutation and epigenomic features) are available in Supplementary Data 2. Gene-level data (for example, mutation, epigenomic and other features, function, expression and selection) are available in Supplementary Data 3. Derived data objects used to create figures can be found as Source Data for individual figures, and additional intermediate data files are available on GitHub (https://github.com/greymonroe/mutation_bias_analysis). Raw mutation data used as our training set were deposited in Figshare (<https://doi.org/10.25386/genetics.6456065>). Previous raw Illumina sequencing reads from 64 *A. thaliana* leaves are available under NCBI SRA BioProject [PRJNA497989](#). Raw Illumina sequencing reads from additional

mutation accumulation experiments (European lines) are available under NCBI SRA BioProject [PRJNA770533](#). Raw reads from the ATAC-seq experiments are available under ENA Project [PRJEB48038](#). Raw reads from resequencing MA73 are available under ENA Project [PRJEB48100](#). Variant data of natural *A. thaliana* accessions are available at <http://1001genomes.org/data/GMI-MPI/releases/v3.1/>. The TAIR10 reference genome and annotation are available at www.arabidopsis.org. The *P. trichocarpa* reference genome, annotation and variant data are available at https://phytozome-next.jgi.doe.gov/info/Ptrichocarpa_v3_1. Chromatin state data are available through the Plant Chromatin State Database (<http://systemsbiology.cau.edu.cn/chromstates>). Tissue-specific expression data are available at <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-7978/>. There are no restrictions on the availability of data used in this study.

Code availability

Functions to characterize Tajima's *D* and polymorphisms in relation to TSSs and TTSs are available on GitHub (<https://github.com/greymonroe/polymorphology>). The annotated code for models and statistical analyses is available on GitHub (https://github.com/greymonroe/mutation_bias_analysis).

References

1. 1.
Futuyma, D. J. *Evolutionary Biology* 2nd edn (Sinauer, 1986).
2. 2.
Martincorena, I. & Luscombe, N. M. Non-random mutation: the evolution of targeted hypermutation and hypomutation. *Bioessays* **35**, 123–130 (2013).
3. 3.

Lynch, M. et al. Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* **17**, 704–714 (2016).

4. 4.

Stoletzki, N. & Eyre-Walker, A. The positive correlation between dN/dS and dS in mammals is due to runs of adjacent substitutions. *Mol. Biol. Evol.* **28**, 1371–1380 (2011).

5. 5.

Hodgkinson, A. & Eyre-Walker, A. Variation in the mutation rate across mammalian genomes. *Nat. Rev. Genet.* **12**, 756–766 (2011).

6. 6.

Chen, X. & Zhang, J. No gene-specific optimization of mutation rate in *Escherichia coli*. *Mol. Biol. Evol.* **30**, 1559–1562 (2013).

7. 7.

Li, C. & Luscombe, N. M. Nucleosome positioning stability is a modulator of germline mutation rate variation across the human genome. *Nat. Commun.* **11**, 1363 (2020).

8. 8.

Li, F. et al. The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutS α . *Cell* **153**, 590–600 (2013).

9. 9.

Schuster-Böckler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507 (2012).

10. 10.

Xia, B. et al. Widespread transcriptional scanning in the testis modulates gene evolution rates. *Cell* **180**, 248–262.e21 (2020).

11. 11.

Chen, X. et al. Nucleosomes suppress spontaneous mutations base-specifically in eukaryotes. *Science* **335**, 1235–1238 (2012).

12. 12.

Weng, M.-L. et al. Fine-grained analysis of spontaneous mutation spectrum and frequency in *Arabidopsis thaliana*. *Genetics* **211**, 703–714 (2019).

13. 13.

Ossowski, S. et al. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**, 92–94 (2010).

14. 14.

Liu, Y. et al. PCSD: a plant chromatin state database. *Nucleic Acids Res.* **46**, D1157–D1167 (2018).

15. 15.

Martincorena, I., Seshasayee, A. S. N. & Luscombe, N. M. Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature* **485**, 95–98 (2012).

16. 16.

Supek, F. & Lehner, B. Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. *Cell* **170**, 534–547.e23 (2017).

17. 17.

Frigola, J. et al. Reduced mutation rate in exons due to differential mismatch repair. *Nat. Genet.* **49**, 1684–1692 (2017).

18. 18.

Belfield, E. J. et al. DNA mismatch repair preferentially protects genes from mutation. *Genome Res.* **28**, 66–74 (2018).

19. 19.

Huang, Y., Gu, L. & Li, G.-M. H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation. *J. Biol. Chem.* **293**, 7811–7823 (2018).

20. 20.

Moore, L. et al. The mutational landscape of human somatic and germline cells. *Nature* **597**, 381–386 (2021).

21. 21.

Wang, L. et al. The architecture of intra-organism mutation rate variation in plants. *PLoS Biol.* **17**, e3000191 (2019).

22. 22.

Bobiwash, K., Schultz, S. T. & Schoen, D. J. Somatic deleterious mutation rate in a woody plant: estimation from phenotypic data. *Heredity* **111**, 338–344 (2013).

23. 23.

Wolfe, K. H., Sharp, P. M. & Li, W. H. Mutation rates differ among regions of the mammalian genome. *Nature* **337**, 283–285 (1989).

24. 24.

Long, H. et al. Evolutionary determinants of genome-wide nucleotide composition. *Nat. Ecol. Evol.* **2**, 237–240 (2018).

25. 25.

Fryxell, K. J. & Zuckerkandl, E. Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Mol. Biol. Evol.* **17**, 1371–1383 (2000).

26. 26.

Shilatifard, A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* **81**, 65–95 (2012).

27. 27.

Herbette, M. et al. The *C. elegans* SET-2/SET1 histone H3 Lys4 (H3K4) methyltransferase preserves genome stability in the germline. *DNA Repair* **57**, 139–150 (2017).

28. 28.

Chong, S. Y. et al. H3K4 methylation at active genes mitigates transcription-replication conflicts during replication stress. *Nat. Commun.* **11**, 809 (2020).

29. 29.

Lim, B., Mun, J., Kim, Y. S. & Kim, S.-Y. Variability in chromatin architecture and associated DNA repair at genomic positions containing somatic mutations. *Cancer Res.* **77**, 2822–2833 (2017).

30. 30.

Zheng, C. L. et al. Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes. *Cell Rep.* **9**, 1228–1234 (2014).

31. 31.

Ha, K., Kim, H.-G. & Lee, H. Chromatin marks shape mutation landscape at early stage of cancer progression. *NPJ Genom. Med.* **2**, 9 (2017).

32. 32.

Exposito-Alonso, M. et al. The rate and potential relevance of new mutations in a colonizing plant lineage. *PLoS Genet.* **14**, e1007155 (2018).

33. 33.

Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**, 264–267 (2016).

34. 34.

Haudry, A. et al. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat. Genet.* **45**, 891–898 (2013).

35. 35.

1001 Genomes Consortium. 1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 481–491 (2016).

36. 36.

Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).

37. 37.

Lloyd, J. P., Seddon, A. E., Moghe, G. D., Simenc, M. C. & Shiu, S.-H. Characteristics of plant essential genes allow for within- and between-species prediction of lethal mutant phenotypes. *Plant Cell* **27**, 2133–2147 (2015).

38. 38.

Lloyd, J. & Meinke, D. A comprehensive dataset of genes with a loss-of-function mutant phenotype in *Arabidopsis*. *Plant Physiol.* **158**, 1115–1129 (2012).

39. 39.

Mergner, J. et al. Mass-spectrometry-based draft of the *Arabidopsis* proteome. *Nature* **579**, 409–414 (2020).

40. 40.

Lynch, M. Evolution of the mutation rate. *Trends Genet.* **26**, 345–352 (2010).

41. 41.

Koonin, E. V. *The Logic of Chance: The Nature and Origin of Biological Evolution* (FT Press, 2011).

42. 42.

Johri, P., Charlesworth, B. & Jensen, J. D. Toward an evolutionarily appropriate null model: jointly inferring demography and purifying selection. *Genetics* **215**, 173–192 (2020).

43. 43.

Shaw, F. H., Geyer, C. J. & Shaw, R. G. A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*. *Evolution* **56**, 453–463 (2002).

44. 44.

Keightley, P. D. & Lynch, M. Toward a realistic model of mutations affecting fitness. *Evolution* **57**, 683–685 (2003).

45. 45.

Belfield, E. J. et al. Thermal stress accelerates *Arabidopsis thaliana* mutation rate. *Genome Res.* **31**, 40–50 (2021).

46. 46.

Watson, J. M. et al. Germline replications and somatic mutation accumulation are independent of vegetative life span in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **113**, 12226–12231 (2016).

47. 47.

Pockrandt, C., Alzamel, M., Iliopoulos, C. S. & Reinert, K. GenMap: ultra-fast computation of genome mappability. *Bioinformatics* **36**, 3687–3692 (2020).

48. 48.

Huang, W., Li, L., Myers, J. R. & Marth, G. T. ART: a next-generation sequencing read simulator. *Bioinformatics* **28**, 593–594 (2012).

49. 49.

Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790–2791 (2013).

50. 50.

Weng, M.-L. et al. Fitness effects of mutation in natural populations of *Arabidopsis thaliana* reveal a complex influence of local adaptation. *Evolution* **75**, 330–348 (2021).

51. 51.

Huang, Y. & Li, G.-M. DNA mismatch repair preferentially safeguards actively transcribed genes. *DNA Repair* **71**, 82–86 (2018).

52. 52.

Wang, Y. et al. Histone H3 lysine 14 acetylation is required for activation of a DNA damage checkpoint in fission yeast. *J. Biol. Chem.* **287**, 4386–4393 (2012).

53. 53.

Yazdi, P. G. et al. Increasing nucleosome occupancy is correlated with an increasing mutation rate so long as DNA repair machinery is intact. *PLoS ONE* **10**, e0136574 (2015).

54. 54.

Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **109**, E2774–E2783 (2012).

55. 55.

Supek, F. & Lehner, B. Scales and mechanisms of somatic mutation rate variation across the human genome. *DNA Repair* **81**, 102647 (2019).

56. 56.

Fryxell, K. J. & Moon, W.-J. CpG mutation rates in the human genome are highly dependent on local GC content. *Mol. Biol. Evol.* **22**, 650–658 (2005).

57. 57.

Elango, N., Kim, S.-H., Vigoda, E. & Yi, S. V. Mutations of different molecular origins exhibit contrasting patterns of regional substitution rate variation. *PLoS Comput. Biol.* **4**, e1000015 (2008).

58. 58.

Hodgkinson, A. & Eyre-Walker, A. The genomic distribution and local context of coincident SNPs in human and chimpanzee. *Genome Biol.*

Evol. **2**, 547–557 (2010).

59. 59.

Arndt, P. F., Hwa, T. & Petrov, D. A. Substantial regional variation in substitution rates in the human genome: importance of GC content, gene density, and telomere-specific effects. *J. Mol. Evol.* **60**, 748–763 (2005).

60. 60.

Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* **10**, 285–311 (2009).

61. 61.

Mugal, C. F. & Ellegren, H. Substitution rate variation at human CpG sites correlates with non-CpG divergence, methylation level and GC content. *Genome Biol.* **12**, R58 (2011).

62. 62.

Youk, J., An, Y., Park, S., Lee, J.-K. & Ju, Y. S. The genome-wide landscape of C:G > T:A polymorphism at the CpG contexts in the human population. *BMC Genomics* **21**, 270 (2020).

63. 63.

Polak, P. et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518**, 360–364 (2015).

64. 64.

Hung, S. et al. Mismatch repair-signature mutations activate gene enhancers across human colorectal cancer epigenomes. *Elife* **8**, e40760 (2019).

65. 65.

Fang, J. et al. Cancer-driving H3G34V/R/D mutations block H3K36 methylation and H3K36me3-MutS α interaction. *Proc. Natl Acad. Sci. USA* **115**, 9598–9603 (2018).

66. 66.

Salzberg, A. C. et al. Genome-wide mapping of histone H3K9me2 in acute myeloid leukemia reveals large chromosomal domains associated with massive gene silencing and sites of genome instability. *PLoS ONE* **12**, e0173723 (2017).

67. 67.

Supek, F. & Lehner, B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**, 81–84 (2015).

68. 68.

Heredia-Genestar, J. M., Marquès-Bonet, T., Juan, D. & Navarro, A. Extreme differences between human germline and tumor mutation densities are driven by ancestral human-specific deviations. *Nat. Commun.* **11**, 2512 (2020).

69. 69.

Quadrana, L. et al. Transposition favors the generation of large effect mutations that may facilitate rapid adaption. *Nat. Commun.* **10**, 3421 (2019).

70. 70.

Choi, J., Lyons, D. B., Kim, M. Y., Moore, J. D. & Zilberman, D. DNA methylation and histone H1 jointly repress transposable elements and aberrant intragenic transcripts. *Mol. Cell* **77**, 310–323.e7 (2020).

71. 71.

Kawakatsu, T. et al. Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* **166**, 492–505 (2016).

72. 72.

Halldorsson, B. V. et al. Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science* **363**, eaau1043 (2019).

73. 73.

O’Hara, R. & Kotze, J. Do not log-transform count data. *Nat. Prec.* <https://doi.org/10.1038/npre.2010.4136.1> (2010).

74. 74.

Knief, U. & Forstmeier, W. Violating the normality assumption may be the lesser of two evils. *Behav. Res. Methods* **53**, 2576–2590 (2021).

75. 75.

Evans, L. M. et al. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nat. Genet.* **46**, 1089–1096 (2014).

76. 76.

McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).

77. 77.

Rand, D. M. & Kann, L. M. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* **13**, 735–748 (1996).

78. 78.

Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618 (2007).

79. 79.

Monroe, G. et al. Drought adaptation in *Arabidopsis thaliana* by extensive genetic loss-of-function. *Elife* **7**, e41038 (2018).

80. 80.

Baggs, E. et al. Convergent loss of an EDS1/PAD4 signaling pathway in several plant lineages reveals co-evolved components of plant immunity and drought response. *Plant Cell* **32**, 2158–2177 (2020).

81. 81.

Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).

82. 82.

Henderson, I. R., Liu, F., Drea, S., Simpson, G. G. & Dean, C. An allelic series reveals essential roles for FY in plant development in addition to flowering-time control. *Development* **132**, 3597–3607 (2005).

83. 83.

Haller, B. C. & Messer, P. W. SLiM 3: forward genetic simulations beyond the Wright–Fisher model. *Mol. Biol. Evol.* **36**, 632–637 (2019).

84. 84.

Bomblies, K. et al. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genet.* **6**, e1000890 (2010).

85. 85.

Platt, A. et al. The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet.* **6**, e1000843 (2010).

86. 86.

Cao, J. et al. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* **43**, 956–963 (2011).

87. 87.

Gossmann, T. I. et al. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol. Biol. Evol.* **27**, 1822–1832 (2010).

88. 88.

Moore, R. C. & Purugganan, M. D. The early stages of duplicate gene evolution. *Proc. Natl Acad. Sci. USA* **100**, 15682–15687 (2003).

Acknowledgements

We thank members of the Weigel laboratory and the broader community for comments on earlier versions of this manuscript, especially A. Britt, P. Flood, K. Krasileva, M. Lynch, D. Petrov, J. Ross-Ibarra, D. Runcie, B. Schmitz and D. Sloan. This work was supported by NSF grants DEB 0844820 and DEB 1257902 (to C.B.F.), NSF DEB 0845413 and DEB 1258053 (to M.T.R.), the UC Davis Department of Plant Sciences (to J.G.M.), and by DFG grant ERA-CAPS 1001G+ and the Max Planck Society (to D.W.).

Funding

Open access funding provided by Max Planck Society.

Author information

Author notes

1. Claude Becker

Present address: Faculty of Biology, Ludwig Maximilian University, Martinsried, Germany

Affiliations

1. Department of Molecular Biology, Max Planck Institute for Biology Tübingen, Tübingen, Germany

J. Grey Monroe, Thanvi Srikant, Pablo Carbonell-Bejerano, Claude Becker, Marie Klein, Julia Hildebrandt, Manuela Neumann & Detlef Weigel

2. Department of Plant Sciences, University of California Davis, Davis, CA, USA

J. Grey Monroe, Mariele Lensink, Marie Klein & Daniel Kliebenstein

3. Department of Plant Biology, Carnegie Institution for Science, Stanford, CA, USA

Moises Exposito-Alonso

4. Department of Biology, Stanford University, Stanford, CA, USA

Moises Exposito-Alonso

5. Department of Biology, Westfield State University, Westfield, MA, USA

Mao-Lun Weng

6. ISEM, University of Montpellier, Montpellier, France

Eric Imbert

7. Department of Ecology and Genetics, EBC, Uppsala University,
Uppsala, Sweden

Jon Ågren

8. Department of Biology, College of Charleston, Charleston, SC, USA

Matthew T. Rutter

9. Oak Lake Field Station, South Dakota State University, Brookings,
SD, USA

Charles B. Fenster

Contributions

All authors contributed to the work presented in this paper. J.G.M. and D.W. conceived the project. M.-L.W., M.T.R. and C.B.F. conducted the mutation accumulation experiments with material from C.B., J.A. and E.I. These lines were sequenced by J.H., M.N. and C.B. T.S. performed ATAC-seq. M.K. and P.C.-B. generated deep resequencing data for siblings. Data analyses were led by J.G.M. with M.-L.W., T.S. and P.C.-B., with major additional contributions from M.L., M.E.-A., M.K., D.K., D.W., C.B.F. and D.W. J.G.M. and D.W. wrote the manuscript with major contributions from T.S., P.C.-B., C.B., M.L., M.E.-A., M.K., D.K., M.-L.W., J.A., M.T.R. and C.B.F. All authors read and provided feedback on the manuscript.

Interpretation of data and results were led by J.G.M. and D.W., with major contributions from T.S., P.C.-B., C.B., M.L., M.E.-A., M.K., D.K., M.-L.W., J.A., M.T.R. and C.B.F.

Corresponding authors

Correspondence to [J. Grey Monroe](#) or [Detlef Weigel](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Olivier Tenaillon, Michael Purugganan 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 Workflow and quality control of *de novo* mutation identification.](#)

a, Filtering pipeline. **b**, High-quality *de novo* mutations called in this study on the original mutation accumulation experiment data (107 replicate lineages of Col-0, total number = mutations from this study plus ref. [12](#)). **c**, Visualization of raw data and properties of high-confidence set. **d**, Estimated probability of mutation calls being erroneous based on alternative and total read depths in the high-confidence set and variants removed by filtering.

[Extended Data Fig. 2 Summary of observed *de novo* mutations and distribution across original Col-0 mutation accumulation \(MA\) lines.](#)

a, *De novo* mutations detected in genic regions (genes \pm 1,000 bp) in individual MA lines. SNVs in light blue, InDels in magenta. Our investigation was focused on mutations in and around genes, so for clarity mutations elsewhere (i.e., near centromeres) are not shown. **b**, Distribution of mutations across genic regions per 2 Mb windows. Vertical black lines in

the lower plot mark the location of genes. **c**, Mutation rates in lethal- and non-lethal-effect genes ($n = 27,206$ genes, mean ± 2 s.e.m., two-sided t-test) **d**, Frequencies of single nucleotide transitions and transversions. **e**, Distribution of frequency of specific mutations across lines. **f**, Number of germline and somatic mutations detected in each MA line. **g**, Distribution of alternative allele read depth for putative somatic mutations. **h**, Relationship between number of detected mutations and total sequencing depth (total number of informative reads in variant sites) in MA lines. **i**, Size distribution of insertions and deletions.

Extended Data Fig. 3 Sequencing depth, mappability, and false positives do not explain observed biases in distributions of natural polymorphisms or observed mutations used to predict mutation probabilities.

a, Sequencing depth around transcription start (TSS) and termination (TTS) sites in one randomly chosen mutation accumulation line. **b**, Mappability around TSS TTS site calculated with GenMap⁴⁷. **c**, Rates of false positive SNP and InDel calls around TSS and TTS determined from 1,000 iterations of simulated Illumina reads. **d**, Simulation of effect of selection on gene bodies. Selection could take the form of mutations being dominant lethal or through somatic competition of mutations with small selection coefficients. $0 = 0\%$, $0.01 = 1\%$, $0.1 = 10\%$, $0.2 = 20\%$, $0.3 = 30\%$ of gene body mutations removed by purifying selection. 30% is estimated to be the approximate upper bound of constrained sites in gene bodies³⁴. **e–i**, Resequencing of 10 siblings of one MA line from ref. ¹². **e**, Overview of experimental design for testing the effect of sequencing depth on calling somatic mutations. **f**, Filtered heterozygous variants (SNVs and InDels) called in sibling #5 sequenced at ~600x depth overlap more with variants called from sibling #5 at ~60x sequencing depth than with other siblings at ~60 sequencing depth. The boxplots show the distribution of 20 iterations of sampling equal numbers of heterozygous variants (to account for differences in total number of variants called in different siblings) for each sibling sequenced at ~60x and compared to sibling 5 sequenced at ~600x. Boxplots show median with maxima and minima reflecting interquartile range (IQR), whiskers = $1.5 * \text{IQR}$ ($n = 20$ iterations). **g**, Frequency

distribution of unfiltered heterozygous variants called in 10 siblings sequenced at ~60x depth each. Note that because these siblings are descendants of 25 generations of self-fertilization, the number of true heterozygous (inherited segregating) calls is expected to be very small compared to heterozygous variants that are chimeric somatic mutations. **h**, Average mappability of variants detected in different numbers of siblings out of the 10 sequenced siblings. **i**, Variants called independently in one sibling or, less so, in two to four siblings show signatures of mutation bias. In contrast, variants called in five or more siblings (which should more likely be false positives due to cryptic duplications or regions with poor mappability) do not show a biased distribution around TSS and TTS, with overall distribution similar to known false positives.

Extended Data Fig. 4 Variants called in additional mutation accumulation datasets.

a, Germline and somatic mutations around gene regions in mutation accumulation lines derived from eight founder genotypes. For each founder, 35-60 lines were propagated for 8-10 generations. **b**, The proportion of somatic variants detected in gene bodies (gene body/(gene body + upstream + downstream)) among descendants of the same founder. F- and p-values from one-way ANOVA. ($n = 400$ unique mutation accumulation lines, mean ± 2 s.e.m.) **c**, Somatic variants detected from reanalysis of 64 individual leaves from two Col-0 plants²¹. **d**, Germline variants detected in a bottlenecked *A. thaliana* lineage following colonization of North America since ~1600 (ref. ³²). **e**, Epigenome-predicted and observed mutation rates across all datasets for InDels and SNVs, comparing gene bodies (GB) with upstream/downstream (U/D) regions. P-values from chi-squared tests.

Extended Data Fig. 5 Relationships between epigenome-predicted mutation probability, observed *de novo* mutations, polymorphisms in natural populations, and Tajima's D in natural populations.

These data show the quantitative relationships apparent in Fig. 2 of the main text. Each point reflects the value in one window of 1,200 calculated

windows across all 33,056 genes, in relation to genome-wide transcription start and termination sites (TSS, TTS). Error bars indicate ± 2 s.e.m. confidence intervals. For epigenome-predicted mutation probability scores, each point reflects the mean ± 2 s.e.m. across all genes. For observed *de novo* mutations, each point reflects the total number of mutations ± 2 s.e.m. (bootstrapped). For polymorphisms, each point reflects the total number of variants ± 2 s.e.m. (bootstrapped). For Tajima's D, each point reflects the mean ± 2 s.e.m. Tajima's D is already predicted by existing theory to be negatively correlated with mutation rate, as regions with higher mutation rate will be enriched for newer and therefore rarer variants.

Extended Data Fig. 6 Effects of mutation rate and selection heterogeneity on polymorphisms and Tajima's D along genes.

Simulation results from SLiM⁸³ for the first 100 genes of chromosome 1 using a population of 1,000 individuals and 10,000 generations. **a–c**, Average correlation between 200 permutations of simulated scenarios and observed patterns of variation in natural *A. thaliana* accessions. **a**, Parameter choice: Difference between mutation rate (dm) in gene bodies and intergenic space (e.g., 0.5 = 50% reduction in mutation rate) and proportion of mutations that are deleterious in the genic (gds) and intergenic (ids) regions. The parameter combinations shown in **d–h** are highlighted with red outlines. **b**, Pearson correlation coefficients (ppc) comparing Tajima's D values from each simulation to that of observed data in natural *A. thaliana* accession. **c**, Pearson correlation coefficients (pcc) comparing number of polymorphisms accumulated in each simulation to that of observed data in wild *Arabidopsis* accessions. **d–k**, Examples of polymorphism (red) and Tajima's D (blue) in relation to gene bodies (TSS, TTS) averaged from 200 permutations of a scenario approximating empirical estimates of mutation rate heterogeneity and selection heterogeneity between gene bodies and intergenic space. Parameters (see **a**) given for each scenario. Strong purifying selection in gene regions alone (with equal mutation rates between gene bodies and intergenic space), which also reduces levels of polymorphism in gene bodies, causes more negative Tajima's D values in gene bodies, which is inconsistent with observed data in natural *A. thaliana* accessions.

Extended Data Fig. 7 Relationships between untranslated regions or introns and mutation rates.

a, Distribution of epigenomic features in genes with different numbers of exons. **b**, Epigenome-predicted Mutation Probability Score (MPS), rates of natural polymorphism, and Tajima’s D in genes with different numbers of exons. **c**, Left: comparison of Mutation Probability Score (MPS) between genes with UTRs and those lacking 5’ or 3’ UTRs. Horizontal lines mark the mean difference between genes with and without UTRs. Vertical lines mark the mean ± confidence intervals of two-sided t-tests. Center: rates of natural polymorphism in natural *A. thaliana* accessions. Right: Tajima’s D in natural accessions. (n = 35,526 gene models). **d**, Left: Pearson’s correlation coefficients for relationship between predicted mutation probabilities and the absolute length of 5’ and 3’ UTRs. Horizontal lines mark the means, and vertical lines mark the mean ± confidence intervals. Center: same for rates of natural polymorphism in natural accessions. Right: same for Tajima’s D in natural accessions. (n = 35,526 gene models). **e**, Left: Relationships between intron number and total intron length with predicted mutation probabilities. Points indicate mean values. Center: same for rates of natural polymorphism in natural accessions. Right: same for Tajima’s D in natural accessions. **f**, Results for 544 *Populus trichocarpa* accessions⁷⁵. Horizontal lines mark the means, and vertical lines mark the mean ± confidence intervals of two-sided t-tests (n = 73,013 gene models).

Extended Data Fig. 8 Epigenomic and other features and mutation rates of lethal-effect and constitutively expressed genes.

a, Enrichment of features in coding sequences of “lethal-effect” genes (n = 2,720 lethal-effect genes, mean ± 2 s.e.m.). **b**, Total mutation rate (±2 s.e.m., bootstrapped) in lethal- and non-lethal-effect genes (n = 27,206 genes). **c**, Enrichment of features in coding sequences of constitutively (across all tissues) expressed genes (n = 9,957 genes, mean ± 2 s.e.m.). **d**, Total mutation rate (±2 s.e.m., bootstrapped) in genes binned according to the number of tissues in which they are expressed (n = 25,987 genes with tissue-specific expression data)³⁹.

Extended Data Fig. 9 Predicted mutation rates and evidence of selection on natural polymorphisms versus *de novo* mutations across gene regions.

a, Relationship between Tajima's D of gene bodies and coding region selection estimated by P_n/P_s ($n = 21,407$ genes) and **b**, D_n/D_s ($n = 21,407$ genes). **c**, Epigenome-predicted mutation probability in different gene features. **d**, Scaled residuals $((\text{Obs}-\text{Pred})/\text{Pred})$ from $S \sim u$. Significantly negative residuals in coding regions are consistent with purifying selection in natural populations acting on new mutations. **e**, Relationships between epigenome-predicted mutation probability and other estimates of constraint. Residuals between predicted mutation rate and observed mutations are positively correlated with predicted mutation rate indicating that genes subject to purifying selection are predicted to mutate less. Genes with low predicted mutation rates are also less likely to have $\alpha > 0$, a measure of variants under positive selection. Genes with low predicted mutation rate are depleted in natural populations for non-synonymous variants that reach fixation, as measured by the Neutrality Index, and for loss-of-function variants.

Extended Data Fig. 10 Estimates of L_{segment} for different regions.

a, Length of sequence space (L_{segment}) reflecting different types of regions. **b**, Test of parameter space that satisfies population genetic theoretical predictions for the possibility for targeted hypomutation to evolve. OOM = orders of magnitude. Selection on intragenomic mutation rate variation will be effective^{5,40} when $N_e * u * s * du * pd * L_{\text{segment}} > 1$ where N_e is the effective population size, u is the mutation rate, s is the average selection coefficient on deleterious mutations, du is the degree of change in mutation rate, pd is the proportion of sites subject to purifying selection, and L_{segment} is the region of the genome affected. Assuming an effective population size of $\sim 300,000$ (ref. 86,87,88), a mutation rate of $\sim 10^{-8}$ (ref. 13), an average selection coefficient of 0.01 (ref. 5), an order-of-magnitude reduction in mutation rate⁵, and functionally constrained regions where 20% of sites are

under selection⁵, the total length of the sequence affected, $L_{segment}$, would have to be at least ~200 kb, which (accounting for differences in effective population size) is similar to previous estimates in humans⁵. For perspective, this minimum $L_{segment}$ is considerably shorter (~1.5%) than the sum of coding regions with elevated levels of H3K4me1 (top quartile is ~13 Mb, or 15% of the genome), a feature enriched in gene bodies and essential genes and associated with lower mutation rate. Thus, selection is expected to act with high efficiency on variants that cause DNA repair and protection mechanisms to preferentially target such regions.

Supplementary information

[Reporting Summary](#)

[Peer Review File](#)

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

Monroe, J.G., Srikant, T., Carbonell-Bejerano, P. *et al.* Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature* **602**, 101–105 (2022). <https://doi.org/10.1038/s41586-021-04269-6>

- Received: 09 November 2020
- Accepted: 17 November 2021
- Published: 12 January 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04269-6>

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

Further reading

- [Tuning mutagenesis by functional outcome](#)

- Darren J. Burgess

Nature Reviews Genetics (2022)

- Important genomic regions mutate less often than do other regions

- Jianzhi Zhang

Nature (2022)

Important genomic regions mutate less often than do other regions

- Jianzhi Zhang

News & Views 12 Jan 2022

Why mutation is not as random as we thought

- Shamini Bundell
- Benjamin Thompson

Nature Podcast 19 Jan 2022

Tuning mutagenesis by functional outcome

- Darren J. Burgess

Research Highlight 01 Feb 2022

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04269-6>

- Article
- Open Access
- [Published: 09 December 2021](#)

Malaria protection due to sickle haemoglobin depends on parasite genotype

- [Gavin Band](#) [ORCID: orcid.org/0000-0002-1710-9024](#)^{1,2,3},
- [Ellen M. Leffler](#) [ORCID: orcid.org/0000-0002-1614-9366](#)^{2,4},
- [Muminatou Jallow](#)^{5,6},
- [Fatoumatta Sisay-Joof](#)⁵,
- [Carolyne M. Ndila](#)⁷,
- [Alexander W. Macharia](#) [ORCID: orcid.org/0000-0001-5364-9610](#)⁷,
- [Christina Hubbart](#) [ORCID: orcid.org/0000-0001-9576-9581](#)¹,
- [Anna E. Jeffreys](#)¹,
- [Kate Rowlands](#)¹,
- [Thuy Nguyen](#)²,
- [Sónia Gonçalves](#)²,
- [Cristina V. Ariani](#)²,
- [Jim Stalker](#)²,
- [Richard D. Pearson](#) [ORCID: orcid.org/0000-0002-7386-3566](#)^{2,3},
- [Roberto Amato](#)²,
- [Eleanor Drury](#) [ORCID: orcid.org/0000-0002-9887-6961](#)²,
- [Giorgio Sirugo](#)^{5,8},
- [Umberto d'Alessandro](#)⁵,
- [Kalifa A. Bojang](#)⁵,
- [Kevin Marsh](#) [ORCID: orcid.org/0000-0001-8377-5466](#)^{7,9},
- [Norbert Peshu](#)⁷,
- [Joseph W. Saelens](#)¹⁰,
- [Mahamadou Diakité](#)¹¹,
- [Steve M. Taylor](#)^{10,12},
- [David J. Conway](#) [ORCID: orcid.org/0000-0002-8711-3037](#)^{5,13},
- [Thomas N. Williams](#) [ORCID: orcid.org/0000-0003-4456-2382](#)^{7,14},

- [Kirk A. Rockett](#) ^{1,2} &
- [Dominic P. Kwiatkowski](#) [ORCID: orcid.org/0000-0002-5023-0176](#)^{1,2,3}

[Nature](#) volume **602**, pages 106–111 (2022)

- 9879 Accesses
- 1 Citations
- 137 Altmetric
- [Metrics details](#)

Subjects

- [Evolutionary genetics](#)
- [Genetic association study](#)
- [Genetic interaction](#)
- [Malaria](#)
- [Parasitic infection](#)

Abstract

Host genetic factors can confer resistance against malaria¹, raising the question of whether this has led to evolutionary adaptation of parasite populations. Here we searched for association between candidate host and parasite genetic variants in 3,346 Gambian and Kenyan children with severe malaria caused by *Plasmodium falciparum*. We identified a strong association between sickle haemoglobin (HbS) in the host and three regions of the parasite genome, which is not explained by population structure or other covariates, and which is replicated in additional samples. The HbS-associated alleles include nonsynonymous variants in the gene for the acyl-CoA synthetase family member^{2,3,4} *PfACS8* on chromosome 2, in a second region of chromosome 2, and in a region containing structural variation on chromosome 11. The alleles are in strong linkage disequilibrium and have frequencies that covary with the frequency of HbS across populations, in particular being much more common in Africa than other parts of the world. The estimated protective effect of HbS against severe malaria, as determined by comparison of cases with population controls, varies greatly according to the parasite genotype at these three loci. These findings open up a new avenue of enquiry into the biological and epidemiological significance of the HbS-associated polymorphisms in the parasite genome and the evolutionary forces that have led to their high frequency and strong linkage disequilibrium in African *P. falciparum* populations.

[Download PDF](#)

Main

Malaria can be viewed as an evolutionary arms race between the host and parasite populations. Human populations in Africa have acquired a high frequency of HbS and other erythrocyte polymorphisms that provide protection against the severe symptoms of *P. falciparum* infection^{1,5}, while *P. falciparum* populations have evolved a complex repertoire of genetic variation to evade the human immune system and to resist antimalarial drugs^{6,7}. This raises the basic question: are there genetic forms of *P. falciparum* that can overcome the human variants that confer resistance to this parasite?

To address this question, we analysed both host and parasite genome variation in samples from 5,096 children from Gambia and Kenya with severe malaria caused by *P. falciparum* (Extended Data Fig. 1, Supplementary Fig. 1, Methods). The samples were collected over the period 1995–2009 as part of a genome-wide association study (GWAS) of human resistance to severe malaria^{5,8,9}. In brief, we sequenced the *P. falciparum* genome using the Illumina X Ten platform using two approaches based on sequencing whole DNA and selective whole-genome amplification¹⁰. We used an established pipeline¹¹ to identify and call genotypes at more than two million single nucleotide polymorphisms (SNPs) and short insertion and deletion variants across the *P. falciparum* genome in these samples (Methods), although the majority of these occurred at low frequency. Our analysis is based on the 4,171 samples that had high quality data for both parasite and human genotypes, of which a subset of 3,346 had human genome-wide genotyping available and were used for discovery analysis. We focused on a set of 51,225 biallelic variants in the *P. falciparum* genome that passed all quality control filters and were observed in at least 25 infections in this subset (Methods). Our analyses exclude mixed-genotype calls that arise in malaria when a host is infected with multiple parasite lineages. Full details of our sequencing and data processing can be found in [Supplementary Methods](#).

We used a logistic-regression approach to test for pairwise association between these *P. falciparum* variants and four categories of human variants that are plausibly associated with malaria resistance: (1) known autosomal protective mutations, including HbS (in *HBB*), the common mutation that determines the O blood group (in *ABO*), regulatory variation associated with protection at *ATP2B4*^{5,8,12} and the structural variant DUP4, which encodes the Dantu blood-group phenotype¹³; (2) variants that showed suggestive but not conclusive evidence for association with severe malaria in our previous GWAS⁸; (3) human leukocyte antigen (HLA) alleles and additional glycophorin structural variants that we previously imputed in these

samples^{8,13}; and (4) variants near genes that encode human blood-group antigens, which we tested against the subset of *P. falciparum* variants lying near genes that encode proteins important for the merozoite stage^{14,15}, as these might conceivably interact during host cell invasion by the parasite. Although several factors could confound this analysis in principle—notably, if there was incidental association between human and parasite population structure—the distribution of test statistics suggested that our test was not affected by systematic confounding after including only an indicator of country as a covariate (Supplementary Fig. 2), and we used this approach for our main analysis. The full set of results is summarized in Fig. 1a, Supplementary Table 1.

Fig. 1: Three regions of the *P. falciparum* genome are associated with HbS.

 figure 1



a, Points show the evidence for association between each *P. falciparum* variant and human genotypes (top row) or between each included human variant and *P. falciparum* genotypes (bottom row). Association evidence is summarized by averaging the evidence for pairwise association (Bayes factor (BF) for test in $n = 3,346$ samples) between each variant (points) and all variants in the other organism against which it was tested ($\log_{10}(\text{BF}_{\text{avg}})$). *P. falciparum* variants are shown grouped by chromosome, and human variants are grouped by inclusion category as described in text and

Methods. Dashed lines and variant annotations reflect pairwise tests with $BF > 10^6$; only the top signal in each association region pair is annotated (Methods). **b**, Detail of the association with HbS in the *Pfsa1*, *Pfsa2* and *Pfsa3* regions of the *P. falciparum* genome. Points show evidence for association with HbS ($\log_{10} (BF_{HbS})$) for each regional variant. Variants that alter protein coding sequence are denoted by plus, and other variants are denoted by circles. Results are computed by logistic regression including an indicator of country as a covariate and assuming an additive model of association, with HbS genotypes based on imputation from genome-wide genotypes as previously described⁸. Mixed and missing *P. falciparum* genotype calls were excluded from the computation. Below, regional genes are annotated, with gene symbols given where the gene has an ascribed name in the PlasmoDB annotation (after removing ‘PF3D7_’ from the name where relevant); the three genes containing the most-associated variants are shown in red. A corresponding plot using directly typed HbS genotypes is presented in Extended Data Fig. 2.

Three *P. falciparum* loci are associated with HbS

The most prominent finding to arise from this joint analysis of host and parasite variation was a strong association between the sickle haemoglobin allele HbS and three separate regions in the *P. falciparum* genome (Fig. 1b). Additional associations with marginal levels of evidence were observed at a number of other loci, including a potential association between *GCNT2* in the host and *PfMSP4* in the parasite and associations involving HLA alleles (detailed in [Supplementary Methods](#), Supplementary Table 1), but here we focus on the association with HbS.

The statistical evidence for association at the HbS-associated loci can be described as follows, focussing on the variant with the strongest association in each region and assuming an additive model of effect of the host allele on parasite genotype on the log-odds scale (Supplementary Table 1). The chr2: 631,190 T>A variant, which lies in *PfACS8*, was associated with HbS with a Bayes factor (BF_{HbS}) of 1.1×10^{15} (computed under a log $F(2,2)$ prior; Methods) and P value of 4.8×10^{-13} (computed using a Wald test; [Supplementary Methods](#)). At a second region on chromosome 2, the chr2: 814,288 C>T variant, which lies in *Pf3D7_0220300*, was associated with $BF_{HbS} = 2.4 \times 10^9$ and $P = 1.6 \times 10^{-10}$. At the chromosome 11 locus, the chr11: 1,058,035 T>A variant, which lies in *Pf3D7_1127000*, was associated with $BF_{HbS} = 1.5 \times 10^{17}$ and $P = 7.3 \times 10^{-12}$. For brevity, we refer to these HbS-associated loci as *Pfsa1*, *Pfsa2* and *Pfsa3*, respectively—for *P. falciparum* sickle-associated—and we use + and – signs to refer to alleles that are positively and negatively correlated with HbS, respectively. For example, *Pfsa1*⁺ denotes the allele that is positively correlated with HbS at the *Pfsa1* locus. All three of the lead variants are nonsynonymous mutations of

their respective genes, as are additional associated variants in these regions (Fig. 1, Supplementary Table 1).

The above results are based on HbS genotypes imputed from surrounding haplotype variation⁸, but we focus below on the larger set of 4,071 cases in which we have previously directly assayed HbS genotypes⁵ (Extended Data Fig. 1). This includes the majority of samples used in our discovery analysis. The *Pfsa1* and *Pfsa3* associations were clearly supported in both populations in this dataset, whereas *Pfsa2*⁺ appears rare in Gambia (Supplementary Tables 2, 3). We also observed convincing replication of the associations in the additional 825 samples that were not part of our discovery phase, with nominal replication of *Pfsa3* in the Gambia (one-tailed $P = 0.026$, $N = 163$) and replication of all three loci in the larger sample from Kenya ($P < 0.001$, $N > 540$) (Supplementary Table 2). Across the full dataset there is thus very strong evidence of association with HbS at all three loci ($BF_{HbS} = 2.0 \times 10^{21}$ for *Pfsa1*, 3.7×10^{12} for *Pfsa2*, and 1.4×10^{24} for *Pfsa3*; Extended Data Fig. 2) with corresponding large effect size estimates (estimated odds ratio (OR) = 12.8 for *Pfsa1*⁺, 7.5 for *Pfsa2*⁺ and 21.7 for *Pfsa3*⁺). As described above, these estimates assume an additive relationship between HbS and the *P. falciparum* genotype at each locus, but we also noted that there is greatest evidence for a dominance effect (Supplementary Tables 2, 3).

We further examined the effect of adjusting for covariates in our data, including human and parasite principal components reflecting population structure, year of sampling, clinical type of severe malaria and technical features related to sequencing (Extended Data Fig. 3). Inclusion of these covariates did not substantially affect results with one exception: we found that parasite principal components computed across the whole *P. falciparum* genome included components that correlated with the *Pfsa* loci, and including these principal components reduced the association signal, particularly in Kenya. Altering the principal components by removing the *Pfsa* regions restored the association, indicating that this is not caused by a general population structure effect that is reflected in genotypes across the parasite genome, and we further discuss the reasons for this finding below. Finally, we analysed available data from a set of 32 uncomplicated infections of Malian children ascertained based on HbS genotypes¹⁶ (Methods); this provided further replication of the associations with *Pfsa1* and *Pfsa3* (Supplementary Table 2). Together, these data indicate that there are genuine differences in the distribution of parasite genotypes between infections of HbS and non-HbS genotype individuals.

HbS protection varies with parasite type

The level of protection afforded by HbS against severe malaria can be estimated by comparing its frequency between cases and population controls. As shown in Fig. 2, the vast majority of children with HbS genotype in our data were infected with

parasites that carry *Pfsa*⁺ alleles. Corresponding to this, our data show little evidence of a protective effect of HbS against severe malaria with parasites of *Pfsa1*⁺, *Pfsa2*⁺ and *Pfsa3*⁺ genotype (estimated relative risk (RR) = 0.83, 95% confidence interval = 0.53–1.30). By contrast, HbS is strongly associated with reduced risk of disease caused by parasites of *Pfsa1*[−], *Pfsa2*[−] and *Pfsa3*[−] genotype (RR = 0.01, 95% confidence interval = 0.007–0.03). These estimates should be interpreted with caution because they are based on just 49 cases of severe malaria that had an HbS genotype, because many of these samples were included in the initial discovery dataset, and because there is some variation evident between populations. However, it can be concluded that the protective effect of HbS is dependent on parasite genotype at the *Pfsa* loci.

Fig. 2: The estimated relative risk for HbS varies by *Pfsa* genotype.

 figure 2

a, Numbers of cases of severe malaria from the Gambia and Kenya with indicated HbS genotype (columns) and carrying the indicated alleles at the *Pfsa1*, *Pfsa2* and *Pfsa3* loci (rows; using $n = 4,054$ samples with directly typed HbS genotype and non-missing genotype at the three *P. falciparum* loci). *Pfsa* alleles positively associated with HbS are denoted + and those negatively associated with HbS are denoted − for the respective loci. Samples with mixed *P. falciparum* genotype calls for at least one of the loci are shown in the bottom row and further detailed in Extended Data Fig. 4. The first row indicates counts of HbS genotypes in population control samples from the same populations⁸. **b**, The estimated relative risk of HbS for severe malaria with *Pfsa* genotypes (rows) as indicated in **a**. Relative risks were estimated using a multinomial logistic regression model with controls as the baseline outcome and assuming complete dominance (that is, that HbAS and HbSS genotypes have the same association with parasite genotype) as described in [Supplementary Methods](#); an indicator of country was included as a covariate. Circles reflect posterior mean

estimates and horizontal lines reflect the corresponding 95% credible intervals (CI). Estimates based on less than 5 individuals with HbAS or HbSS genotypes are represented by smaller circles. To reduce overfitting we used Stan⁴⁶ to fit the model assuming a mild regularising Gaussian prior with mean zero and standard deviation of 2 on the log-odds scale (that is, with 95% of mass between 1/50 and 50 on the relative risk scale) for each parameter, and between-parameter correlations set to 0.5.

Population genetics of the *Pfsa* loci

The *Pfsa1*⁺, *Pfsa2*⁺ and *Pfsa3*⁺ alleles had similar frequencies in Kenya (approximately 10–20%) whereas in Gambia *Pfsa2*⁺ had a much lower allele frequency than *Pfsa1*⁺ or *Pfsa3*⁺ (below 3% in all years studied, versus 25–60% for the *Pfsa1*⁺ or *Pfsa3*⁺ alleles; Fig. 3a). To explore the population genetic features of these loci in more detail, we analysed the MalariaGEN Pf6 open resource, which provides *P. falciparum* genome variation data for 7,000 worldwide samples¹¹ (Fig. 3b). This showed considerable variation in the frequency of these alleles across Africa, the maximum observed value being 61% for *Pfsa3*⁺ in the Democratic Republic of Congo, and indicated that these alleles are rare outside Africa. Moreover, we found that within Africa, population frequencies of the *Pfsa*⁺ alleles are strongly correlated with the frequency of HbS (Fig. 3c; estimated using data from the Malaria Atlas Project¹⁷).

Fig. 3: The relationship between *Pfsa* and HbS allele frequencies across populations.

 **figure 3**

a. Bars show the estimated frequency of each *Pfsa*⁺ allele in severe cases of malaria from each country. Details of allele frequencies and sample counts across years are presented in Extended Data Fig. 5. **b.** Estimated frequency of each *Pfsa*⁺ allele in worldwide populations from the MalariaGEN Pf6 resource¹¹, which contains samples collected during the period 2008–2015. Only countries with at least 50 samples are shown (this excludes Columbia, Peru, Benin, Nigeria, Ethiopia, Madagascar and Uganda). **c.** Estimated population-level *Pfsa*⁺ allele frequency (as in **a**, **b**) against HbS allele frequency in populations from MalariaGEN Pf6 (coloured as in **b**; selected populations are also labelled). *Pfsa*⁺ allele frequencies were computed from the relevant genotypes, after excluding mixed or missing genotype calls. HbS allele frequencies were computed from frequency estimates previously published by the Malaria Atlas Project¹⁷ for each country, by averaging over the locations of MalariaGEN Pf6 sampling sites weighted by the sample size. DR, Democratic Republic; PNG, Papua New Guinea.

This analysis also revealed a further feature of the *Pfsa*⁺ alleles: although *Pfsa1* and *Pfsa2* are separated by 180 kb, and the *Pfsa3* locus is on a different chromosome, they are in strong linkage disequilibrium (LD). This can be seen from the co-occurrence of these alleles in severe cases (Fig. 2), and from the fact that they covary over time in our sample (Extended Data Fig. 5) and geographically across populations (Fig. 3b). We computed LD metrics between the *Pfsa*⁺ alleles in each population (Supplementary Table 4) after excluding HbS-carrying individuals to avoid

confounding with the association outlined above. *Pfsa1*⁺ and *Pfsa2*⁺ were strongly correlated in Kenyan severe cases ($r = 0.75$) and *Pfsa1*⁺ and *Pfsa3*⁺ were strongly correlated in both populations ($r = 0.80$ in Kenya; and $r = 0.43$ in severe cases from the Gambia). This high LD was not explained by population structure or other covariates in our data (Methods), and was also observed in multiple populations in MalariaGEN Pf6 (for example, $r = 0.20$ between *Pfsa1*⁺ and *Pfsa3*⁺ in the Gambia; $r = 0.71$ in Kenya; and $r > 0.5$ in all other African populations surveyed; Supplementary Table 4), showing that the LD is not purely an artefact of our sample of severe malaria cases.

This observation of strong correlation between alleles at distant loci is unexpected, because the *P. falciparum* genome undergoes recombination in the mosquito vector and typically shows very low levels of LD in malaria-endemic regions^{11,18,19}. To confirm that this is unusual, we compared LD between the *Pfsa* loci with the distribution computed from all common biallelic variants on different chromosomes (Fig. 4). In Kenyan samples, the *Pfsa* loci have the highest between-chromosome LD of any pair of variants in the genome. In Gambia, between-chromosome LD at these SNPs is also extreme, but another pair of extensive regions on chromosomes 6 and 7 also show strong LD. These regions contain the chloroquine resistance-linked genes *PfCRT* and *PfAAT1*^{20,21} and contain long stretches of DNA sharing identical by descent, consistent with positive selection of antimalarial-resistant haplotypes²². Moreover, we noted that these signals are among a larger set of HbS-associated and drug-resistance loci that appear to have increased between-chromosome LD in these data (Supplementary Table 4).

Fig. 4: HbS-associated variants show extreme between-chromosome correlation in severe *P. falciparum* infections.

 **figure 4**

Empirical distribution of absolute genotype correlation ($|r|$) between pairs of variants on different *P. falciparum* chromosomes in the Gambia (top) and Kenya (bottom). To avoid capturing direct effects of the HbS association, correlation values are computed after excluding HbS-carrying individuals. All pairs of biallelic variants with estimated minor allele frequency at least 5% and at least 75% of samples having non-missing and non-mixed genotype call are shown (totalling 16,487 variants in the Gambia and 13,766 variants in Kenya). Colours indicate the subset of comparisons between HbS-associated variants in *Pfsa* regions relevant for the population (red) and between variants in LD with the CRT K76T mutation. Labelled points denote the variant pairs

showing the highest and second-highest pairwise correlation in each population after grouping correlated variants into regions; for this purpose regions were defined to include all nearby pairs of correlated variants with minor allele frequency $\geq 5\%$ and $r^2 > 0.05$, such that no other such pair of variants within 10 kb of the given region boundaries is present (Methods). A longer list of regions showing increased between-chromosome LD is presented in Supplementary Table 5.

Combining these new findings with other population genetic evidence from multiple locations across Africa, including observations of frequency differentiation within and across *P. falciparum* populations^{11,23,24} and other metrics at these loci indicative of selection^{22,25,26}, it appears likely that the allele frequencies and strong LD between *Pfsa1*, *Pfsa2* and *Pfsa3* are maintained by some form of natural selection. However, the mechanism for this is unclear. Given our findings, an obvious hypothesis is that the *Pfsa1*⁺, *Pfsa2*⁺ and *Pfsa3*⁺ alleles are positively selected in hosts with HbS, but since the frequency of HbS carriers^{5,17} is typically $<20\%$ it is not clear whether this alone is a sufficient explanation to account for the high population frequencies or the strong LD observed in non-HbS carriers. Equally, since the *Pfsa*⁺ alleles have not reached fixation (Fig. 3) and do not appear to be rapidly increasing in frequency (Extended Data Fig. 5), an opposing force may also be operating to maintain their frequency. However, the above data do not suggest strong fitness costs for *Pfsa*⁺-carrying parasites in HbAA individuals (Fig. 2), and the *Pfsa2*⁺ allele also appears to be present only in east Africa, further complicating these observations. It thus remains entirely possible that additional selective factors are involved, such as epistatic interactions between these loci, or further effects on fitness in the host or vector in addition to those observed here in relation to HbS.

The genomic context of the *Pfsa* variants

The biological function of these parasite loci is an area of considerable interest for future investigation. At the *Pfsa1* locus, the signal of association includes non-synonymous changes in the *PfACS8* gene, which encodes an acyl-CoA synthetase³ that belongs to a gene family that has expanded in the Laverania relative to other *Plasmodium* species⁴ and lies close to a parologue *PfACS9* on chromosome 2. *PfACS8* has been predicted to localize to the apicoplast²⁷, but it also contains a *Plasmodium* export element (PEXEL)-like motif^{28,29,30}, which may instead indicate export to the host cytosol where other acyl-CoA synthetase family members have been observed³¹. The functions of the proteins encoded by *PF3D7_0220300* (an exported protein, at the *Pfsa2* locus) and *PF3D7_1127000* (a putative tyrosine phosphatase, at *Pfsa3*) are not known; however, the protein encoded by *PF3D7_0220300* has been observed to localise to the host membrane and to colocalise with host stomatin³², whereas the protein encoded by *PF3D7_1127000* has been observed in the food vacuole³³. All

three genes appear to be expressed at multiple parasite lifecycle stages ([Supplementary Text](#)) in 3D7 parasites and are not essential for in vitro growth³⁴.

We noted two further features that may point to the functional role of the *Pfsa*⁺ alleles themselves. The associated variants at *Pfsa2* and *Pfsa3* each include SNPs immediately downstream of a PEXEL motif (detailed in [Supplementary Information](#)), which mediates export through a pathway that involves protein cleavage at the motif³⁵. This process leaves the downstream amino acids at the N terminus of the mature protein, and it is therefore possible that these variants influence successful export^{36,37}. However, another possibility is that the *Pfsa*⁺ alleles affect levels of transcription of the relevant genes. In this context, we noted a recent study¹⁶ that found that *PF3D7_1127000* is among the most differentially over-expressed genes in trophozoite-stage infections of children with HbAS compared with those with HbAA (more than 32-fold increase in transcripts per million (TPM) at the trophozoite stage; $n = 12$; unadjusted $P = 5.6 \times 10^{-22}$). We reanalysed these data in light of genotypes at the *Pfsa* loci (Supplementary Table [6](#)), and found that the *Pfsa3*⁺ mutations plausibly explain this increased expression. In particular, read ratios at the second-most-associated *Pfsa3* SNP (chr11:1,057,437 T > C) (Supplementary Table [1](#)) appear especially strongly correlated with increased expression at trophozoite stage (Extended Data Fig. [6](#)). Further support for this observation comes from an in vitro time-course experiment conducted in the same study¹⁶, in which the increased expression was also observed in HbAA erythrocytes infected with a *Pfsa*⁺-carrying isolate (Extended Data Figs. [7](#), [8](#), Methods). The mechanism of upregulation is not known, but a further relevant observation is that the *Pfsa3*⁺ alleles appear to be linked to a neighbouring copy number variant that includes duplication of the 5' end of the small nuclear ribonucleoprotein gene *SNRPF*, upstream of *PF3D7_1127000* (based on analysis of available genome assemblies of *P. falciparum* isolates³⁸; Extended Data Fig. [9](#), Supplementary Fig. [3](#)). We caution that these findings are tentative, and the manner in which *Pfsa* alleles affect genome function is a subject for future research. Understanding this functional role could provide important clues into how HbS protects against malaria and help to distinguish between the various proposed mechanisms, which include enhanced macrophage clearance of infected erythrocytes³⁹, inhibition of intraerythrocytic growth dependent on oxygen levels⁴⁰, altered cytoadherence of infected erythrocytes⁴¹ due to cytoskeleton remodelling⁴², and immune-mediated mechanisms⁴³.

Discussion

A fundamental question in the biology of host-pathogen interactions is whether the genetic makeup of infections is determined by the genotype of the host. While there is some previous evidence of this in malaria—for example, allelic variants of the *PfCSP*

gene have been associated with HLA type⁴⁴ and HbS has itself previously been associated with MSP-1 alleles⁴⁵ (described further in [Supplementary Information](#))—our findings provide clear evidence of an interaction between genetic variants in the parasite and the host. Our central discovery is that among African children with severe malaria there is a strong association between HbS in the host and three loci in different regions of the parasite genome. Based on estimation of relative risk, HbS has no apparent protective effect against severe malaria in the presence of the *Pfsa1*⁺, *Pfsa2*⁺ and *Pfsa3*⁺ alleles. These alleles, which are much more common in Africa than elsewhere, are positively correlated with HbS allele frequencies across populations. However, they are also found in substantial numbers of individuals without HbS, reaching up to 60% allele frequency in some populations. The *Pfsa1*, *Pfsa2* and *Pfsa3* loci also show remarkably high levels of long-range between-locus LD relative to other loci in the *P. falciparum* genome, which is equally difficult to explain without postulating ongoing evolutionary selection. Although it seems clear that HbS has a key role in this selective process, there is a need for further population surveys (that include asymptomatic and uncomplicated cases of malaria) to gain a more detailed understanding of the genetic interactions between HbS and these parasite loci, and how they affect the overall protective effect of HbS against severe malaria.

Methods

Ethics and consent

Sample collection and design of our case-control study⁸ was approved by Oxford University Tropical Research Ethics committee (OXTREC), Oxford, United Kingdom (OXTREC 020-006). Informed consent was obtained from parents or guardians of patients with malaria, and from mothers for population controls. Local approving bodies were the MRC/Gambia Government Ethics Committee (SCC 1029v2 and SCC670/630) and the KEMRI Research Ethics Committee (SCC1192).

Building a combined dataset of human and *P. falciparum* genotypes for severe cases

We used Illumina sequencing to generate two datasets jointly reflecting human and *P. falciparum* genetic variation, using a sample of severe malaria cases from the Gambia and Kenya for which human genotypes have previously been reported^{5,8}. A full description of our sequencing and data processing is given in [Supplementary Methods](#) and summarized in Extended Data Fig. 1. In brief, following a process of sequence data quality control and merging across platforms, we generated (1) a dataset of microarray and imputed human genotypes, and genome-wide *P. falciparum* genotypes, in 3,346 individuals previously identified as without close relationships⁸; and (2) a

dataset of HbS genotypes directly typed on the Sequenom iPLEX Mass-Array platform (Agena Biosciences)⁵, and genome-wide *P. falciparum* genotypes, in 4,071 individuals without close relationships⁸. Parasite DNA was sequenced from whole DNA in samples with high parasitaemia, and using selective whole-genome amplification (SWGA) to amplify *P. falciparum* DNA in all samples. *P. falciparum* genotypes were called using an established pipeline¹¹ based on GATK, which calls single nucleotide polymorphisms and short insertion–deletion variants relative to the Pf3D7 reference sequence. This pipeline deals with mixed infections by calling parasite variants as if the samples were diploid; in practice this means that variants with substantial numbers of reads covering reference and alternate alleles are called as heterozygous genotypes.

For the analyses presented in main text, we used the 3,346 samples with imputed human genotypes for our initial discovery analysis, and the 4,071 individuals with directly typed HbS genotypes for all other analysis. The individuals in these two datasets substantially overlap (Extended Data Fig. 1), but a subset of 825 individuals have directly typed for HbS but were not in the discovery data and we used these for replication.

Inference of genetic interaction from severe malaria cases

To describe our approach, we first consider a simplified model of infection in which parasites have a single definite (measurable) genotype, acquired at time of biting, that is relevant to disease outcome—that is, we neglect any effects of within-host mutation, co- and super-infection at the relevant genetic variants. We consider a population of individuals who are susceptible to being bitten by an infected mosquito. A subset of infections go on to cause severe disease. Among individuals who are bitten and infected with a particular parasite type $I=x$, the association of a human allele $E=e$ with disease outcome can be measured by the relative risk,

$$\text{RR} = \frac{P(\{\text{disease}\} | E=e, I=\{x\})}{P(\{\text{disease}\} | E=0, I=\{x\})} \quad (1)$$

where we have used $E=0$ to denote a chosen baseline human genotype against which risks are measured. If the strength of association further varies between parasite types (say between $(I=x)$ and a chosen infection type $(I=0)$) then these relative risks will vary, and thus the ratio of relative risks (RRR) will differ from 1. If the host genotype e confers protection against severe malaria, the ratio of relative risks will therefore capture variation in the level of protection compared between different parasite types.

Although phrased above in terms of a relative risk for human genotypes, the RRR can be equivalently expressed as a ratio of relative risks for a given parasite genotype compared between two human genotypes ([Supplementary Methods](#)). It is thus conceptually symmetric with respect to human and parasite alleles, and would equally well capture variation in the level of pathogenicity conferred by a particular parasite type compared between different human genotypes.

The OR for specific human and parasite alleles computed in severe malaria cases is formally similar to the ratio of relative risks but with the roles of the genotypes and disease status interchanged. We show in [Supplementary Methods](#) that in fact

$$\text{OR} = \text{RRR} \times \text{OR}^{\text{biting}} \quad (2)$$

where $\text{OR}^{\text{biting}}$ is a term that reflects possible association of human and parasite genotypes at the time of mosquito biting. Thus, under this model and in the absence of confounding factors, $(\text{OR} \neq 1)$ implies either that host and parasite genotypes are not independent at time of biting, or that there is an interaction between host and parasite genotypes in determining disease risk. The former possibility may be considered less plausible because it would seem to imply that relevant host and parasite genotypes can be detected by mosquitos prior to or during biting, but we stress that this cannot be tested formally without data on mosquito-borne parasites. A further discussion of these assumptions can be found in [Supplementary Methods](#).

Testing for genome-to-genome correlation

We developed a C++ program (HPTEST) to efficiently estimate the odds ratio (equation (2)) across multiple human and parasite variants, similar in principle to approaches that have been developed for human-viral and human-bacterial GWAS^{[47,48,49](#)}. HPTEST implements a logistic regression model in which genotypes from one file are included as the outcome variable and genotypes from a second file on the same samples are included as predictors. Measured covariates may also be included, and the model accounts for uncertainty in imputed predictor genotypes using the approach from SNPTEST^{[50](#)}. The model is fit using a modified Newton-Raphson with line search method. For our main analysis we applied HPTEST with the parasite genotype as outcome and the host genotype as predictor, assuming an additive effect of the host genotype on the log-odds scale, and treating parasite genotype as a binary outcome (after excluding mixed and missing genotype calls.).

To mitigate effects of finite sample bias, we implemented regression regularised by a weakly informative $\log F(2,2)$ prior distribution^{[51](#)} on the effect of the host allele (similar to a Gaussian distribution with standard deviation 1.87; [Supplementary](#)

[Methods](#)). Covariate effects were assigned a $\log F(0.08,0.08)$ prior, which has similar 95% coverage interval to a gaussian with zero mean and standard deviation of 40. We summarised the strength of evidence using a Bayes factor against the null model that the effect of the host allele is zero. A P -value can also be computed under an asymptotic approximation by comparing the maximum posterior estimate of effect size to its expected distribution under the null model ([Supplementary Methods](#)). For our main results we included only one covariate, an indicator of the country from which the case was ascertained (Gambia or Kenya); additional exploration of covariates is described below.

Choice of genetic variants for testing

For our initial discovery analysis we concentrated on a set of 51,552 *P. falciparum* variants that were observed in at least 25 individuals in our discovery set, after excluding any mixed or missing genotype calls. These comprised: 51,453 variants that were called as biallelic and passed quality filters (detailed in [Supplementary Methods](#); including the requirement to lie in the core genome⁵²); an additional 98 biallelic variants in the region of *PfEBL1* (which lies outside the core genome but otherwise appeared reliably callable); and an indicator of the *PfEBA175* ‘F’ segment, which we called based on sequence coverage as described in [Supplementary Methods](#) and Supplementary Fig. 6. We included *PfEBL1* and *PfEBA175* variation because these genes encode known or putative receptors for *P. falciparum* during invasion of erythrocytes¹⁵.

We concentrated on a set of human variants chosen as follows: we included the 94 autosomal variants from our previously reported list of variants with the most evidence for association with severe malaria⁸, which includes confirmed associations at *HBB*, *ABO*, *ATP2B4* and the glycophorin locus. We also included three glycophorin structural variants¹³, and 132 HLA alleles (62 at 2-digit and 70 at 4-digit resolution) that were imputed with reasonable accuracy (determined as having minor allele frequency $> 5\%$ and IMPUTE info at least 0.8 in at least one of the two populations in our dataset). We tested these variants against all 51,552 *P. falciparum* variants described above. We also included all common, well-imputed human variants within 2 kb of a gene determining a blood-group antigen (defined as variants within 2 kb of a gene in the HUGO blood-group antigen family⁵³ and having a minor allele frequency of 5% and an IMPUTE info score of at least 0.8 in at least one of the two populations in our dataset; this includes 39 autosomal genes and 4,613 variants in total). We tested these against all variants lying within 2 kb of *P. falciparum* genes previously identified as associated or involved in erythrocyte invasion^{14,15} (60 genes, 1740 variants in total). In total we tested 19,830,288 distinct human-parasite variant pairs in the discovery dataset (Fig. 1a).

Definition of regions of pairwise association

We grouped all associated variant pairs (defined as pairs (v,w) having $\text{BF}(v,w) > 100$, where $\text{BF}(v,w)$ is the association test Bayes factor for the variant pair) into regions using an iterative algorithm as follows. For each associated pair (v,w) , we found the smallest enclosing regions (R_v, R_w) such that any other associated pair either lay with (R_v, R_w) or lay further than 10 kb from (R_v, R_w) in the host or parasite genomes, repeating until all associated pairs were assigned to regions. For each association region pair, we then recorded the region boundaries and the lead variants (defined as the regional variant pair with the highest Bayes factor), and we identified genes intersecting the region and the gene nearest to the lead variants using the NCBI refGene⁵⁴ and PlasmoDB v44⁵⁵ gene annotations. Due to our testing a selected list of variant pairs as described above, in some cases these regions contain a single human or parasite variant. Supplementary Table 1 summarises these regions for variant pairs with $\text{BF} > 1,000$.

Interpretation of association test results

We compared association test P -values to the expectation under the null model of no association using a quantile-quantile plot, both before and after removing comparisons with HbS (Supplementary Fig. 2; HbS is encoded by the ‘A’ allele at rs334, chr11:5,248,232 T -> A). A simple way to interpret individual points on the QQ-plot is to compare each P -value to its expected distribution under the relevant order statistic (depicted by the grey area in Supplementary Fig. 2); for the lowest P -value this is similar to considering a Bonferroni correction. However, we caution that thresholds determined by this approach are dependent on the set of tests carried out. A more defensible approach is to ask what P -value threshold is needed to generate confidence that a particular pair of variants is genuinely associated. This depends on both the prior probability of association and on the statistical power, with the relationship.

$$\text{\$}\$\{\text{\rm\{posterior\}}\},\{\text{\rm\{odds\}}\},(\{\text{\rm\{association\}}\}|p < T)=\{\text{\rm\{prior\}}\},\{\text{\rm\{odds\}}\},\times \frac{\{\text{\rm\{power\}}\}}{T}\$\$ \\ (3)$$

for any P -value threshold T (Supplementary Methods)⁵⁰. The term on the left is the odds of true association given observation of a P -value below the given threshold; the corresponding probability is therefore equal to one minus the positive false discovery rate⁵⁶. Interpretation of (3) requires knowledge of both a relevant prior odds of association, and the power, which in turn depends on the true effect size distribution and the underlying frequencies of the variants.

A similar approach conditional on the observed data ([Supplementary Methods](#)) leads to an analogous formula involving the Bayes factor instead of the power term and T . If the distribution of truly associated variant effect sizes is similar to the $\log F(2,2)$ distribution we have used to calculate Bayes factors, and if variant pairs have approximately similar prior probability of association, then a fixed threshold on the Bayes factor would provide an approximately constant posterior probability of association.

We illustrate a possible computation as follows. The 51,552 *P. falciparum* variants represent around 20,000 1 kb regions of the *P. falciparum* genome, which might be thought of as approximately independent given LD decay rates¹¹; similarly the human genome may be thought of as consisting of around 2 million approximately independent regions. If we take the view that a small number—say up to ten—of pairs of regions might be associated, this dictates prior odds on the order of 1 in 4 billion. A Bayes factor around 10^{10} would therefore be needed to generate substantial posterior odds of association, while a Bayes factor an order of magnitude higher would provide compelling evidence (posterior probability $> 95\%$). In [Supplementary Fig. 4](#) and [Supplementary Methods](#) we detail the analogous calculation applied to P-values. For large effect sizes on the order of $OR \approx 4$, this suggests that P-values on the order of 1×10^{-10} to 1×10^{-12} might provide compelling evidence for association, depending on the allele frequencies, but weaker effects would require lower thresholds and would be less easily detectable.

It might be considered that the human variants and genes that we have considered here are among those with the highest prior plausibility for association with parasites, and thus the above choice of prior may be considered somewhat conservative. However, even under stronger prior odds on the order of 1 in 2 million (for example, assuming 10 associations among the variant pairs tested in our study), our results do not identify any associations additional to the HbS–*Pfsa* associations with very strong evidence. Particular variants may however be of further interest due to specific prior plausibility; in [Supplementary Methods](#) we give further details on putative associations with $BF > 10^5$ and those involving known malaria-protective mutations in the human genome.

Summarizing evidence for each variant

For each human variant v , we further summarised the evidence that v is associated with variation in the parasite genome using the mean Bayes factor $\langle \{ \{ \text{rm} \{ BF \} \} \}_v \{ \{ \text{rm} \{ avg \} \} \}(v) \rangle$, computed as the average of the Bayes factor $BF(v,w)$ across all the parasite variants w tested against v . Under the restrictive assumption that at most one parasite variant is associated with w , $\langle \{ \{ \text{rm} \{ BF \} \} \}_v \{ \{ \text{rm} \{ avg \} \} \}(v) \rangle$ can be interpreted as a model-averaged Bayes factor reflecting the evidence for association of v with parasite variants; more

generally $\{\rm{BF}\}_{\rm{avg}}$ provides a pragmatic way to combines evidence across all tested variants. We similarly define $\{\rm{BF}\}_{\rm{avg}}(w)$ for each parasite variant w averaged over all human variants tested against v . $\rm{BF}_{\rm{avg}}$ is plotted for human and parasite variants in Fig. 1a.

A direct interpretation of these average Bayes factors can be carried in a similar way to the individual Bayes factors as described above. To illustrate, if as above we assume that around 10 of the 20,000 1 kb regions of the *P. falciparum* genome might be associated with human genetic variants among those tested, an average Bayes factor $> 10,000$ would be needed to indicate $> 80\%$ posterior odds of association; this is only achieved for the *Pfsa* variants in our data (Fig. 1a). This calculation can be adjusted as appropriate to take into account specific information about individual variants.

Investigation of additional associations

In addition to the HbS-*Pfsa* associations, we also observed moderate evidence for association at a number of other variant pairs. These include associations between variation in the human gene *GCNT2* and *PfMSP4* with $BF = 2.8 \times 10^6$, and between HLA variation and multiple parasite variants with BF in the range 10^5 – 10^6 (Fig. 1a, Supplementary Table 1). A fuller description of the context of these SNPs can be found in [Supplementary Methods](#). Our interpretation is that the statistical evidence for these associations is not sufficiently strong on its own to make these signals compelling without additional evidence.

Assessment of possible confounding factors

To assess whether the observed association between HbS and *P. falciparum* alleles might be driven by confounding factors we conducted additional pairwise association tests as follows using HPTEST, based on directly typed HbS genotypes and working separately in the two populations. Results are shown in Extended Data Fig. 3. First, we repeated the pairwise association test including only individuals overlapping the discovery dataset, and separately in the remaining set of 825 individuals. For discovery samples a set of population-specific principal components (PCs) reflecting human population structure were previously computed⁸ and we included these as covariates (including 20 PCs in total). Second, across all 4,071 individuals with directly typed HbS data, we repeated tests including measured covariates as additional predictors. Specifically, we considered: (1) the age of individual at time of ascertainment (measured in years; range 0–12; treated as a categorical covariate), sex, reported ethnic group, and year of admission (range 1995–2010, treated as a categorical covariate); (2) technical covariates including an indicator of method of sequencing (SWGA or whole DNA), mean depth of coverage of the *P. falciparum* genome, mean insert size computed from aligned reads, and percentage of mixed calls;

and (3) an indicator of the clinical form of severe malaria with which the sample was ascertained ('SM subtype'; either cerebral malaria, severe malarial anaemia, or other).

To assess the possibility that parasite population structure might impact results, we also included PCs computed in parasite populations as follows. Working in each population separately, we started with the subset of biallelic SNPs with minor allele frequency at least 1% from among the 51,552 analysed variants (50,547 SNPs in Gambia and 48,821 SNPs in Kenya respectively). We thinned variants by iteratively picking variants at random from this list and excluding all others closer than 1 kb (leaving 12,036 SNPs in Gambia and 11,902 SNPs in Kenya). We used QCTOOL to compute PCs using this list of SNPs. Several of the top PCs had elevated loadings from SNPs in specific genomic regions. This was especially noticeable in Kenya and included the widely reported extensive regions of LD around the *AATI* and *CRT* regions on chromosomes 6 and 7, and also the HbS-associated chromosome 2 and 11 loci. We therefore also considered separate sets of PCs computed after excluding SNPs in chromosomes 6 and 7 (leaving 9,933 and 9,812 SNPs respectively), after excluding chromosomes 2 and 11 (10,521 and 10,421 SNPs respectively) or after excluding 100 kb regions centred on the lead HbS-associated SNPs (11,866 and 11,732 SNPs respectively). For each set of PCs, we repeated association tests including 20 PCs as fixed covariates.

For each subset of individuals, each HbS-associated variant and each set of covariates described above, we plotted the estimated effect size and 95% posterior interval, annotated with the total number of samples, the number carrying the non-reference allele at the given variant, and the number carrying heterozygous or homozygous HbS genotypes (Extended Data Fig. 3). Corresponding genotype counts can be found in Supplementary Table 3. To assess mixed genotype calls, we also plotted the ratio of reads with reference and nonreference alleles at each site; this can be found in Extended Data Fig. 4.

Interpretation in terms of causal relationships

Observing $\text{OR} \neq 1$ implies non-independence between host and parasite genotypes in individuals with severe disease, but does not determine the mechanism by which this could occur. Assuming $(\{\text{rm}\{\text{OR}\}\})^{\wedge}(\{\text{rm}\{\text{biting}\}\})=1$, we show in [Supplementary Methods](#) that $(\{\text{rm}\{\text{OR}\}\}=1)$ is equivalent to a multiplicative model in which host and parasite genotypes separately and multiplicatively affect disease risk (equations (S6) and (S7) in [Supplementary Methods](#)). In general deviation from this model could arise in several ways, including through within-host selection, interaction effects determining disease tolerance, as well as potential effects not specific to individual variants but relating to disease diagnosis (similar to Berkson's paradox⁵⁷). Our study provides only limited data to distinguish these possible

mechanisms. For the HbS association described in main text, we note in [Supplementary Methods](#) that there is little evidence that the *Pfsa*⁺ variants are themselves associated with increased disease risk, and little evidence that the *Pfsa*⁺ variants associate with other host protective variants, suggesting that the observed interaction is specific to HbS. The appearance of the association in uncomplicated cases further suggests that the effect is not specifically related to infections manifesting as severe disease.

Genotype inference from uncomplicated cases from Mali

We called genotypes from previously published RNA-seq data from 32 children ascertained with uncomplicated malaria from Mali¹⁶ (NCBI BioProject PRJNA685106). We used BAM files of reads aligned to the 3D7 reference genome as generated previously¹⁶. All data were from 50 bp paired end RNA-seq of total RNA. For each *Pfsa* lead variant, we counted the number of reads aligning to the variant with the reference, alternate, or other allele, ignoring those with lower than 20 mapping quality. These read counts are presented in Supplementary Table 6 for the lead *Pfsa* mutations. We treated genotypes as mixed (and excluded these from association analysis in Supplementary Table 2) if > 10% of reads covering both alleles were observed; otherwise, we assigned a specific genotype. At *Pfsa3*, we also tabulated genotypes for the second-most associated SNP (chr11:1,057,437 T>C) because we noted one sample (AS15) that had different genotype between this and the chr11:1,085,035 T>A variant.

Comparison of severe cases to human population controls

Equation (3) represents the odds ratio in severe cases as a ratio of relative risks for a given human allele *e* conditional on the parasite genotype at infection time. A closely related interpretation involves the relative risk of severe malaria with the observed (disease-time) parasite type,

$$\text{Equation (4)}: \text{RR}(y) = \frac{P(\{\text{severe}\}, \{\text{disease}\}, \{\text{with}\}, \{\text{parasite}\}, \{\text{type}\}, y|E=e)}{P(\{\text{severe}\}, \{\text{disease}\}, \{\text{with}\}, \{\text{parasite}\}, \{\text{type}\}, y|E=0)}$$

A calculation shows that the OR in severe cases is also equal to the ratio $\text{RR}(y)/\text{RR}(0)$. The relative risk (equation (4)) can be computed by comparing the frequencies of host genotypes between disease cases with infection genotype *y* and in the whole population. In [Supplementary Methods](#) we extend this to show that $\text{RR}(y)$ can be estimated using multinomial logistic regression with population controls and cases stratified by parasite type as outcome levels, and with

the human genotype and any covariates included as predictors. We apply this approach in Fig. 2 to estimate $\text{rm}\{\text{RR}\}\{(\text{y})\}$, where y ranges over combined genotypes at the three *Pfsa* loci, conditional on the country of sampling.

Assessing sequencing performance in HbS-associated regions

We assessed sequencing performance at the chr2:631,190, chr2:814,288 and chr11:1,058,035 loci by computing counts of reads aligning to each position ('coverage') and comparing this to the distribution of coverage across all biallelic sites in our dataset, treating each sample separately (Supplementary Fig. 5). In general coverage at the three sites was high; we noted especially high coverage at chr2:814,288 in sWGA sequencing data (for example, >90% of samples have coverage among the top 80% of that at biallelic variants genome-wide) but somewhat lower coverage in WGS samples at the chr11:1,058,035 locus. Variation in coverage between loci and samples is expected due to variation in DNA quantities, DNA amplification and sequencing processes, but we did not observe systematic differences in coverage between the different *Pfsa* genotypes at these loci. To further establish alignment accuracy, we also inspected alignment metrics. Across all analysis samples, over 99% of reads at each location carried either the reference or the identified non-reference allele, and over 99% of these reads had mapping quality at least 50 (representing confident read alignment). These results suggest sequence reads provide generally accurate genotype calls at these sites.

Assessing the distribution of between-chromosome LD

We developed a C++ program (LDBIRD) to efficiently compute LD between all pairs of *P. falciparum* variants. LDBIRD computes the frequency of each variant, and computes the correlation between genotypes at each pair of variants with sufficiently high frequency. It then generates a histogram of correlation values and reports pairs of variants with squared correlation above a specified level. We applied LDBIRD separately to *P. falciparum* data from Gambian and Kenyan severe malaria cases. We restricted attention to comparisons between biallelic variants that had frequency at least 5% in the given population and with at least 75% of samples having non-missing genotypes at both variants in the pair, after treating mixed genotype calls as missing, and output all pairs with r^2 at least 0.01 for further consideration. To avoid confounding of LD by the HbS association signal, we also repeated this analysis after excluding individuals that carry the HbS allele (with the latter results presented in Fig. 4 and Supplementary Table 4).

To summarise between-chromosome LD results we grouped signals into regions as follows. First, we observed that most variant pairs have $|r| < 0.15$ and hence $r^2 > 0.05$ is typically a substantially outlying degree of inter-chromosomal LD (Fig. 4). We

therefore focussed on variant pairs (v_1, v_2) with $r^2 > 0.05$. To each such pair (v_1, v_2) we assigned a pair of LD regions (R_1, R_2) with the property that R_1 and R_2 capture all other nearby variants with high r^2 . Specifically, R_1 and R_2 are defined as the smallest regions containing v_1 and v_2 respectively, such that no other variant pair within 10 kb of (R_1, R_2) has $r^2 > 0.05$. To compute R_1 and R_2 , we implemented an iterative algorithm that successively expands the initial pair until no additional nearby pairs with high r^2 can be found. For each LD region pair we recorded the region boundaries and the most-correlated pair of variants. A full list of region pairs with $r^2 > 0.05$ is given in Supplementary Table 5; the highest LD pairs for *Pfsa* regions and for *PfCRT–PfAAT1* are shown in Fig. 4.

Assessing the influence of covariates on LD estimates

To investigate whether the observed between-locus LD might arise due to population structure effects or due to other artefacts captured by measured covariates in our data, we used HPTEST to fit a logistic regression model of association with the genotypes at one *Pfsa* locus as outcome and the genotypes at a second *Pfsa* locus as predictor, repeating for each pair of *Pfsa* regions, in each population separately. We fit the model including each of a set of covariates as follows: (1) no covariates; (2) 20 parasite PCs; (3) technical covariates including an indicator of the type of sequencing and sequence depth (as in Extended Data Fig. 3); (4) year of admission, or (5) all of the above combined. For each set of covariates we compared the estimated odds ratio indicating the strength of association to the unadjusted odds ratio. In Kenya, across covariate sets, the minimum and unadjusted estimates were 128.0 and 128.0 (*Pfsa1*⁺ vs *Pfsa2*⁺; minimum with no covariates), 218.0 vs 219.4 (*Pfsa1*⁺ vs *Pfsa3*⁺; minimum when including technical covariates) and 40.2 vs 47.2 (*Pfsa2*⁺ vs *Pfsa3*⁺; minimum when including parasite PCs). In Gambia the minimum and unadjusted estimates were 7.0 and 7.7 (*Pfsa1*⁺ and *Pfsa3*⁺, minimum when including parasite PCs). These results therefore suggest the observed LD is not substantially explained by population structure or other features of our sample that are captured by these covariates.

Genotype inference from the Uganda Palo Alto isolate

To determine the genotype of the Uganda Palo Alto (FUP/H) isolate (Extended Data Figures 7–8), we downloaded Illumina sequence read data from the Sequence Read Archive (accessions [SRR530503](#), [SRR629055](#) and [SRR629078](#), generated by the Broad Institute). All reads were 101 base pair paired end. We aligned the reads to the Pf3D7_v3 genome using bwa mem and inspected read pileups to determine the genotypes at HbS-associated mutations. These data clearly indicate that FUP/H carries the alternate allele at the *Pfsa1*–3 lead SNPs as well as at chr:1,057,437 T>C (based on >98% of reads carrying the non-reference allele).

Analysis of transcript expression from time course experiments

We analysed data from three previously reported experiments that measured the transcription of genes in 3D7 at different time stages following invasion of erythrocytes (accessions [PRJEB2015](#)⁵⁸ and [PRJEB31535](#)⁵⁹). Data were processed using a similar pipeline to that described previously¹⁶. In brief, reads were aligned to a concatenated human GRCh38 / Pf3D7 genome using STAR v2.7.3a, informed by the Gencode v38 human and the PlasmoDB v52 Pf3D7 gene annotations. Reads aligning to Pf3D7 were then extracted and transcript abundance (TPM) was estimated using Salmon v1.5.1. Estimated TPM values for genes in the *Pfsa* regions, along with previously computed TPM values from Saelens et al.¹⁶, are shown in Extended Data Fig. 7.

Assessing the structure of *Pfsa* regions in available genome assemblies

We extracted 101 bp and 1001 bp flanking sequence centred at the chr2:631,190, chr2:814,288 and chr11:1,058,035 loci from the Pf3D7 reference sequence. We then used minimap2⁶⁰ to align these sequences to a previously generated set of genome assemblies from *P. falciparum* isolates and laboratory strains³⁸ (Supplementary Table 7), allowing for multiple possible mapping locations. Each flanking sequence aligned to a single location on the corresponding chromosome in all included genomes, with the exception that sequence flanking the chromosome 11 locus aligned to two locations in the ML01 sample. This sample was excluded from previous analysis³⁸ as it represents a multiple infection; we comment further on this below.

To further inspect sequence identity, we used MAFFT to generate a multiple sequence alignment (MSA) corresponding to the 1001 bp sequence centred at each locus. Four isolates (GA01 from The Gabon, SN01 from Senegal, Congo CD01 and ML01 from Mali) carry the non-reference ‘A’ allele at the chr11:1,058,035 SNP; two of these (GA01 and CD01) also carry the non-reference allele at the chr2:631,190 SNP and one (CD01) carries the non-reference allele at all three SNPs. However, expansion of alignments to include a 10,001 bp segment indicated that these four samples also carry a structural rearrangement at the chr11 locus. Specifically, GA01, SN01, CD01 and ML01 genomes include a ~1 kb insertion present approximately 900 bp to the right of chr11:1,058,035, and also a ~400 bp deletion approximately 2,400 bp to the left of chr11:1,058,035. To investigate this, we generated *k*-mer sharing ‘dot’ plots for *k* = 50 across the region (Supplementary Fig. 3, Extended Data Fig. 9), revealing a complex rearrangement carrying both deleted and duplicated segments. The duplicated sequence includes a segment (approximate coordinates 1,054,000–1,055,000 in Pf3D7) that contains the gene *SNRPF* (‘small nuclear ribonucleoprotein F, putative’) in the Pf3D7 reference. Inspection of breakpoints did not reveal any other predicted gene copy number changes in this region, including for *Pf3D7_1127000*.

As noted above, the chromosome 11 region aligns to a second contig in ML01 (contig chr0_142, Supplementary Table 2). This contig appears to have a different tandem duplication of a ~4 kb segment lying to the right of the associated SNP (approximately corresponding to the range 11:1,060,100–1,064,000 in Pf3D7; Supplementary Fig. 3). This segment contains a number of genes including dUTPase, which has been under investigation as a potential drug target⁶¹. We interpret this second contig as arising due to the multiple infection in this sample³⁸, and given challenges inherent in genome assembly of mixed samples it is unclear whether this duplication represents an assembly artefact or a second genuine regional structural variant.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

Sequence read data from whole DNA and SWGA sequencing of *P. falciparum* genomes (as detailed in Extended Data Fig. 1) are available from the European Nucleotide Archive (study accession [ERP000190](#)). A full list of relevant sample accessions can be found at <http://www.malariagen.net/resource/32>. Human genotype data used in this study have been described previously^{5,8} and are available under managed-access terms from the European Genome–Phenome Archive (study accession [EGAS00001001311](#)), as detailed at <https://www.malariagen.net/resource/25>. A dataset of the human and *P. falciparum* genotypes for 3,346 severe cases of malaria used in our discovery scan (Fig. 1), and HbS genotypes and *P. falciparum* genotypes in the larger set of 4,071 severe cases with direct HbS typing (Fig. 2), is available from Zenodo (<https://doi.org/10.5281/zenodo.4973476>). Association test summary statistics from our discovery data (Fig. 1) are also available from Zenodo (<https://doi.org/10.5281/zenodo.5722497>). A full list of data generated by this study and associated resources can be found at <http://www.malariagen.net/resource/32>.

Code availability

Source code for HPTEST and LDBIRD is available at <https://code.enkre.net/qctool>. A snapshot has also been deposited at Zenodo (<https://doi.org/10.5281/zenodo.5685581>).

References

1. 1.
Kariuki, S. N. & Williams, T. N. Human genetics and malaria resistance. *Hum. Genet.* **139**, 801–811 (2020).
2. 2.

Bethke, L. L. et al. Duplication, gene conversion, and genetic diversity in the species-specific acyl-CoA synthetase gene family of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **150**, 10–24 (2006).

3. 3.

Matesanz, F., Téllez, M. A.-D.-M. & Alcina, A. The *Plasmodium falciparum* fatty acyl-CoA synthetase family (PfACS) and differential stage-specific expression in infected erythrocytes. *Mol. Biochem. Parasitol.* **126**, 109–112 (2003).

4. 4.

Otto, T. D. et al. Genomes of all known members of a *Plasmodium* subgenus reveal paths to virulent human malaria. *Nat. Microbiol.* **3**, 687–697 (2018).

5. 5.

Malaria Genomic Epidemiology Network. Reappraisal of known malaria resistance loci in a large multicenter study. *Nat. Genet.* **46**, 1197–1204 (2014).

6. 6.

Cowell, A. N. & Winzeler, E. A. The genomic architecture of antimalarial drug resistance. *Brief. Funct. Genomics* **18**, 314–328 (2019).

7. 7.

Gomes, P. S., Bhardwaj, J., Rivera-Correa, J., Freire-De-Lima, C. G. & Morrot, A. Immune escape strategies of malaria parasites. *Front. Microbiol.* **7**, 1617 (2016).

8. 8.

Band, G. et al. Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania. *Nat.*

Commun. **10**, 5732 (2019).

9. 9.

Band, G. et al. A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature* **526**, 253–257 (2015).

10. 10.

Oyola, S. O. et al. Whole genome sequencing of *Plasmodium falciparum* from dried blood spots using selective whole genome amplification. *Malaria J.* **15**, 597 (2016).

11. 11.

Ahouidi, A. et al. An open dataset of *Plasmodium falciparum* genome variation in 7,000 worldwide samples. *Wellcome Open Research* **6**, 42 (2021).

12. 12.

Timmann, C. et al. Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature* **489**, 443–446 (2012).

13. 13.

Leffler, E. M. et al. Resistance to malaria through structural variation of red blood cell invasion receptors. *Science* **356**, aam6393 (2017).

14. 14.

Cowman, A. F., Berry, D. & Baum, J. The cellular and molecular basis for malaria parasite invasion of the human red blood cell. *J. Cell Biol.* **198**, 961–971 (2012).

15. 15.

Cowman, A. F., Tonkin, C. J., Tham, W. H. & Duraisingh, M. T. The molecular basis of erythrocyte invasion by malaria parasites. *Cell Host*

Microbe **22**, 232–245 (2017).

16. 16.

Saelens, J. W. et al. Impact of sickle cell trait hemoglobin on the intraerythrocytic transcriptional program of *Plasmodium falciparum*. *mSphere* **6**, e0075521 (2021).

17. 17.

Piel, F. B. et al. Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *Lancet* **381**, 142–151 (2013).

18. 18.

Manske, M. et al. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487**, 375–379 (2012).

19. 19.

Mzilahowa, T., McCall, P. J. & Hastings, I. M. “Sexual” population structure and genetics of the malaria agent *P. falciparum*. *PLoS ONE* **2**, e613–e613 (2007).

20. 20.

Tindall, S. M. et al. Heterologous expression of a novel drug transporter from the malaria parasite alters resistance to quinoline antimalarials. *Sci. Rep.* **8**, 2464 (2018).

21. 21.

Wang, Z. et al. Genome-wide association analysis identifies genetic loci associated with resistance to multiple antimalarials in *Plasmodium falciparum* from China–Myanmar border. *Sci. Rep.* **6**, 33891 (2016).

22. 22.

Amambua-Ngwa, A. et al. Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa. *Science* **365**, 813–816 (2019).

23. 23.

Moser, K. A. et al. Describing the current status of *Plasmodium falciparum* population structure and drug resistance within mainland Tanzania using molecular inversion probes. *Mol. Ecol.* **30**, 100–113 (2021).

24. 24.

Verity, R. et al. The impact of antimalarial resistance on the genetic structure of *Plasmodium falciparum* in the DRC. *Nat. Commun.* **11**, 2107 (2020).

25. 25.

Chang, H.-H. et al. Genomic sequencing of *Plasmodium falciparum* malaria parasites from senegal reveals the demographic history of the population. *Mol. Biol. Evol.* **29**, 3427–3439 (2012).

26. 26.

Park, D. J. et al. Sequence-based association and selection scans identify drug resistance loci in the *Plasmodium falciparum* malaria parasite. *Proc. Natl Acad. Sci. USA* **109**, 13052 (2012).

27. 27.

Ralph, S. A. et al. Metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nat. Rev. Microbiol.* **2**, 203–216 (2004).

28. 28.

Hiller, N. L. et al. A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306**, 1934–1937 (2004).

29. 29.

Jonsdottir, T. K., Gabriela, M., Crabb, B. S., F. de Koning-Ward, T. & Gilson, P. R. Defining the essential exportome of the malaria parasite. *Trends Parasitol.* **37**, 664–675 (2021).

30. 30.

Marti, M., Good, R. T., Rug, M., Knuepfer, E. & Cowman, A. F. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306**, 1930–1933 (2004).

31. 31.

Matesanz, F., Durán-Chica, I. & Alcina, A. The cloning and expression of Pfacs1, a *Plasmodium falciparum* fatty acyl coenzyme A synthetase-1 targeted to the host erythrocyte cytoplasm. *J. Mol. Biol.* **291**, 59–70 (1999).

32. 32.

Butler, T. K. *An Exported Malaria Protein Regulates Glucose Uptake During Intraerythrocytic Infection*. PhD thesis, Washington Univ. in St Louis (2014).

33. 33.

Lamarque, M. et al. Food vacuole proteome of the malarial parasite *Plasmodium falciparum*. *Proteomics Clin. Appl.* **2**, 1361–1374 (2008).

34. 34.

Zhang, M. et al. Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. *Science* **360**, eaap7847 (2018).

35. 35.

Russo, I. et al. Plasmepsin V licenses *Plasmodium* proteins for export into the host erythrocyte. *Nature* **463**, 632–636 (2010).

36. 36.

Boddey, J. A. et al. Role of plasmepsin V in export of diverse protein families from the *Plasmodium falciparum* exportome. *Traffic* **14**, 532–550 (2013).

37. 37.

Grüning, C. et al. Uncovering common principles in protein export of malaria parasites. *Cell Host Microbe* **12**, 717–729 (2012).

38. 38.

Otto, T. D. et al. Long read assemblies of geographically dispersed *Plasmodium falciparum* isolates reveal highly structured subtelomeres. *Wellcome Open Res.* **3**, 52 (2018).

39. 39.

Luzzatto, L. Sickle cell anaemia and malaria. *Mediterr. J. Hematol. Infect. Dis.* **4**, e2012065 (2012).

40. 40.

Archer, N. M. et al. Resistance to *Plasmodium falciparum* in sickle cell trait erythrocytes is driven by oxygen-dependent growth inhibition. *Proc. Natl Acad. Sci. USA* **115**, 7350–7355 (2018).

41. 41.

Cholera, R. et al. Impaired cytoadherence of *Plasmodium falciparum*-infected erythrocytes containing sickle hemoglobin. *Proc. Natl Acad. Sci. USA* **105**, 991–996 (2008).

42. 42.

Cyrklaff, M. et al. Hemoglobins S and C interfere with actin remodeling in *Plasmodium falciparum* infected erythrocytes. *Science* **334**, 1283–1286 (2011).

43. 43.

Williams, T. N. et al. An immune basis for malaria protection by the sickle cell trait. *PLoS Med.* **2**, e128 (2005).

44. 44.

Gilbert, S. C. et al. Association of malaria parasite population structure, HLA, and immunological antagonism. *Science* **279**, 1173–1177 (1998).

45. 45.

Ntoumi, F. et al. Imbalanced distribution of *Plasmodium falciparum* MSP-1 genotypes related to sickle-cell trait. *Mol. Med.* **3**, 581–592 (1997).

46. 46.

Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual*. <https://mc-stan.org> (2021).

47. 47.

Ansari, M. A. et al. Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. *Nat. Genet.* **49**, 666–673 (2017).

48. 48.

Bartha, I. et al. A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *Elife* **2**, e01123 (2013).

49. 49.

Lees, J. A. et al. Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis. *Nat. Commun.* **10**, 2176 (2019).

50. 50.

Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).

51. 51.

Greenland, S. & Mansournia, M. A. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Stat. Med.* **34**, 3133–3143 (2015).

52. 52.

Miles, A. et al. Indels, structural variation, and recombination drive genomic diversity in *Plasmodium falciparum*. *Genome Res.* **26**, 1288–1299 (2016).

53. 53.

Braschi, B. et al. Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res.* **47**, D786–D792 (2019).

54. 54.

O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D74(2016).

55. 55.

Aurrecoechea, C. et al. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.* **37**, D539–D543 (2009).

56. 56.

Storey, J. D. The positive false discovery rate: a Bayesian interpretation and the q -value. *Ann. Stat.* **31**, 2013–2035 (2003).

57. 57.

Berkson, J. Limitations of the application of fourfold table analysis to hospital data. *Int. J. Epidemiol.* **43**, 511–515 (2014).

58. 58.

Otto, T. D. et al. New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-seq. *Mol. Microbiol.* **76**, 12–24 (2010).

59. 59.

Wichers, J. S. et al. Dissecting the gene expression, localization, membrane topology, and function of the *Plasmodium falciparum* STEVOR protein family. *mBio* **10**, e01500–e01519 (2019).

60. 60.

Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

61. 61.

Pérez-Moreno, G. et al. Validation of *Plasmodium falciparum* dUTPase as the target of 5'-tritylated deoxyuridine analogues with anti-malarial activity. *Malar. J.* **18**, 392 (2019).

Acknowledgements

We thank the patients and staff of Kilifi County Hospital and the KEMRI-Wellcome Trust Research Programme, Kilifi for their help with this study, members of the Human Genetics Group in Kilifi for help with sample collection and processing, and the patients and staff at the Paediatric Department of the Royal Victoria Hospital in Banjul, Gambia for their help with the study. The human genetic data from severe cases and controls used in this study have previously been reported by the Malaria Genomic Epidemiology Network, and we thank all our colleagues who contributed to

this previous work as part of MalariaGEN Consortial Project 1. A full list of consortium members is provided at <https://www.malariagen.net/projects/consortial-project-1/malariagen-consortium-members>. The MalariaGEN Pf6 open resource¹¹ was generated through the Malaria Genomic Epidemiology Network *Plasmodium falciparum* Community Project (<https://www.malariagen.net/resource/26>). The Malaria Genomic Epidemiology Network study of severe malaria was supported by Wellcome (<https://wellcome.ac.uk/>) (WT077383/Z/05/Z (MalariaGEN)) and the Bill & Melinda Gates Foundation (<https://www.gatesfoundation.org/>) through the Foundations of the National Institutes of Health (<https://fnih.org/>) (566 (MalariaGEN)) as part of the Grand Challenges in Global Health Initiative. The Resource Centre for Genomic Epidemiology of Malaria is supported by Wellcome (090770/Z/09/Z and 204911/Z/16/Z (MalariaGEN)). This research was supported by the Medical Research Council (<https://mrc.ukri.org/>) (G0600718, G0600230 and MR/M006212/1 (MalariaGEN)). Wellcome also provides core awards to the Wellcome Centre for Human Genetics (203141/Z/16/Z (WCHG)) and the Wellcome Sanger Institute (206194 (WSI)). Genome sequencing was carried out at the Wellcome Sanger Institute and we thank the staff of the Wellcome Sanger Institute Sample Logistics, Sequencing, and Informatics facilities for their contribution. T.N.W. is supported through a Senior Fellowship from Wellcome (202800/Z/16/Z). This paper is published with permission from the Director of the Kenya Medical Research Institute (KEMRI). Sequencing and initial analysis of the Mali uncomplicated malaria cases was funded by NIAID (R21AI125988 to S.M.T.). J.W.S. received support from the National Center for Advancing Translational Sciences (UL1TR002553). This research was funded in whole or in part by Wellcome as detailed above. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank D. Goldberg (Washington University in St.Louis) and M. Lee (Wellcome Sanger Institute) for their assistance with revisions.

Author information

Affiliations

1. Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

Gavin Band, Christina Hubbart, Anna E. Jeffreys, Kate Rowlands, Kirk A. Rockett & Dominic P. Kwiatkowski

2. Wellcome Sanger Institute, Hinxton, Cambridge, UK

Gavin Band, Ellen M. Leffler, Thuy Nguyen, Sónia Gonçalves, Cristina V. Ariani, Jim Stalker, Richard D. Pearson, Roberto Amato, Eleanor Drury, Kirk A. Rockett & Dominic P. Kwiatkowski

3. Big Data Institute, Li Ka Shing Centre for Health and Information Discovery, University of Oxford, Oxford, UK

Gavin Band, Richard D. Pearson & Dominic P. Kwiatkowski

4. Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT, USA

Ellen M. Leffler

5. Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical Medicine, Fajara, The Gambia

Muminatou Jallow, Fatoumatta Sisay-Joof, Giorgio Sirugo, Umberto d'Alessandro, Kalifa A. Bojang & David J. Conway

6. Edward Francis Small Teaching Hospital (formerly Royal Victoria Teaching Hospital), Banjul, The Gambia

Muminatou Jallow

7. KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya

Carolyne M. Ndila, Alexander W. Macharia, Kevin Marsh, Norbert Peshu & Thomas N. Williams

8. Division of Translational Medicine and Human Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Giorgio Sirugo

9. Nuffield Department of Medicine, University of Oxford, Oxford, UK

Kevin Marsh

10. Division of Infectious Diseases, Duke University School of Medicine, Durham, NC, USA

Joseph W. Saelens & Steve M. Taylor

11. Malaria Research and Training Center, University of Sciences, Techniques, and Technologies of Bamako, Bamako, Mali

Mahamadou Diakité

12. Duke Global Health Institute, Duke University, Durham, NC, USA

Steve M. Taylor

13. Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK

David J. Conway

14. Institute for Global Health Innovation, Department of Surgery and Cancer, Imperial College London, London, UK

Thomas N. Williams

Contributions

Conceptualization: G.B., E.M.L., T.N.W., K.A.R. and D.P.K. Data curation: G.B., E.M.L., T.N., M.J., C.M.N., R.D.P., R.A. and K.A.R. Formal analysis: G.B., E.M.L. and K.A.R. Funding acquisition: D.P.K. Investigation: C.H., A.E.J., K.R., E.D. and K.A.R. Methodology: G.B., K.A.R. and D.P.K.

Project administration: S.G., E.D., K.A.R. and D.P.K. Resources: S.G., E.D., J.S., C.V.A., R.A., R.D.P., M.J., F.S.-J., K.A.B., G.S., C.M.N., A.W.M., N.P., J.W.S., M.D., S.M.T., C.H., A.E.J., K.R., E.D. and K.A.R. Software and visualisation: G.B. Supervision: D.J.C., U.d.A., K.M., T.N.W., S.G., K.A.R. and D.P.K. Writing: G.B., E.M.L., S.M.T., D.J.C., T.N.W., K.A.R. and D.P.K. in collaboration with all authors.

Corresponding authors

Correspondence to [Gavin Band](#), [Kirk A. Rockett](#) or [Dominic P. Kwiatkowski](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Daniel Falush, John Lees and the other, anonymous reviewers 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 Flowchart showing generation and processing of *P.falciparum* \(*Pf*\) sequence data from 5,096 severe malaria cases.](#)

Flowchart shows sample processing from initial selection for whole DNA and Selective Whole genome Amplification (SWGA) pipelines (top) to the curated analysis datasets (bottom). Numbers in each box show counts of severe cases in The Gambia (blue) and Kenya (orange) with the number of individuals sequenced multiple times indicated in brackets. Following curation and QC of data (large box), available *Pf* data was intersected with two existing human genotype datasets to for the analyses described in main text. The combined pf/human imputed dataset has 3,346 samples and the combined pf/human direct typing dataset contains 4,071 individuals. These two datasets have substantial overlap; 825 individuals were represented in the directly-typed data but not the imputed data and were used for replication.

Extended Data Fig. 2 Evidence for association with HbS in three regions of the *Pf* genome using directly-typed HbS genotypes.

Points show evidence for association with HbS (\log_{10} Bayes Factor for test in $N = 4,071$ samples, y axis) based on direct typing of HbS for variants in the *Pfsa1*, *Pfsa2* and *Pfsa3* regions of the *Pf* genome (panels). Variants which alter protein coding sequence are denoted by plusses, while other variants are denoted by circles. Results are computed by logistic regression including an indicator of country as a covariate and assuming an additive model of association; missing and mixed Pf genotype calls were excluded. A corresponding plot using imputed HbS genotypes can be found in Fig. 1. The variant with the strongest association in each region is annotated and the panels show regions of length 50kb centred at this variant. Below, regional genes are annotated, with gene symbols given where the gene has an ascribed name in the PlasmoDB annotation (after removing 'PF3D7_' from the name where relevant); the three genes containing the most-associated variants are shown in red.

Extended Data Fig. 3 Odds ratios for association of HbS with the *Pfsa* variants in severe malaria cases.

Plot shows parameter estimates (points) and 95% posterior credible intervals (horizontal line segments) for the association of HbS with *Pf* genotype at each of the three *Pfsa* lead variants (columns), using several combinations of sample subsets and covariates (rows) in The Gambia and Kenya. Estimates are computed separately for each SNP using logistic regression with the given covariates included as fixed-effect terms, and are based on directly-typed HbS genotypes assuming a dominance model of HbS on *Pf* genotype. Samples with mixed *Pf* genotype calls are excluded from the regression. All estimates are made using a weakly-informative log-F(2,2) prior ([Supplementary Methods](#)) on the genetic effect; a diffuse log-F(0.08,0.08) prior is also applied to covariate effects. Row names are as follows: "Discovery": samples with human genome-wide imputed data that were included in our initial scan (Fig. 1); "Replication": the 825 additional samples that are not closely related to discovery samples (as determined previously⁸); "Combined": all samples with direct typing (as in Fig. 2 and Extended Data Fig. 2); "technical": indicators of sequencing performance including indicator of SWGA or whole DNA sequencing method for the sample, sequence read depth, insert size, and proportion of mixed genotype calls; "SM subtype": indicator of clinical presentation (cerebral malaria, severe malarial anaemia or other severe malaria) the individual was ascertained with; "Pf PCs": principal components (PCs) computed using all called biallelic SNPs having minor allele frequency at least 1% in each population and thinned to exclude variants closer than 1kb; additional rows are shown for PCs computed after excluding SNPs in chromosomes 2 and 11, or from the three regions of association shown in Fig. 1 plus a 25kb margin. Numbers to the right of each estimate show the total regression sample size, the number of samples having the non-reference allele at the given *Pf* SNP, and the number heterozygous or homozygous for HbS.

Extended Data Fig. 4 Allele read ratio versus HbS genotype at the three HbS-associated loci.

For each sample (points) and each of the three HbS-associated loci (rows), the figure shows the proportion of sequencing reads that carry the nonreference allele (y axis). Points are separated by country (columns) and HbS genotype (x axis); the x axis values are jittered to visually separate.

The called *Pf* genotype of each sample is indicated by the shape, with mixed calls indicated by squares.

Extended Data Fig. 5 *Pfsa*+ allele frequency and sample size by year of ascertainment.

a) points show the sample allele frequency (y axis) for each *Pfsa* variant (rows) in severe malaria cases by year of ascertainment (x axis) and country (colour). Vertical line segments show the 95% confidence interval corresponding to each estimate. Horizontal dashed lines show the overall estimate across all years in our data, as in Fig. 3a. b) Bars show the total number of severe case samples in our dataset (y axis) in each country (colour) by year of ascertainment (x axis).

Extended Data Fig. 6 *Pfsa3*+ genotypes are correlated with *PF3D7_1127000* transcript levels in trophozoite-stage infections.

Plot shows the ratio of RNA-seq reads carrying the non-reference allele at the chr11:1,057,437 T > C mutation (x axis) against the estimated transcript abundance of *PF3D7_1127000* (log2 TPM, y axis), for 32 children from Mali ascertained with HbAA or HbAS genotype (colours, as detailed in the legend). Underlying data is as published by Saelens et al¹⁶ and is further detailed in Supplementary Table 6. The plot is separated by parasite stage (panel labels) as previously inferred¹⁶. Among trophozoite-stage infections, we noted one infection of an HbAS individual (AS08) that has low expression of *PF3D7_1127000* (TPM = 28.3); this sample appears to have *Pfsa3*- genotype although we caution that only two reads with reasonable mapping quality were observed. Conversely, one ring-stage infection of an HbAA genotype individual (AA01) has relatively high expression (TPM = 253.2) of *PF3D7_1127000*; this sample is likely mixed as it appears to express gene copies with both *Pfsa3*- and *Pfsa3*+ genotypes, with *Pfsa3*+ predominant.

Extended Data Fig. 7 Estimated abundance of *Pfsa* region gene transcripts from *in vitro* intraerythrocytic time course

experiments.

Plot shows the estimated relative transcript abundance (\log_{10} TPM, y axis) against hours post-infection of erythrocytes (x axis) for the three *Pfsa* region genes containing nonsynonymous sickle-associated polymorphisms (Supplementary Table 1). Data is from three studies which analysed the 3D7 isolate (Otto et al⁵⁸, Wickers et al⁵⁹ and Saelens et al¹⁶) as indicated by columns; the Saelens et al study also analysed the FUP/H isolate (dashed lines). Time points can be roughly interpreted⁵⁸ as: ring stage (~0-16h post-invasion); trophozoite stage (16-40h post-invasion); schizont stage (40-48h post-invasion). Replicate experiments are indicated by multiple lines in each panel; colours indicate the HbS genotype of the erythrocytes used as noted in the legend. TPM values are estimated based on reads aligning to the 3D7 reference genome; for the Saelens et al study these were used as reported previously while for the other studies we recomputed TPM as described in **Methods**.

Extended Data Fig. 8 Estimated abundance of all transcripts in 3D7 and FUP/H parasites across the intraerythrocytic time course.

Plot shows the estimated relative transcript abundance (TPM) of *Pfalciparum* genes measured in 3D7 (x axis) and in FUP/H (Uganda Palo Alto, y axis) parasites, using the data reported by Saelens et al¹⁶. Transcript abundance is measured *in vitro* using erythrocytes from two HbAA genotype individuals (rows), and at multiple time points post-infection (columns). TPM is measured by alignment to the 3D7 genome followed transcript quantification as described by Saelens et al¹⁶. The genes *PF3D7_0215300* (*PfACS8*, *Pfsa1* locus), *PF3D7_0220300* (*Pfsa2* locus), and *PF3D7_1127000* (*Pfsa3* locus) are denoted by coloured points as shown in the legend. Both *PF3D7_1127000* and to a lesser extent *PF3D7_0220300* show an increase in expression at trophozoite stage in FUP/H parasites. We determined the genotypes of FUP/H. We determined the *Pfsa* genotypes of FUP/H as + + + (using the notation of Fig. 2) by aligning available short-read sequencing reads (SRA accessions SRR530503, SRR629055, and SRR629078; Broad Institute 2014).

Extended Data Fig. 9 Structural variation at the *Pfsa3* locus.

Plot shows all DNA segments of length 50 (50-mers) that are shared identically between the 3D7 genome assembly (x axis) and CD01 genome assembly³⁸ (y axis) in the *Pfsa3* region. Points near the diagonal indicate similar structure, while sequences of off-diagonal points indicate structural differences between genomes. Coloured regions indicate approximate regions of 3D7 that contain increased copy number (light blue) or deletions (light yellow) in CD01 relative to 3D7. Segment endpoints are determined by inspection of shared kmer locations and are: 1,053,925 - 1055024 (duplication); 1,055,395-1,055,784 (deletion); 1,058,765-1,059,087 (deletion); 1,059,675-1,059,777 (triplication). The CD01 assembly carries *Pfsa1*+, *Pfsa2*+ and *Pfsa3*+ alleles. Comparisons of 3D7 to other available assembled *Pf* genomes in *Pfsa* regions can be found in Supplementary Fig. 3.

Supplementary information

Supplementary Information

This file contains Supplementary text, Figs. 1–6, Methods and references.

Reporting Summary

Peer Review File

Supplementary Tables 1–8

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

Band, G., Leffler, E.M., Jallow, M. *et al.* Malaria protection due to sickle haemoglobin depends on parasite genotype. *Nature* **602**, 106–111 (2022). <https://doi.org/10.1038/s41586-021-04288-3>

- Received: 30 March 2021
- Accepted: 29 November 2021
- Published: 09 December 2021
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04288-3>

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/s41586-021-04288-3>

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

- Article
- [Published: 19 January 2022](#)

Androgens increase excitatory neurogenic potential in human brain organoids

- [Iva Kelava](#) ¹nAff³,
- [Ilaria Chiaradia](#)¹,
- [Laura Pellegrini](#)¹,
- [Alex T. Kalinka](#)² &
- [Madeline A. Lancaster](#) [ORCID: orcid.org/0000-0003-2324-8853](#)¹

[Nature](#) volume 602, pages 112–116 (2022)

- 8013 Accesses
- 115 Altmetric
- [Metrics details](#)

Subjects

- [Developmental neurogenesis](#)
- [Neural stem cells](#)

Abstract

The biological basis of male–female brain differences has been difficult to elucidate in humans. The most notable morphological difference is size, with male individuals having on average a larger brain than female

individuals^{1,2}, but a mechanistic understanding of how this difference arises remains unknown. Here we use brain organoids³ to show that although sex chromosomal complement has no observable effect on neurogenesis, sex steroids—namely androgens—lead to increased proliferation of cortical progenitors and an increased neurogenic pool. Transcriptomic analysis and functional studies demonstrate downstream effects on histone deacetylase activity and the mTOR pathway. Finally, we show that androgens specifically increase the neurogenic output of excitatory neuronal progenitors, whereas inhibitory neuronal progenitors are not increased. These findings reveal a role for androgens in regulating the number of excitatory neurons and represent a step towards understanding the origin of sex-related brain differences in humans.

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: Androgens lead to expanded basal progenitors through increased proliferation of radial glial stem cells.

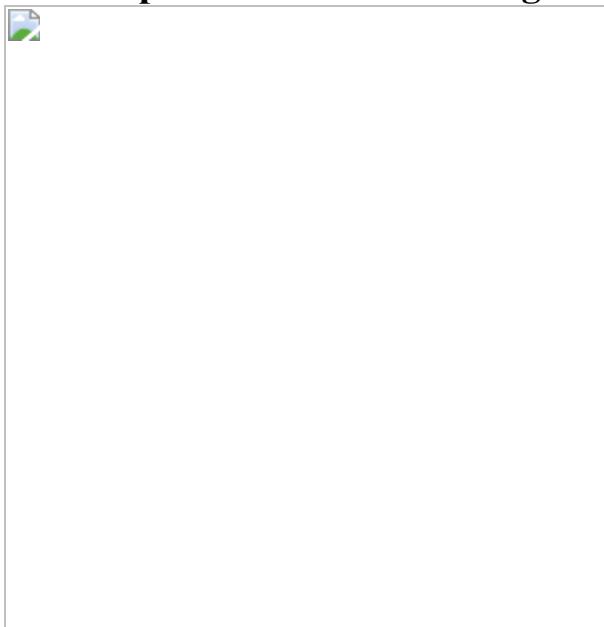
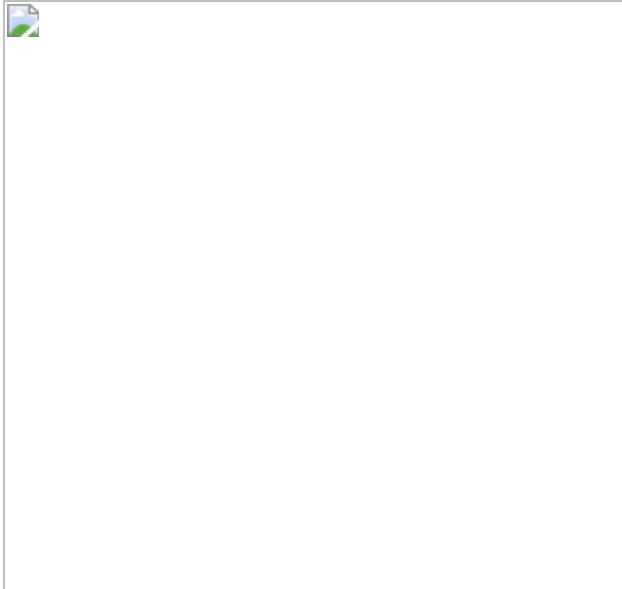


Fig. 2: Transcriptomics reveals the involvement of HDACs and mTOR in the androgen-mediated phenotype.



Fig. 3: Specific increase in excitatory neurogenesis after androgen surge.



Data availability

All sequencing data (bulk and single cell) have been deposited at the NCBI Gene Expression Omnibus (GEO) under accession code [GSE187877](#). The following open access databases were used in this study: SZDB: A Database for Schizophrenia Genetic Research (<http://szdb.org/>) and SFARI gene (<https://gene.sfari.org/>).

Code availability

Delboy code is available at <https://github.com/alextkalinka/delboy>.

References

1. 1.
Ritchie, S. J. et al. Sex differences in the adult human brain: evidence from 5216 UK Biobank participants. *Cereb. Cortex* **28**, 2959–2975 (2018).
2. 2.
Ruigrok, A. N. V. et al. A meta-analysis of sex differences in human brain structure. *Neurosci. Biobehav. Rev.* **39**, 34–50 (2014).
3. 3.
Lancaster, M. A. et al. Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
4. 4.
Green, T., Flash, S. & Reiss, A. L. Sex differences in psychiatric disorders: what we can learn from sex chromosome aneuploidies. *Neuropsychopharmacol.* **44**, 9–21 (2018).
5. 5.
McCarthy, M. M. Multifaceted origins of sex differences in the brain. *Phil. Trans. R. Soc. B* **371**, 20150106 (2016).
6. 6.
Loomes, R., Hull, L. & Mandy, W. P. L. What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *J. Am. Acad. Child Adolesc. Psychiatry* **56**, 466–474 (2017).
7. 7.
Abel, K. M., Drake, R. & Goldstein, J. M. Sex differences in schizophrenia. *Int. Rev. Psychiatry* **22**, 417–428 (2010).
8. 8.

Hines, M. Neuroscience and sex/gender: looking back and looking forward. *J. Neurosci.* **40**, 37–43 (2020).

9. 9.

McCarthy, M. M. & Arnold, A. P. Reframing sexual differentiation of the brain. *Nat. Neurosci.* **14**, 677–683 (2011).

10. 10.

Woodson, J. C. & Gorski, R. A. in *Sexual Differentiation of the Brain* (ed. Matsumoto, A.) Ch. 13 (CRC Press, 1999).

11. 11.

Rabinowicz, T., Dean, D. E., Petetot, J. M.-C. & Courten-Myers, G. M. D. E. Gender differences in the human cerebral cortex: more neurons in males; more processes in females. *J. Child Neurol.* **14**, 98–107 (1999).

12. 12.

Knickmeyer, R. C. et al. Impact of sex and gonadal steroids on neonatal brain structure. *Cereb. Cortex* **24**, 2721–2731 (2014).

13. 13.

Arnold, A. P. A general theory of sexual differentiation. *J. Neurosci. Res.* **95**, 291–300 (2017).

14. 14.

Arnold, A. P. The organizational–activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Horm. Behav.* **55**, 570–578 (2009).

15. 15.

McCarthy, M. M. Estradiol and the developing brain. *Physiol. Rev.* **88**, 91–134 (2008).

16. 16.

Wallen, K. Hormonal influences on sexually differentiated behavior in nonhuman primates. *Front. Neuroendocrin.* **26**, 7–26 (2005).

17. 17.

Miller, W. L. & Auchus, R. J. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr. Rev.* **32**, 81–151 (2011).

18. 18.

Pollen, A. A. et al. Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015).

19. 19.

Davey, R. A. & Grossmann, M. Androgen receptor structure, function and biology: from bench to bedside. *Clin. Biochem. Rev.* **37**, 3–15 (2016).

20. 20.

Wang, F. et al. RNAscope: a novel *in situ* RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* **14**, 22–29 (2012).

21. 21.

Quartier, A. et al. Genes and pathways regulated by androgens in human neural cells, potential candidates for the male excess in autism spectrum disorder. *Biol. Psychiat.* **84**, 239–252 (2018).

22. 22.

Qu, Y. et al. Constitutively active AR-V7 plays an essential role in the development and progression of castration-resistant prostate cancer. *Sci. Rep.* **5**, 7654 (2015).

23. 23.

Jeselsohn, R. et al. Emergence of constitutively active estrogen receptor- α mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin. Cancer Res.* **20**, 1757–1767 (2014).

24. 24.

Kalinka, A. T. Improving the sensitivity of differential-expression analyses for under-powered RNA-seq experiments. Preprint at <https://doi.org/10.1101/2020.10.15.340737> (2020).

25. 25.

Tang, T. et al. HDAC1 and HDAC2 regulate intermediate progenitor positioning to safeguard neocortical development. *Neuron* **101**, 1117–1133 (2019).

26. 26.

Li, L., Jin, J. & Yang, X.-J. Histone deacetylase 3 governs perinatal cerebral development via neural stem and progenitor cells. *Iscience* **20**, 148–167 (2019).

27. 27.

Dey, A. et al. YB-1 is elevated in medulloblastoma and drives proliferation in Sonic hedgehog-dependent cerebellar granule neuron progenitor cells and medulloblastoma cells. *Oncogene* **35**, 4256–4268 (2016).

28. 28.

Kielar, M. et al. Mutations in Eml1 lead to ectopic progenitors and neuronal heterotopia in mouse and human. *Nat. Neurosci.* **17**, 923–933

(2014).

29. 29.

Abrahams, B. S. et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36–36 (2013).

30. 30.

Hackinger, S. et al. Evidence for genetic contribution to the increased risk of type 2 diabetes in schizophrenia. *Transl. Psychiatry* **8**, 252 (2018).

31. 31.

Udawela, M. et al. SELENBP1 expression in the prefrontal cortex of subjects with schizophrenia. *Transl. Psychiatry* **5**, e615 (2015).

32. 32.

Fatemi, S. H., Folsom, T. D. & Thuras, P. D. Deficits in GABA_B receptor system in schizophrenia and mood disorders: a postmortem study. *Schizophr. Res.* **128**, 37–43 (2011).

33. 33.

Narayan, S., Head, S. R., Gilmartin, T. J., Dean, B. & Thomas, E. A. Evidence for disruption of sphingolipid metabolism in schizophrenia. *J. Neurosci. Res.* **87**, 278–288 (2009).

34. 34.

Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).

35. 35.

Göttlicher, M. et al. Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. *EMBO J.* **20**, 6969–6978 (2001).

36. 36.

Boissinot, M. et al. Induction of differentiation and apoptosis in leukaemic cell lines by the novel benzamide family histone deacetylase 2 and 3 inhibitor MI-192. *Leukemia Res.* **36**, 1304–1310 (2012).

37. 37.

Saito, A. et al. A synthetic inhibitor of histone deacetylase, MS-27-275, with marked in vivo antitumor activity against human tumors. *Proc Natl Acad. Sci. USA* **96**, 4592–4597 (1999).

38. 38.

Alexeyenko, A. et al. Comparative interactomics with Funcoup 2.0. *Nucleic Acids Res.* **40**, D821–D828 (2012).

39. 39.

Mayer, C. & Grummt, I. Ribosome biogenesis and cell growth: mTOR coordinates transcription by all three classes of nuclear RNA polymerases. *Oncogene* **25**, 6384–6391 (2006).

40. 40.

Kim, W.-Y. Brain size is controlled by the mammalian target of rapamycin (mTOR) in mice. *Commun. Integr. Biol.* **8**, e994377 (2015).

41. 41.

Saxton, R. A. & Sabatini, D. M. mTOR signaling in growth, metabolism, and disease. *Cell* **169**, 361–371 (2017).

42. 42.

Voss, M. H. et al. Phase 1 study of mTORC1/2 inhibitor sapanisertib (TAK-228) in advanced solid tumours, with an expansion phase in renal, endometrial or bladder cancer. *Br. J. Cancer* **123**, 1590–1598 (2020).

43. 43.

Choi, Y. J. et al. Inhibitory effect of mTOR activator MHY1485 on autophagy: suppression of lysosomal fusion. *PLoS ONE* **7**, e43418 (2012).

44. 44.

Sohal, V. S. & Rubenstein, J. L. R. Excitation–inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Mol. Psychiatr.* **24**, 1248–1257 (2019).

45. 45.

Marín, O. & Müller, U. Lineage origins of GABAergic versus glutamatergic neurons in the neocortex. *Curr. Opin. Neurobiol.* **26**, 132–141 (2014).

46. 46.

O’Shaughnessy, P. J. et al. Alternative (backdoor) androgen production and masculinization in the human fetus. *PLoS Biol.* **17**, e3000002 (2019).

47. 47.

Martínez-Cerdeño, V., Noctor, S. C. & Kriegstein, A. R. Estradiol stimulates progenitor cell division in the ventricular and subventricular zones of the embryonic neocortex. *Eur. J. Neurosci.* **24**, 3475–3488 (2006).

48. 48.

Eliot, L., Ahmed, A., Khan, H. & Patel, J. Dump the “dimorphism”: comprehensive synthesis of human brain studies reveals few male–female differences beyond size. *Neurosci. Biobehav. Rev.* **125**, 667–697 (2021).

49. 49.

Knickmeyer, R. C. & Baron-Cohen, S. Fetal testosterone and sex differences. *Early Hum. Dev.* **82**, 755–760 (2006).

50. 50.

Bahari-Javan, S. et al. HDAC1 links early life stress to schizophrenia-like phenotypes. *Proc. Natl Acad. Sci. USA* **114**, E4686–E4694 (2017).

51. 51.

Ryskalin, L., Limanaqi, F., Frati, A., Busceti, C. L. & Fornai, F. mTOR-related brain dysfunctions in neuropsychiatric disorders. *Int. J. Mol. Sci.* **19**, 2226 (2018).

52. 52.

Lancaster, M. A. et al. Guided self-organization and cortical plate formation in human brain organoids. *Nat. Biotechnol.* **35**, 659–666 (2017).

53. 53.

Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).

54. 54.

Bagley, J. A., Reumann, D., Bian, S., Lévi-Strauss, J. & Knoblich, J. A. Fused cerebral organoids model interactions between brain regions. *Nat. Methods* **14**, 743–751 (2017).

55. 55.

Hines, M., Constantinescu, M. & Spencer, D. Early androgen exposure and human gender development. *Biol. Sex Differ.* **6**, 3 (2015).

56. 56.

Reyes, F. I., Boroditsky, R. S., Winter, J. D. S. & Faiman, C. Studies on human sexual development. II. Fetal and maternal serum gonadotropin and sex steroid concentrations. *J. Clin. Endocrinol. Metab.* **38**, 612–617 (1974).

57. 57.

McManus, J. M. & Sharifi, N. Structure-dependent retention of steroid hormones by common laboratory materials. *J. Steroid Biochem. Mol. Biol.* **198**, 105572 (2020).

58. 58.

Shoskes, J. J., Wilson, M. K. & Spinner, M. L. Pharmacology of testosterone replacement therapy preparations. *Transl. Androl. Urol.* **5**, 834–843 (2016).

59. 59.

Wright, A. S., Thomas, L. N., Douglas, R. C., Lazier, C. B. & Rittmaster, R. S. Relative potency of testosterone and dihydrotestosterone in preventing atrophy and apoptosis in the prostate of the castrated rat. *J. Clin. Invest.* **98**, 2558–2563 (1996).

60. 60.

Iacopino, F. et al. Valproic acid activity in androgen-sensitive and -insensitive human prostate cancer cells. *Int. J. Oncol.* **32**, 1293–1303 (1992).

61. 61.

Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

62. 62.

Giandomenico, S. L. et al. Cerebral organoids at the air–liquid interface generate diverse nerve tracts with functional output. *Nat. Neurosci.* **22**, 669–679 (2019).

63. 63.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

64. 64.

Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

65. 65.

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

66. 66.

Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).

67. 67.

Gerard, D. Data-based RNA-seq simulations by binomial thinning. *BMC Bioinformatics* **21**, 206 (2020).

68. 68.

Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).

69. 69.

Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

70. 70.

Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

71. 71.

Benito-Kwiecinski, S. et al. An early cell shape transition drives evolutionary expansion of the human forebrain. *Cell* **184**, 2084–2102 (2021).

72. 72.

Noctor, S. C., Martínez-Cerdeño, V. & Kriegstein, A. R. Distinct behaviors of neural stem and progenitor cells underlie cortical neurogenesis. *J. Comp. Neurol.* **508**, 28–44 (2008).

73. 73.

Giandomenico, S. L., Sutcliffe, M. & Lancaster, M. A. Generation and long-term culture of advanced cerebral organoids for studying later stages of neural development. *Nat. Protoc.* **16**, 579–602 (2021).

Acknowledgements

We thank D. Srivastava, S. Munro and M. Hegde for comments; the MRC LMB Light Microscopy facility for help with microscopy and

Bioinformatics (P. Freire-Pritchett) for help with RNA-seq analysis, as well as the CRUK Genomics facility for library preparation and sequencing; M. Mancini, M. Marcelli and E. Wilson for depositing their plasmids in Addgene; and the other members of the Lancaster laboratory and the MRC LMB Cell Biology division for discussions. Work in the Lancaster laboratory is supported by the Medical Research Council (MC_UP_1201/9) and the European Research Council (ERC STG 757710).

Author information

Author notes

1. Iva Kelava

Present address: Wellcome Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK

Affiliations

1. Medical Research Council Laboratory of Molecular Biology, Cambridge, UK

Iva Kelava, Ilaria Chiaradia, Laura Pellegrini & Madeline A. Lancaster

2. Milner Therapeutics Institute, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge, UK

Alex T. Kalinka

Contributions

I.K. conceived the study, performed experiments, analysed data and wrote the paper. I.C. performed experiments and analysed the data. L.P. performed experiments. A.T.K. performed bioinformatics analysis. M.A.L. supervised the study, analysed data and co-wrote the paper.

Corresponding authors

Correspondence to [Iva Kelava](#) or [Madeline A. Lancaster](#).

Ethics declarations

Competing interests

M.A.L. is an inventor on several patents related to cerebral organoids, is co-founder and member of the scientific advisory board of a:head bio, and is a member of the scientific advisory board of the Roche Institute for Translational Bioengineering.

Peer review information

Nature thanks Giorgia Quadrato 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 Treatment and phenotypic characterization of cerebral organoids.](#)

a) Illustration of the determination of the progenitor zone for quantifications. The progenitor zone (VZ+SVZ) was determined based on immunostaining and histological landmarks (see [Methods](#)). Immunostaining of XX 35d organoids stained for TBR2 (yellow, white), co-stained with DAPI (blue). Dashed line indicates the progenitor zone. Schematic of an organoid ventricle with different progenitor populations of interest. **b)** Quantification of thicknesses of different morphological zones between male (XY) and female (XX) brain organoids at 35d (above) and 52d

(below). Significance values (Mann-Whitney, two-tailed) of the measurements from XX organoids, as compared to XY organoids, are indicated by different shades of grey. VZ-ventricular zone, SVZ-subventricular zone. **c**) Actual measured concentration of testosterone (left) and oestradiol (right) in the medium after the addition of 100nM (calculated) on day 0, as measured by ELISA assay. **d**) Actual measured concentration of testosterone over 4 days of organoid culture after addition of 100nM (calculated) on day 0, as measured by ELISA. **e**) Sections of whole XX organoids at 35d (left), and XY organoids at 52d (right), stained for TBR2 (white). Co-stained with DAPI (blue). **f**) XY 35d organoids stained for TBR2 (yellow, white), co-stained with DAPI (blue), treated with DHT, T or E. **g**) Quantification of non-apical mitotic, PH3+ cells in XY and XX, 35d and 52d organoids, treated with DHT, T and E. Scale bars: **e**) 500 μ m, **a**), **f**) 100 μ m. See [Methods](#) for details of statistics and Supplementary Table [5](#) for details of *n* numbers.

Extended Data Fig. 2 Basal progenitors are increased after androgen treatment.

a) Quantification of TBR2+ cells per mm² progenitor zone (see Extended Data Fig. [1a](#)) at 35d. **b)** Image of the ventricular zone/subventricular zone (VZ/SVZ) border in XX 35d organoids stained for Ki67 (green, white) and TBR2 (white). Cells positive for both Ki67 and TBR2 are indicated with magenta arrowheads. Double positive mitotic cells are indicated with yellow arrowheads. A white dashed line delineates VZ/SVZ border with the ventricular zone below. **c)** Immunostaining for proliferation marker Ki67 (white) on XX 35d control (left) and DHT treated (right) organoids. Dashed yellow lines represent the apical surface (bottom line) and the VZ/SVZ boundary. Note the difference in Ki67+ cells in the SVZ of DHT-treated organoids. **d)** Quantification of Ki67/TBR2 double positive cells, out of all TBR2+ cells, in XX 35d and 52d organoids. **e)** Immunostaining for HOPX (fire LUT) of XX 52d organoids. Yellow dashed line indicates the ventricular surface. Note the HOPX signal in radial glia. Images are single, 1.2 μ m optical planes. **f)** Quantification of HOPX+ basal radial glia per mm² SVZ of XX 52d organoids. **g)** Immunostaining for phosphorylated histone H3 (PH3) (green) on XX 35d and XY 52d old organoids. Co-stained with DAPI (white). **h)** Quantification of apical mitotic cells (PH3+),

normalized per mm length of ventricle in male (XY) and female (XX) organoids, at 35d and 52d, treated with DHT, T and E. **i**) Quantification of ventricular length in male (XY) and female (XX) organoids, at 35d and 52d, treated with DHT, T and E. Scale bars: **g**) 100 μ m, **e**) 50 μ m, **b**), **c**) 20 μ m. See [Methods](#) for details of statistics and Supplementary Table [5](#) for details of *n* numbers.

[Extended Data Fig. 3 Clonal labelling reveals increased radial glial proliferation after androgen treatment.](#)

a) Timeline and schematic of the Sendai/Lenti virus lineage tracing and analysis, with virus encoding emGFP injected at day 45 and virus encoding RFP injected at day 53. **b**) Representative images of GFP-labelled clones (white) 8 days post Sendai emGFP labelling in control (above) and DHT-treated (below) XX organoids. DAPI is in blue. **c**) Quantification of GFP+ clone size (left side) at 51d and 53d and RFP+ clone size at 61d. **d**) XX control and DHT organoids injected with Lenti-RFP at 53d, and fixed at 61d. RFP (magenta). Co-stained with DAPI (blue). **e**) Quantification of TBR2+ cells per mm² of progenitor layer in control and DHT-treated XX organoids, at 8 days post-labelling with Sendai emGFP. Note that TBR2+ intermediate progenitors were not yet increased at this time point, as the radial glia which produce them were still in the VZ. Scale bars: **d**) 50 μ m, **b**) 25 μ m. See [Methods](#) for details of statistics and Supplementary Table [5](#) for details of *n* numbers.

[Extended Data Fig. 4 AR activity in radial glia promotes their proliferation rather than differentiation.](#)

a) RNA Scope (fluorescent *in situ*) for AR of XY organoids at 45d. Single AR mRNA puncta (white). Co-stained with DAPI (blue). Yellow dashed lines demarcate ventricular zone (VZ) and the subventricular zone (SVZ). **a'**) Portion of the VZ from a. **a''**) Portion of the SVZ from a. **b**) RNA Scope for AR of XX organoids at 45d. Single AR mRNA puncta (white). Co-stained with DAPI (blue). **b'**) Portion of the VZ from b. **b''**) Portion of the SVZ from b. Note the difference in the amount of AR mRNA in the VZ and SVZ in both cell lines. **c**) RNA Scope for AR of XX 21d (top) and 35d

(bottom) organoids. Single AR mRNA puncta (white). Immunostaining for TBR2 (magenta). Co-stained with DAPI (blue). Yellow dashed line demarcates the apical surface c') Portion of the ventricular zone (VZ) from XX 21d organoid. c") Portion of the VZ from XX 35d organoid. **d)** RNA Scope negative (above) and positive (below) control (fire look up table) in female (XX) 35d organoids. DAPI is in white. **e)** Quantification of AR mRNA distribution in GFP+ pairs of cells, depending on the stage of cell division. **f)** RNAscope of AR mRNA (white puncta), together with GFP signal (green) from EmGFP Sendai Fluorescence Reporter showing examples of three daughter cell pairs. Yellow dashed lines indicate GFP+ cell body. **g)** Western blot for AR on control, DHT and E treated organoids at 17-25 days. AR specific band is predicted to be around 110 kDa. Note the increased AR signal in DHT-treated organoids. Asterisk indicates a lower amount of protein loaded for the 75d organoid lane (see [Methods](#)). For gel source data, see Supplementary Fig. 1. **h)** Quantification of levels of AR protein from **g**), normalized by GAPDH expression. **i)** Droplet digital PCR (ddPCR) results showing the decrease in AR transcription between days 17 and 25. ref = housekeeping gene EIF2B2. **j)** Representative images of the cortical wall at 21, 35 and 52d, with DAPI-labelled nuclei visible, showing the relative reduction in the radial glial progenitor layer (ventricular zone - VZ) and an increase in the thickness of the neuronal layer over time. **k)** Quantification of the percentage (%) of cells (AR+ DAPI), containing AR mRNA puncta (as detected by RNA Scope), out of all cells (DAPI) at 21, 35 and 52d. Cells counted: 21d – 1222, 35d – 1067, 52d – 1201. Co-stained with DAPI (white). **l)** XX organoids electroporated (EP) at 45d and fixed at 50d. EGFP-C1: control; EGFP-C1-AR-V7: constitutively active AR. Immunostaining for GFP and TBR2. Yellow dashed lines demarcate the apical and basal boundaries of the VZ. Note increased GFP+ nuclei in the VZ in EGFP-C1-AR-V7. **m)** Immunostaining for GFP (green) and Ki67 (magenta) on XX 50d organoids, electroporated at 45d. Co-stained with DAPI (blue). EGFP-C1: control plasmid; EGFP-C1-AR-V7: plasmid expressing constitutively active AR; EGFP-C1-ERaY537S: plasmid expressing constitutively active ERa. **n)** Quantification of the proportion of GFP+ cells co-staining for the proliferation marker Ki67 at 5 days post electroporation in XX organoids electroporated with the indicated plasmid at 45d. **o)** Quantification of the proportion of GFP+ cells co-staining for the intermediate progenitor marker TBR2 at 2- and 5 days post electroporation

in XX organoids electroporated at 45d. At 5 days post electroporation, most cells electroporated with EGFP-C1-ERaY537S died indicating a later effect on cell survival. **p)** Quantification of the proportion of GFP+ cells co-staining for the neuronal marker HuC/D 2- and 5 days post electroporation in XX organoids electroporated at 45d. At 5 days post electroporation, most cells electroporated with EGFP-C1-ERaY537S died. **q)** XX 47d organoids, electroporated at 45d. Immunostaining for GFP (green), NGN2 (cyan) and BRN2 (magenta). Co-stained with DAPI (white). Yellow arrowheads: GFP+/NGN2+ cells. White arrowheads: GFP+/BRN2+ cells. Yellow dashed line demarcates the apical surface. Note the increased expression of differentiation markers upon expression of ERaY537S perhaps indicating a premature cell cycle exit and relating to the cell death observed. **r)** Quantification of the proportion of GFP+ cells co-staining for the neurogenic marker NGN2 2 days post electroporation in XX organoids electroporated at 45d. **s)** Quantification of the proportion of GFP+ cells co-staining for the upper layer neurogenesis marker BRN2 2 days post electroporation in XX organoids electroporated at 45d. Scale bars: **a), b), l), m)** 50 μ m, **c), d), q)** 25 μ m, **a'), a''), b'), b'')**, **j)** 20 μ m, **c'), c'')**, **f)** 5 μ m. See [Methods](#) for details of statistics and Supplementary Table [5](#) for details of *n* numbers.

[Extended Data Fig. 5 Live imaging of radial glial division modes after activation and androgen signalling.](#)

a) Still images from live imaging of organoids electroporated at 31d and imaged beginning at 2 days post-electroporation (see [Methods](#) for details of image acquisition). pCAG-mCherry+ cells are shown in magenta, and EGFP-C1-AR-V7+ are shown in green. Green arrowheads: GFP+ cells. Yellow arrowheads: mCherry+/GFP– cells. Filled in arrowheads: one of the tracked daughter cells. **a)** Comparison of the behaviour of mCherry+ and GFP+ only cells at 33d. Basal surface is up. mCherry+ only cell, upon division (06:11) produces two daughter cells, one of which migrates basally, indicating a more differentiated identity, whereas the other stays in the ventricular zone (VZ), representing an example of asymmetric division. GFP+ cell divides (07:28), but both of the daughter cells continue to reside in the VZ. **b)** Still images from live imaging of organoids electroporated at 54d and imaged at 56d. Basal surface is down. After division (15:50), both

of the daughter cells remain in the VZ. Yellow dashed lines demarcate the apical surfaces. Time scale: hours:minutes. **c**) Schematic of the proposed mechanism leading to increased basal progenitors. Radial glia can divide symmetrically, increasing their numbers, or asymmetrically, generating one radial glia and one basal progenitor. Application of androgen supports symmetric, proliferative divisions, thus increasing the size of radial glial clones. After androgen is withdrawn, these radial glia can start producing basal progenitors in increased numbers, as shown in Fig. [1b, c](#). Basal progenitors represented as a single population (yellow cells), for clarity. Scale bars: **a**, **b**) 20 μ m. See [Methods](#) for details of statistics and Supplementary Table [5](#) for details of n numbers.

[Extended Data Fig. 6 Transcriptomics reveals key signalling pathways and cell types.](#)

a) Principle component analysis (PCA) plots of XX 35d bulk RNA-seq, uncorrected for batch effects. **b)** Principle component analysis (PCA) plots of XX 35d bulk RNA-seq, corrected for batch effects using the R package ComBat. Filtering of lowly expressed genes was performed prior to batch correction, resulting in different PC value scales between a and b. See [Methods](#). **c)** Quantification of SRD5A1, an upregulated DEG under androgen treatment. SRD5A1 was quantified as the number of SRD5A1+ puncta per 10 μ m² of the ventricular zone (VZ) in XX 35d control and DHT-treated organoids. **d)** Venn diagram representing the overlap of up- and downregulated DEGs in androgen treated XX 35d organoids with genes differentially expressed in patients with schizophrenia^{[73](#)}. *P* values represent significance of a hypergeometric test of the intersection with up- and downregulated genes ($P=3.6E-4$ and $P=0.0012$, respectively). DE: differentially expressed. **e)** Top significant GO^{[34](#)} term enrichments in XX 35d androgen upregulated genes. **f)** Left, UMAP plot showing different clusters identified by scRNA-seq in 35d old treated and control XX organoids. Right, UMAP plot showing the distribution of cells belonging to control, DHT and E treatments. Note the high degree of overlap indicating reproducibility. **g)** Heat map showing scaled expression levels of cluster specific genes identified through differential gene expression analysis across clusters. The top 10 genes ordered by average log fold change are shown for each cluster. **h)** Dot plot showing the relative expression of cell

type-specific markers. **i**) Feature scatter plot for percent ribosome (percent.ribo) by percent mitochondrial (percent.mt) reads. Left, plotted data points grouped by RG subcluster. Right, plotted data points grouped by treatment status. -0.27 refers to the correlation coefficient. **j**) Pseudotime analysis using Monocle3 mapped onto the reclustered seurat neural clusters (all clusters except ChP/Hem) UMAP. **k**) Feature plot of the neurogenic marker BTG2 (Tis21) and more mature marker synaptotagmin (SYT1) which was also identified as upregulated in RG2 cluster. **l**) Dot plot of the subclustering of the IP/N cell population showing the relative expression of cell type-specific markers and clear separation of different cell subpopulations. ImmN = immature neurons, MatN = maturing neurons. **m**) Stacked bar plot of radial glial (RG) cluster proportions detected by scRNA-seq at 35d. Upper right: chi-square with Monte Carlo simulation for overall distribution (Supplementary Table 4). Note increased RG1 and bRG, but decrease in more committed RG2, with no significant difference compared with expected counts in total RG (combined four RG clusters) and non-RG clusters (IP/N and ChP/Hem) in DHT treatment as determined by chi-square test (not shown) indicating specificity for RG subclusters. * $P<0.05$ and ** $P<0.00005$ **n**) Point range plot of Monte Carlo permutation test of statistical significance between cell cluster proportions in DHT compared with control (left) and E compared with control (right) displaying bootstrapped confidence intervals for the difference in cluster proportion (reported as observed log2 fold-change) (see [Methods](#)). **o**) Dot plot showing the relative expression of bRG markers in different treatments, detected by scRNA-seq, in 35d old organoids. See [Methods](#) for details of statistics and Supplementary Table 5 for details of n numbers.

[Extended Data Fig. 7 HDAC activity interacts with androgen signalling to influence progenitor behaviour.](#)

a) Dot plot showing the relative expression of HDACs in different treatments, detected by scRNA-seq, in 35d old organoids. **b)** UMAP plots showing the relative expression of HDAC2 in cells from control, DHT and E treated organoids. Right - UMAP plot showing the radial glia clusters identified by scRNA-seq. Yellow circle demarcates the bRG subpopulation. Note the increased expression of HDAC2 in DHT bRG cluster. **c)** Immunostaining for TBR2 (green) and HuC/D (magenta) in XX 35d

organoids, treated with VPA from 17d. Co-stained with DAPI (white, blue). c') c'') c''') insets from **c**), as indicated. Yellow asterisks indicate neural rosettes forming above the VZ in some VPA-treated organoids. **d)** Quantification of ventricular length for XX 35d old organoids treated with VPA and VPA+DHT. **e)** Immunostaining for CTIP2 (magenta), and SATB2 (green, white) on XX 35d control and VPA-treated organoids. Co-stained with DAPI (blue). e'), e'') insets from **e**) showing just SATB2 in fire look-up table. e'') Note the SATB2+ cells (yellow arrowheads) and SATB2+ staining present in the VZ in VPA-treated organoids. Yellow dashed lines demarcate the apical and basal boundaries of the VZ. **f)** Quantification of SATB2+ cells in the cortical wall of control, VPA- and VPA+DHT-treated organoids in XX organoids at 35d. **g)** Quantification of the thickness of VZ for XX 35d organoids treated with VPA and VPA+DHT. **h)** Immunostaining for TBR2 (yellow) in 35d old XX organoids with the following treatments: DHT, MI-192 (HDAC2/3 inhibitor), MI-192+DHT, MS-275 (HDAC1/3 inhibitor), MS-275+DHT. White dashed line demarcates the apical surface. **i)** Quantification of TBR2+ cells per mm² of the control and the following treatments: DHT, MI-192 (HDAC2/3 inhibitor), MI-192+DHT, MS-275 (HDAC1/3 inhibitor), MS-275+DHT in 35d old XX organoids. Scale bar: **c)** 100 µm, c'), c''), c'''), **e)**, **h)** 50µm, e'), e'') 20µm. See [Methods](#) for details of statistics and Supplementary Table 5 for details of *n* numbers.

Extended Data Fig. 8 Downstream mTOR activity influences progenitor expansion.

a) Heat map showing scaled expression of differentially expressed genes in DHT-treated compared to control organoids, detected by scRNA-seq. Cell identities were assigned to their treatment group and differential expression analysis performed in Seurat as performed for comparison across clusters (see [Methods](#)). **b)** Dot plot showing the relative expression of bRG markers, overlapping markers of RG1 cells and DE transcripts in DHT, and genes of the mTOR pathway in cells from control, DHT and E treated organoids, detected by scRNA-seq. **c)** Protein interaction network of upregulated DEG directly interacting genes, obtained by bulk RNA seq. Translation/ribosome biogenesis genes are in red. **d)** Immunostaining for phosphorylated S6 (PS6) (fire look-up table), an indicator of mTOR activity, in 35d old XX organoids, treated with DHT and E. White arrowheads indicate PS6+ cell

bodies in the ventricular zone (VZ). **e**) Quantification of PS6+ cells in the VZ of control, DHT- and E-treated 35d old XX organoids. **f**) Immunostaining for TBR2 (white) in control and MHY-1485-treated 35d XX organoids. Co-stained with DAPI (blue). **g**) Quantification of TBR2+ cells per mm² progenitor layer in control and MHY-1485-treated 35d XX organoids. Scale bars: **d), f)** 50μm. See [Methods](#) for details of statistics and Supplementary Table [5](#) for details of *n* numbers.

[Extended Data Fig. 9 Ventral progenitors and mouse organoids exhibit differential responses to those of human excitatory neurogenic progenitors.](#)

a) Schematic of neurogenesis in the human developing brain at ~12 gestation weeks. Excitatory neurons (green arrows) are born within the dorsal cortex and migrate basally to form the cortical plate. Inhibitory neurons (orange arrow) are born in the ventral telencephalon and migrate towards the dorsal side to incorporate themselves into the cortical plate. Yellow: progenitor zone, dark pink: cortical plate. **b)** RNA Scope (fluorescent *in situ* hybridization) for AR mRNA (white) in XX and XY 45d old ventral organoids. Co-stained with DAPI (blue). b') and b'') - portions of the VZ, as indicated. Note the decreased levels compared with Extended Data Fig. [4a,b](#). **c)** Above: Heat map of steroidogenic enzyme expression from XX 35d bulk RNA-seq. Values are in tpm. Below: Schematic of the steroidogenic pathway. Purple – SRD5A1. ESR1, ESR2-estrogen receptors a and b, respectively. **d)** Comparison of SRD5A1 (fire LUT) immunostaining on XX and XY dorsal 35d and ventral 45d organoids. Both anti-SRD5A1 and anti-TBR2 primary antibodies were used in XY samples, and the secondary antibody recognized both primary antibodies but are discernible by their different subcellular localisations. Co-stained with DAPI (white). Yellow dashed lines demarcate the apical and basal boundaries of the VZ. d') Portion of the VZ in XX dorsal 35d organoids stained for SRD5A1. Note the SRD5A1+ puncta. d'') Portion of the VZ in XX ventral 45d organoids stained for SRD5A1. d''') Portion of the VZ in XY dorsal 35d organoids stained for SRD5A1. Note the SRD5A1+ puncta. d''''') Portion of the VZ in XY ventral 45d organoids stained for SRD5A1. White arrowhead - SRD5A1+ punctum. **e)** Timeline of the mouse organoid generation protocol. See [Methods](#) for details. **f)**

Immunostaining for TBR2 (yellow) on 11d old mouse organoids in control, DHT- and E- treated organoids. Co-stained with DAPI (blue). **g)** Quantification of TBR2+ cells per mm² of progenitor layer in mouse organoids at 9d and 11d. Scale bars: **b), d), f)** 50µm, **b'), b'')** 20µm, **d'), d'')**, **d'''), d'''')** 10µm. See [Methods](#) for details of statistics and Supplementary Table [5](#) for details of *n* numbers.

Extended Data Fig. 10 Excitatory neurons are increased after release of androgen signalling.

a) Immunostaining of XX 52d continuously treated control and DHT organoids stained for NEUROD2 (white). **b)** Quantification of NEUROD2+ cells per mm² of cortical wall at 35 days, in XY and XX organoids. **c)** Quantification of NEUROD2+ cells per mm² of cortical wall at 52 days, in XY and XX organoids. **d)** Timeline of the pulse-chase experiment. Hormones were administered between 17-35d, then removed. Organoids were fixed at 52d and 75d. **e)** NEUROD2+ cells per mm² cortical wall after pulse-chase treatment. **f)** Immunostaining for CTIP2 (white) of XX 52d organoids after pulse-chase treatment (control and DHT). Co-stained with DAPI (blue). **g)** Quantification of CTIP2+ cells per mm² cortical wall after pulse-chase treatment. **h)** Quantification of SATB2+ cells per 0.5mm² cortical wall at 75d after pulse-chase treatment. Scale bars: **a)** 100µm, **f)** 50µm. See [Methods](#) for details of statistics and Supplementary Table [5](#) for details of *n* numbers.

Supplementary information

Supplementary Information

This file contains a Supplementary Guide, Supplementary Tables 1–5 and Supplementary Fig. 1.

Reporting Summary

Peer Review File

Supplementary Video 1

Live imaging of XX organoids, electroporated at 31 d, and imaged at 33 d for >48 h. Cells labelled with pCAG-mCherry are in magenta. Cells electroporated with EGFP-C1-AR-V7 are in green. Still images from this video are shown in Extended Data Fig. 5a. Time scale: hours:minutes.

Supplementary Video 2

Live imaging of XX organoids, electroporated at 54 d, and imaged at 56 d for >48 h. Cells labelled with pCAG-mCherry are in magenta. Cells electroporated with EGFP-C1-AR-V7 are in green. Still images from this video are shown in Extended Data Fig. 5b. Time scale: hours:minutes.

Supplementary Video 3

Green channel (EGFP-C1-AR-V7) from Supplementary Video 2 for easier tracking of electroporated, GFP⁺ cells. Time scale: hours:minutes.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Kelava, I., Chiaradia, I., Pellegrini, L. *et al.* Androgens increase excitatory neurogenic potential in human brain organoids. *Nature* **602**, 112–116 (2022). <https://doi.org/10.1038/s41586-021-04330-4>

- Received: 24 October 2020
- Accepted: 09 December 2021
- Published: 19 January 2022

- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04330-4>

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/s41586-021-04330-4>

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

- Article
- [Published: 05 January 2022](#)

A speech planning network for interactive language use

- [Gregg A. Castellucci](#) ORCID: [orcid.org/0000-0001-7311-2829^{1,2}](https://orcid.org/0000-0001-7311-2829),
- [Christopher K. Kovach](#) ORCID: [orcid.org/0000-0002-0117-151X³](https://orcid.org/0000-0002-0117-151X),
- [Matthew A. Howard III³](#),
- [Jeremy D. W. Greenlee](#) ORCID: [orcid.org/0000-0002-8481-8517³](https://orcid.org/0000-0002-8481-8517) &
- [Michael A. Long](#) ORCID: [orcid.org/0000-0002-9283-3741^{1,2}](https://orcid.org/0000-0002-9283-3741)

[Nature](#) volume **602**, pages 117–122 (2022)

- 5580 Accesses
- 1 Citations
- 522 Altmetric
- [Metrics details](#)

Subjects

- [Cognitive control](#)
- [Cooperation](#)

Abstract

During conversation, people take turns speaking by rapidly responding to their partners while simultaneously avoiding interruption^{1,2}. Such interactions display a remarkable degree of coordination, as gaps between

turns are typically about 200 milliseconds³—approximately the duration of an eyeblink⁴. These latencies are considerably shorter than those observed in simple word-production tasks, which indicates that speakers often plan their responses while listening to their partners². Although a distributed network of brain regions has been implicated in speech planning^{5,6,7,8,9}, the neural dynamics underlying the specific preparatory processes that enable rapid turn-taking are poorly understood. Here we use intracranial electrocorticography to precisely measure neural activity as participants perform interactive tasks, and we observe a functionally and anatomically distinct class of planning-related cortical dynamics. We localize these responses to a frontotemporal circuit centred on the language-critical caudal inferior frontal cortex¹⁰ (Broca's region) and the caudal middle frontal gyrus—a region not normally implicated in speech planning^{11,12,13}. Using a series of motor tasks, we then show that this planning network is more active when preparing speech as opposed to non-linguistic actions. Finally, we delineate planning-related circuitry during natural conversation that is nearly identical to the network mapped with our interactive tasks, and we find this circuit to be most active before participant speech during unconstrained turn-taking. Therefore, we have identified a speech planning network that is central to natural language generation during social interaction.

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: Cortical responses during interactive speech.

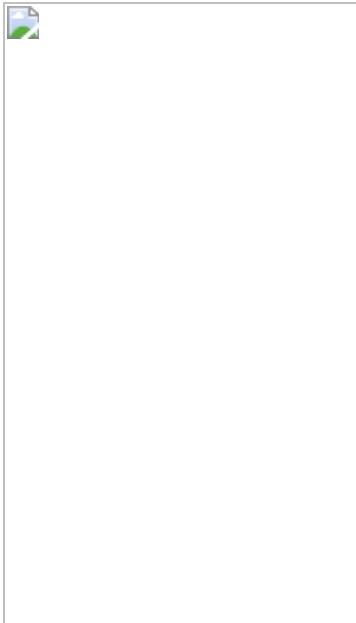


Fig. 2: Spatial organization of speech planning responses.

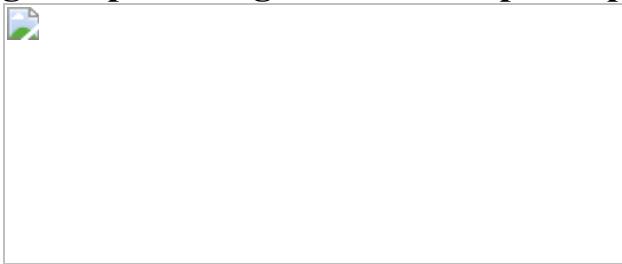


Fig. 3: Linguistic selectivity of planning responses.

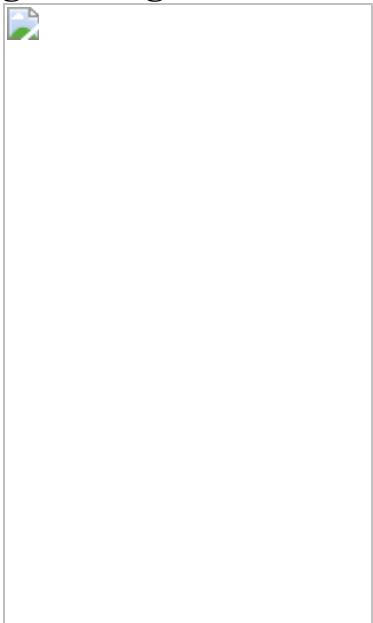
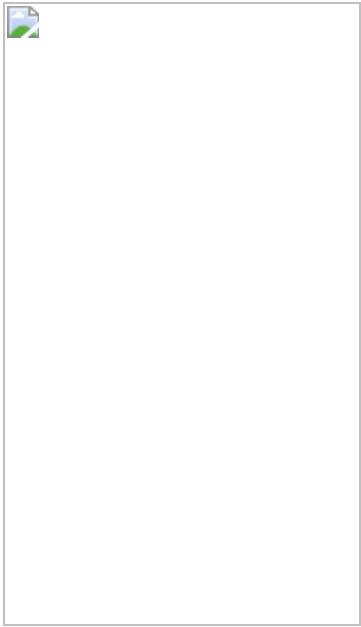


Fig. 4: Planning activity during natural conversation.



Data availability

The data used in these analyses are not publicly available owing to concerns regarding patient privacy; however, the corresponding author will provide deidentified primary data upon request.

Code availability

The corresponding author will provide the MATLAB code used in this study for analysis of ECoG and behavioural data upon request.

References

1. 1.

Sacks, H., Schegloff, E. A. & Jefferson, G. A simplest systematics for the organization of turn-taking for conversation. *Language* **50**, 696–735 (1974).

2. 2.

Levinson, S. C. & Torreira, F. Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* **6**, 731 (2015).

3. 3.

Stivers, T. et al. Universals and cultural variation in turn-taking in conversation. *Proc. Natl Acad. Sci. USA* **106**, 10587–10592 (2009).

4. 4.

Schiffman, H. R. *Sensation and Perception: An Integrated Approach* (Wiley, 2001).

5. 5.

Flinker, A. et al. Redefining the role of Broca's area in speech. *Proc. Natl Acad. Sci. USA* **112**, 2871–2875 (2015).

6. 6.

Basilakos, A., Smith, K. G., Fillmore, P., Fridriksson, J. & Fedorenko, E. Functional characterization of the human speech articulation network. *Cereb. Cortex* **28**, 1816–1830 (2018).

7. 7.

Mirman, D., Kraft, A. E., Harvey, D. Y., Brecher, A. R. & Schwartz, M. F. Mapping articulatory and grammatical subcomponents of fluency deficits in post-stroke aphasia. *Cogn. Affect. Behav. Neurosci.* **19**, 1286–1298 (2019).

8. 8.

Guenther, F. H. *Neural Control of Speech* (MIT, 2016).

9. 9.

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D. & Halgren, E. Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* **326**, 445–449 (2009).

10. 10.

Broca, P. Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bull. Mem. Soc. Anat. Paris* **36**, 330–356 (1861).

11. 11.

Chang, E. F. et al. Pure apraxia of speech after resection based in the posterior middle frontal gyrus. *Neurosurgery* **87**, E383–E389 (2020).

12. 12.

Brass, M. & von Cramon, D. Y. The role of the frontal cortex in task preparation. *Cereb. Cortex* **12**, 908–914 (2002).

13. 13.

Sierpowska, J. et al. Involvement of the middle frontal gyrus in language switching as revealed by electrical stimulation mapping and functional magnetic resonance imaging in bilingual brain tumor patients. *Cortex* **99**, 78–92 (2018).

14. 14.

Levinson, S. C. Turn-taking in human communication—origins and implications for language processing. *Trends Cogn. Sci.* **20**, 6–14 (2016).

15. 15.

Indefrey, P. The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* **2**, 255 (2011).

16. 16.

Schuhmann, T., Schiller, N. O., Goebel, R. & Sack, A. T. The temporal characteristics of functional activation in Broca's area during overt picture naming. *Cortex* **45**, 1111–1116 (2009).

17. 17.

Ferpozzi, V. et al. Broca's area as a pre-articulatory phonetic encoder: gating the motor program. *Front. Hum. Neurosci.* **12**, 64 (2018).

18. 18.

Alario, F. X., Chainay, H., Lehericy, S. & Cohen, L. The role of the supplementary motor area (SMA) in word production. *Brain Res.* **1076**, 129–143 (2006).

19. 19.

Ramanarayanan, V., Goldstein, L., Byrd, D. & Narayanan, S. S. An investigation of articulatory setting using real-time magnetic resonance imaging. *J. Acoust. Soc. Am.* **134**, 510–519 (2013).

20. 20.

Bogels, S., Magyari, L. & Levinson, S. C. Neural signatures of response planning occur midway through an incoming question in conversation. *Sci Rep.* **5**, 12881 (2015).

21. 21.

Ferreira, F. & Swets, B. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *J. Mem. Lang.* **46**, 57–84 (2002).

22. 22.

Wagner, V., Jescheniak, J. D. & Schriefers, H. On the flexibility of grammatical advance planning during sentence production: effects of cognitive load on multiple lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* **36**, 423–440 (2010).

23. 23.

Dubey, A. & Ray, S. Cortical electrocorticogram (ECoG) is a local signal. *J. Neurosci.* **39**, 4299–4311 (2019).

24. 24.

Cheung, C., Hamiton, L. S., Johnson, K. & Chang, E. F. The auditory representation of speech sounds in human motor cortex. *eLife* **5**, e12577 (2016).

25. 25.

Glanz Iljina, O. et al. Real-life speech production and perception have a shared premotor-cortical substrate. *Sci. Rep.* **8**, 8898 (2018).

26. 26.

Cisek, P. & Kalaska, J. F. Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* **33**, 269–298 (2010).

27. 27.

Ray, S. & Maunsell, J. H. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* **9**, e1000610 (2011).

28. 28.

Flinker, A., Chang, E. F., Barbaro, N. M., Berger, M. S. & Knight, R. T. Sub-centimeter language organization in the human temporal lobe. *Brain Lang.* **117**, 103–109 (2011).

29. 29.

Bouchard, K. E., Mesgarani, N., Johnson, K. & Chang, E. F. Functional organization of human sensorimotor cortex for speech articulation. *Nature* **495**, 327–332 (2013).

30. 30.

Cogan, G. B. et al. Sensory-motor transformations for speech occur bilaterally. *Nature* **507**, 94–98 (2014).

31. 31.

Kotz, S. A. et al. Lexicality drives audio-motor transformations in Broca's area. *Brain Lang.* **112**, 3–11 (2010).

32. 32.

Fadiga, L. & Craighero, L. Hand actions and speech representation in Broca's area. *Cortex* **42**, 486–490 (2006).

33. 33.

Knudsen, B., Creemers, A. & Meyer, A. S. Forgotten little words: how backchannels and particles may facilitate speech planning in conversation? *Front. Psychol.* **11**, 593671 (2020).

34. 34.

Long, M. A. et al. Functional segregation of cortical regions underlying speech timing and articulation. *Neuron* **89**, 1187–1193 (2016).

35. 35.

Tate, M. C., Herbet, G., Moritz-Gasser, S., Tate, J. E. & Duffau, H. Probabilistic map of critical functional regions of the human cerebral cortex: Broca's area revisited. *Brain* **137**, 2773–2782 (2014).

36. 36.

Long, M. A. & Fee, M. S. Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* **456**, 189–194 (2008).

37. 37.

Okobi, D. E., Jr, Banerjee, A., Matheson, A. M. M., Phelps, S. M. & Long, M. A. Motor cortical control of vocal interaction in neotropical singing mice. *Science* **363**, 983–988 (2019).

38. 38.

Tremblay, P. & Dick, A. S. Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain Lang.* **162**, 60–71 (2016).

39. 39.

Hosman, T. et al. Auditory cues reveal intended movement information in middle frontal gyrus neuronal ensemble activity of a person with tetraplegia. *Sci Rep.* **11**, 98 (2021).

40. 40.

Catani, M. et al. Short frontal lobe connections of the human brain. *Cortex* **48**, 273–291 (2012).

41. 41.

Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).

42. 42.

Mathis, A. et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).

43. 43.

Deger, K. & Ziegler, W. Speech motor programming in apraxia of speech. *J. Phon.* **30**, 321–335 (2002).

44. 44.

Jackson, E. S. et al. A fNIRS investigation of speech planning and execution in adults who stutter. *Neuroscience* **406**, 73–85 (2019).

45. 45.

Bogels, S., Casillas, M. & Levinson, S. C. Planning versus comprehension in turn-taking: fast responders show reduced anticipatory processing of the question. *Neuropsychologia* **109**, 295–310 (2018).

46. 46.

Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).

47. 47.

Fischl, B. et al. Automatically parcellating the human cerebral cortex. *Cereb. Cortex* **14**, 11–22 (2004).

48. 48.

Klein, A. & Tourville, J. 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* **6**, 171 (2012).

49. 49.

Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).

50. 50.

Avants, B. B. et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011).

51. 51.

Tyszka, J. M. & Pauli, W. M. In vivo delineation of subdivisions of the human amygdaloid complex in a high-resolution group template. *Hum. Brain Mapp.* **37**, 3979–3998 (2016).

52. 52.

Kovach, C. K. & Gander, P. E. The demodulated band transform. *J. Neurosci. Methods* **261**, 135–154 (2016).

53. 53.

Liu, Y., Coon, W. G., Pesters, A., de, B. P. & Schalk, G. The effects of spatial filtering and artifacts on electrocorticographic signals. *J. Neural Eng.* **12**, 056008 (2015).

54. 54.

Friston, K. J. et al. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* **2**, 189–210 (1995).

55. 55.

Qian, T., Wu, W., Zhou, W., Gao, S. & Hong, B. in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 2347–2350.

56. 56.

Tilsen, S. et al. Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units. *PLoS ONE* **11**, e0146813 (2016).

Acknowledgements

We thank A. Flinker, E. Jackson, J. Krivokapić, D. Schneider, N. Tritsch and members of the Long laboratory for comments on earlier versions of this manuscript; A. Ramirez-Cardenas, H. Chen, K. Ibayashi, H. Kawasaki, K. Nourski, H. Oya, A. Rhone and B. Snoad for help with data collection;

and F. Guenther and N. Majaj for helpful conversations. This research was supported by R01 DC019354 (M.A.L.), R01 DC015260 (J.D.W.G.) and Simons Collaboration on the Global Brain (M.A.L.).

Author information

Affiliations

1. NYU Neuroscience Institute and Department of Otolaryngology, New York University Langone Medical Center, New York, NY, USA

Gregg A. Castellucci & Michael A. Long

2. Center for Neural Science, New York University, New York, NY, USA

Gregg A. Castellucci & Michael A. Long

3. Department of Neurosurgery, University of Iowa, Iowa City, IA, USA

Christopher K. Kovach, Matthew A. Howard III & Jeremy D. W. Greenlee

Contributions

G.A.C. and M.A.L. conceived the study and designed the experiments; G.A.C., C.K.K., J.D.W.G. and M.A.L. conducted the research; G.A.C., C.K.K. and M.A.L. performed data analyses; G.A.C., C.K.K. and M.A.L. created the figures; G.A.C. and M.A.L. wrote the initial draft of the manuscript; G.A.C., C.K.K., M.A.H., J.D.W.G. and M.A.L. edited and reviewed the final manuscript. J.D.W.G. and M.A.L. acquired funding; J.D.W.G., M.A.H. and M.A.L. supervised the project.

Corresponding author

Correspondence to [Michael A. Long](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Gregory Cogan, Uri Hasson and Frederic Theunissen 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 Behaviour during the CI task.

a, Description of subprocesses assumed to occur during the perception, planning, and production windows of the CI task. **b**, Histograms of reaction times (RT) in early and late CI trials for all participants. **c**, Median RT values for early and late CI trials for all participants. **d**, **e**, Histograms depicting the distribution of average peak-to-trough response amplitudes for all electrodes displaying planning-related responses when aligned to CI onset in early and late trials (**d**) and different CI question types (**e**); median values for each distribution are indicated. Observed data (in black) are compared with a null distribution (in grey) consisting of randomly chosen timepoints ([Methods](#)). **f**, Schematics displaying GLM regressor structure for an early (top) and a late (bottom) variant of an example CI task question.

Extended Data Fig. 2 GLM temporal jittering analysis.

a, Full model R values for GLM fits of jittered high gamma activity from participant 436; each line represents data from an individual electrode. **b**, Example distribution of pooled D values with the fit of two Gaussians overlaid (black). The Gaussian distributions corresponding to well fit (blue) and poorly fit electrodes (red) as well as the 95th percentile of the D distribution for poorly fit electrodes (dashed line) are indicated. D values above the 95th percentile of the pooled distribution were deemed outliers (white bars) and not fitted. **c**, Table summarizing the number of electrodes rejected by the jittering analysis in each participant. **d**, Table reporting the anatomical locations of electrodes rejected by the jittering analysis and electrodes displaying significant activity in the CI task. **e**, Scatterplot depicting the proportion of rejected electrodes within a region as a function of the proportion of responsive electrodes in a region.

Extended Data Fig. 3 Analysis of neural activity in the CI task.

a, Scatterplot depicting the distribution of all simulated task-responsive electrodes from the continuum model in three-dimensional GLM weight space; cluster membership indicated by greyscale colour. **b**, **c**, Distribution of simulated electrodes from the continuum model displaying responses in one window (i.e., unmixed) of the CI task (**b**) or multiple windows (**c**); response class indicated by colour in **b** and **c** and unmixed electrodes denoted by small black points in **c**. In **b**, simulated unmixed electrodes located outside the cluster primarily containing electrodes of the same type (i.e., ‘misclustered’) are indicated with an ‘X’. **d**, **e**, Histograms depicting the distribution of the proportion of misclustered electrodes responsive during a single task window (i.e., unmixed electrodes) (**d**), and the proportion of electrodes displaying more than one significant positive weight (i.e., mixed electrodes) (**e**) across 100,000 iterations of the continuum model simulation. The median of each distribution as well as the values observed in the actual data (dashed line) are indicated. Gold arrows indicate the bin of each distribution containing the measurements corresponding to the example iteration depicted in panels p, r, and t of Fig. 1. **f**, Table reporting the number of electrodes displaying perception-related responses using either the full model or the reduced GLM lacking a planning regressor. **g**, **h**, Scatterplots depicting perception (**g**) and planning (**h**) GLM weights in the full model and reduced models lacking a planning

regressor or perception regressor, respectively. Significant positive weights are denoted with filled points and nonsignificant or significant negative weights are denoted with unfilled points; the x -coordinates of each point are randomly jittered by 25% to better visualize filled versus unfilled status. No planning electrodes displayed significant perception responses in the reduced GLM lacking a planning regressor, and no perception electrodes displayed significant planning responses in the reduced GLM lacking a perception regressor.

Extended Data Fig. 4 Additional analyses of task-related activity changes.

a, Table reporting the number of perception, planning, and production-related electrodes displaying significant positive and negative weights for each GLM regressor. **b**, Histogram depicting mean high gamma amplitude in the first 500 ms of CI questions for all unmixed perception, planning, and production electrodes. **c, d**, Canonical cortical surfaces displaying electrodes with significant positive (coloured) or negative (black) GLM weights in the perception (**c**), production (**d**), and planning (**e**) windows of the CI task across all participants. Electrode diameter is scaled to the absolute magnitude of the GLM weight, and electrodes not displaying a significant weight for a given regressor are indicated with small white circles.

Extended Data Fig. 5 Anatomical analysis of responses.

a, Cortical reconstructions for all participants displaying the location of all electrodes; the size of each electrode depicts the actual size of its recording area on the cortical surface. GLM classification is indicated by electrode colour. **b**, Canonical cortical surfaces showing electrode locations from all participants as standard-sized white circles. **c**, Number of electrodes sampling each area of the canonical cortical surface (1 cm diameter spatial smoothing) after pooling electrodes from all participants. **d**, Proportion of electrodes displaying significant production-related responses in the CI task (1-cm-diameter spatial smoothing). **e**, Canonical cortical surfaces displaying electrodes with significant responses related to speech perception, production, and planning in patients with tumour (top) and

patients with epilepsy (bottom) separately; electrode diameter scaled to GLM regressor weight. Electrodes not displaying a significant response for a process are depicted as small white circles.

Extended Data Fig. 6 Additional conversation-related analyses.

a, Table reporting additional turn-taking behavioural measures for each participant. **b**, Histograms of gap durations (time between experimenter turn offset and participant turn onset) during unconstrained conversation for each participant; bins are centred on 100 ms increments with a width of 100 ms. **c**, Scree plots for the PCA analysis of high gamma signals in the task (left) and conversation (right) periods of the recordings; data from each participant are represented by thin lines and the average across participants is denoted with a thick black line. The 95% confidence interval of the linear decay phase across participants ([Methods](#)) is also indicated. **d**, The observed number of electrodes whose cluster membership was not stable (i.e., switched clusters) between the task and conversation with a histogram depicting the distribution of electrode cluster switches expected by chance. **e**, The observed percentage of electrodes in perception, planning, and production clusters (in conversation-derived PC coefficient space) displaying significant perception, planning, and production responses (per the GLM), respectively, with histograms depicting the percentages expected by chance for each cluster type. **f**, Canonical cortical surfaces displaying the locations of all electrodes in perception, planning, and production clusters across participants ($n = 6$) in the task (left) and conversation (right). **g**, Table reporting summary statistics for PC activity (i.e., time-varying PC score) during unconstrained conversation for each participant.

Extended Data Fig. 7 PCA results for individual participants.

a–f, For 6 participants possessing sufficient numbers of electrodes belonging to multiple GLM classes ([Methods](#)): scatterplots depicting electrode distributions in PC coefficient space in the task and conversation periods (top row). Bar graphs depicting the PC coefficients for all electrodes in perception, planning, or production clusters from the PCA performed on task data and conversation data (bottom rows). Participant number given at top of each panel. **g, h**, For 2 participants possessing

mainly planning electrodes ([Methods](#), Extended Data Table 1): bar graphs depicting the PC coefficients for all planning-related electrodes from the PCA performed on task data and conversation data. In the bar graphs, the functional categorization of PCs is indicated by filled bars coloured either green (perception), blue (planning), or red (production). Any clusters rejected due to a high proportion (50%) of mixed electrodes are indicated with grey filled bars.

Extended Data Table 1 Participant information

Extended Data Table 2 DKT parcellation of all CI task-responsive electrodes

Supplementary information

Reporting Summary

Peer Review File

Supplementary Data 1

List of all task stimuli.

Supplementary Data 2

All electrode locations and GLM classifications.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Castellucci, G.A., Kovach, C.K., Howard, M.A. *et al.* A speech planning network for interactive language use. *Nature* **602**, 117–122 (2022).
<https://doi.org/10.1038/s41586-021-04270-z>

- Received: 29 September 2020
- Accepted: 19 November 2021
- Published: 05 January 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04270-z>

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/s41586-021-04270-z>

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

Toroidal topology of population activity in grid cells

- [Richard J. Gardner](#) [ORCID: orcid.org/0000-0002-3242-8840¹](#),
- [Erik Hermansen²](#),
- [Marius Pachitariu³](#),
- [Yoram Burak^{4,5}](#),
- [Nils A. Baas](#) [ORCID: orcid.org/0000-0003-0710-3507²](#),
- [Benjamin A. Dunn](#) [ORCID: orcid.org/0000-0002-3287-4744^{1,2}](#),
- [May-Britt Moser](#) [ORCID: orcid.org/0000-0001-7884-3049¹](#) &
- [Edvard I. Moser](#) [ORCID: orcid.org/0000-0003-0226-5566¹](#)

Nature volume 602, pages 123–128 (2022)

- 19k Accesses
- 252 Altmetric
- [Metrics details](#)

Subjects

- [Network models](#)
- [Neural circuits](#)

Abstract

The medial entorhinal cortex is part of a neural system for mapping the position of an individual within a physical environment¹. Grid cells, a key component of this system, fire in a characteristic hexagonal pattern of locations², and are organized in modules³ that collectively form a population code for the animal’s allocentric position¹. The invariance of the correlation structure of this population code across environments^{4,5} and behavioural states^{6,7}, independent of specific sensory inputs, has pointed to

intrinsic, recurrently connected continuous attractor networks (CANs) as a possible substrate of the grid pattern^{1,8,9,10,11}. However, whether grid cell networks show continuous attractor dynamics, and how they interface with inputs from the environment, has remained unclear owing to the small samples of cells obtained so far. Here, using simultaneous recordings from many hundreds of grid cells and subsequent topological data analysis, we show that the joint activity of grid cells from an individual module resides on a toroidal manifold, as expected in a two-dimensional CAN. Positions on the torus correspond to positions of the moving animal in the environment. Individual cells are preferentially active at singular positions on the torus. Their positions are maintained between environments and from wakefulness to sleep, as predicted by CAN models for grid cells but not by alternative feedforward models¹². This demonstration of network dynamics on a toroidal manifold provides a population-level visualization of CAN dynamics in grid cells.

[Download PDF](#)

Main

The idea of a CAN has become one of the most influential concepts in theoretical systems neuroscience^{13,14,15}. A CAN is a network in which recurrent synaptic connectivity constrains the joint activity of cells to a continuous low-dimensional repertoire of possible coactivation patterns in the presence of a wide range of external inputs. Few systems are more suitable for analysis of CAN dynamics than the spatial mapping circuits of the rodent brain, owing to the continuous, low-dimensional nature of space, and the availability and interpretability of data from these circuits^{1,2,3,4,5,6}. In medial entorhinal cortex (MEC) and surrounding areas, head direction cells¹⁶ encode orientation whereas grid cells² encode position. CAN models conceptualize the neural representations of these variables as spanning periodic one- or two-dimensional (1D or 2D) continua on a ring^{17,18,19} or a torus^{1,8,9,10,11}, respectively. In this scheme, activity within the neural network stabilizes as a localized bump when cells are ordered according to their preferred firing directions or locations in physical space. The activity bump may be smoothly translated along the network continuum by speed and direction inputs, or by external sensory cues.

In agreement with CAN models^{1,8,9,10,11}, head direction cells^{16,20,21} and modules of grid cells^{4,5,6,7} maintain fixed correlation structures. In head direction cells, cell samples of a few dozen have been sufficient to demonstrate that the network activity traverses a ring^{22,23,24}, but for grid cells, the number of possible locations in the two-dimensional state space has been too large for the topology of the manifold to be uncovered. Here we take advantage of recently developed high-site-count Neuropixels silicon probes^{25,26} to determine in many hundreds of simultaneously recorded grid

cells whether, as predicted by two-dimensional CAN models^{8,9,10,11}, the population activity in an individual grid-cell module resides on a toroidal manifold, independently of behavioural tasks and states and decoupled from the position of the animal in physical space. We focused on individual modules because (i) these are the unit networks of CAN models^{1,8,9,10}; and (ii) topological analysis of multi-module representations would require even larger numbers of cells²⁷.

Visualization of toroidal manifold

We recorded extracellular spikes of a total of 7,671 single units in layers II and III of the MEC–parasubiculum region in freely moving rats with unilateral or bilateral implants (total of 4 recordings, in 2 rats with bilateral single-shank probes and 1 rat with a unilateral 4-shank probe; from 546 to 2,571 cells per recording; Extended Data Fig. 1). During recordings, the rats were engaged in foraging behaviour in a square open-field (OF) enclosure or on an elevated track, or they slept in a small resting box. Using a clustering-based approach, we identified six grid modules across all rats (4 recording sessions, from 140 to 544 grid cells per session; 7.8% to 25.6% of total number of cells; Extended Data Fig. 2a–d, g, h). Each grid module cluster contained a mixture of nondirectional ('pure') grid cells and conjunctive grid \times direction cells²⁸, from 66 to 189 grid cells per module (total pure and conjunctive grid cells; Extended Data Fig. 2g). We initially limited our analyses to the subset of pure grid cells because (i) the expected toroidal topology might be distorted by additional directional modulation; and (ii) detection of topology in conjunctive cells may require a larger number of cells than recorded here²⁷.

To visually inspect the structure of the population activity of grid cells for signatures of toroidal topology, we constructed a three dimensional (3D) embedding of the n -dimensional population activity of a module of $n = 149$ pure grid cells (Fig. 1a). For this, we applied a two-stage dimensionality reduction procedure on the matrix of firing rates. First, to improve robustness to noise, we conducted a principal component analysis (PCA). We retained the first six principal components, which explained a particularly large fraction of the variance for all grid modules in the OF condition (with a similar tendency seen during sleep; Extended Data Fig. 4a). Next, we applied uniform manifold approximation and projection (UMAP) to reduce the six principal components into a 3D visualization. This visualization revealed a torus-like structure (Fig. 1b, Supplementary Video 1). Movement of the rat in the OF was accompanied by similarly continuous movement of the population activity across the toroidal manifold (Fig. 1b). When the activity of individual cells was plotted with reference to the 3D population representation, spikes for each cell were localized within a single patch of the population state space (Fig. 1c). The offsets between the firing locations of

individual cells in the arena corresponded with the relative firing locations of the cells in the toroidal state space.

Fig. 1: Signatures of toroidal structure in the activity of a module of grid cells.

 figure 1



a, Firing rates of 149 grid cells co-recorded from the same module and shown, in order of spatial information content, as a function of rat position in OF arena (rates colour-coded, max 0.2–35.0 Hz; rat ‘R’ day 1, module 2; Extended Data Fig. 2b). **b**, Nonlinear dimensionality reduction reveals torus-like structure in the population activity of a single grid module (same 149 cells; 3 different views of same point cloud). Each dot represents the population state at one time point (dots coloured by first principal component). Bold line shows a 5-s trajectory, demonstrating smooth movement over the toroidal manifold. Right, corresponding trajectory in OF. **c**, Toroidal positions of spikes from three grid cells from the module in **a**. Each panel shows the same 3D point cloud of population states as in **b**, with black dots indicating when the cell fired. Insets show: left: the cell’s 2D firing locations in OF (black dots on grey trajectory); middle: colour-coded firing rate map in OF (range 0 to max); right: colour-coded autocorrelogram of the rate map (range –1 to +1). Maximum rate and grid score (GS) are indicated. **d**, Same as in **c** (same cells) but with the rat running on an elevated, wheel-shaped track (‘wagon-wheel track’; WW). Note preserved toroidal field locations. **e, f**, Barcodes indicate toroidal topology of grid-cell

population activity. Results of persistent cohomology analyses (30 longest bars in the first three dimensions: H^0 , H^1 and H^2) are shown for three grid modules from one rat (R1–R3 day 1, $n = 93$, 149 and 145 cells, respectively), in OF (e) and WW (f). Grey shading indicates longest lifetimes among 1,000 iterations in shuffled data (aligned to lower values of original bars). Arrows show four most prominent bars across all dimensions (all longer than in shuffled data). One prominent bar in dimension 0, two in dimension 1 and one in dimension 2 indicates cohomology equal to that of a torus.

[Source data](#)

Quantification of toroidal topology

Although the UMAP projection allowed a toroidal point cloud to be visualized, the method does not lend itself to straightforward quantification of the topology of the state space or comparison of representations across experiments. We therefore turned to the framework of persistent cohomology, a toolset from topological data analysis in which the structure of neural data can be classified by identifying holes of varying dimensionality in topological spaces assigned to point clouds of the cells' firing rates^{22,23}. In applying this toolset, we replace each point of the point cloud by a ball of common radius. The union of balls results in a topological space in which the number of holes of different dimensions can be counted. By increasing this radius from zero until all the balls intersect, we observe the lifetime of each hole—the range of radii from when the hole first appears until it disappears (see Extended Data Fig. 3C). The lifetimes of the holes are depicted as bars and the totality of bars referred to as the barcode. For a torus, the barcode must display four bars of substantial length: a 0D hole (a single component connecting all points); two 1D holes (describing circular features); and a 2D hole (a cavity; Extended Data Fig. 3B).

Persistent cohomology analyses allowed us to classify the shape of the six-dimensional representation that serves as an intermediate step in UMAP (Extended Data Fig. 3A). We constructed barcodes for each of the six individual modules of grid cells recorded in the OF arena (three modules from rat 'R', 2 from rat 'Q' and 1 from rat 'S', henceforth named R1, R2, R3, Q1, Q2 and S1). The barcodes showed clear indications of toroidal characteristics. For all six modules, we detected four long-lived bars representing a single 0D hole, two 1D holes and a 2D hole. Their lifetimes were significantly longer than the lifetime of any bar obtained in 1,000 shuffles of the data in which spike times were randomly rotated (Fig. 1e,f, Extended Data Fig. 6Aa; $P < 0.001$). The findings suggest that network dynamics during OF foraging resides on a low-dimensional manifold with the same barcode as a torus. We noted the appearance of additional short bars in the barcodes for all modules, but these are expected for toroidal point clouds²⁷, as we confirmed with simulated data from several CAN

models^{10,11} and point clouds sampled from idealized tori, which in each case exhibited similar features (see Extended Data Fig. 7).

Tori persist despite grid distortions

The appearance of a torus in the point cloud, and the mapping of the activity of individual grid cells onto the torus (Fig. 1c), are consistent with a relationship between position in 2D physical space and position in the dimensionality-reduced neural state space. However, in many environments, this relationship may not be isometric, as the grid pattern is distorted by geometrical features of the environment, such as walls and corners^{29,30,31} or discrete landmarks and reward locations^{32,33}. We thus asked whether such geometric features could similarly distort the toroidal organization of network activity in the point cloud. We tested rats on an elevated running track shaped like a wagon wheel with four radial spokes ('wagon-wheel track' (WW); Fig. 1d,f). Spatial autocorrelation analyses confirmed that the strict periodicity of the grid pattern was compromised in this task (Extended Data Fig. 2e,f). Despite these distortions of the grid pattern in individual cells, toroidal tuning was maintained in the transformed population activity (Fig. 1d). The persistent cohomology analysis continued to identify one 0D hole, two 1D holes and one 2D hole with lifetimes that substantially exceeded those of shuffled data (Fig. 1f, Extended Data Fig. 6Ab). We also determined how the neural population activity mapped onto the torus by calculating angular coordinates from each of the two 1D holes identified by the barcode ('cohomological decoding'; Extended Data Fig. 5). The two angular coordinates defined directions intersecting at 60°, identifiable as a twisted torus (Fig. 2a). Consistent with CAN models, the vast majority of grid cells were tuned to a single location on the torus in each module and across environments, independent of geometry and local landmarks (Fig. 2b, Extended Data Fig. 4f, [Supplementary Information](#)).

Fig. 2: Cohomological decoding of position on an inferred state space torus.

 figure 2

a, b, Individual grid cells have distinct firing fields on the inferred torus (Extended Data Fig. 5). Toroidal coordinates for population activity vectors were decoded from

the two significant 1D holes (red circles in **a**) in the barcodes in Fig 1e,f. **a**, Left, 3D embedding of the toroidal state space displaying colour-coded mean firing rate of one grid cell as a function of toroidal position. Right, a 2D torus may be formed by gluing opposite sides of a rhombus. **b**, Representative grid cells from module R2 day 1 showing tuning to toroidal coordinates (all R2 cells; Supplementary Fig. 1). Each row of four plots corresponds to one cell. Left to right, colour-coded maps of cells' firing rates across the environment (OF or WW) and on the inferred torus (toroidal OF, toroidal WW, aligned to common axes). **c**, **d**, Toroidal information content (**c**) and explained deviance (**d**) for toroidal position (T) versus spatial position (S) in OF (top) and WW (bottom). Explained deviance is an R^2 -statistic (range 0–1) expressing goodness-of-fit of GLM models for S or T. Left, scatterplots with dots showing individual cells; colour indicates module (inset). Right, mean \pm s.e.m. for each module. $n = 93$ (R1), 149 (R2), 145 (R3), 94 (Q1), 65 (Q2) and 73 (S1) cells. **e**, **f**, Distances between toroidal firing field locations. **e**, Field locations of all R2 cells in OF and WW. Lines connect fields of the same cell. Toroidal OF and WW axes were aligned either separately ('separate') or commonly to OF ('common'). **f**, Left, cumulative frequency distribution of field distances (all R2 cells; green curve, separate alignment; grey lines, common alignment (to either OF or WW); black curve, shuffled data, $n = 1,000$ shuffles). Right, mean distance between field centres (\pm s.e.m.) for all modules. n cells as in **c**, **d**. **g**, Same as **f**, but showing Pearson correlations between pairs of toroidal rate maps.

Source data

To test how faithfully location in the environment is mapped onto the toroidal representation, we next asked whether grid-cell activity is predicted better by the cells' tuning to the inferred torus than by their tuning to physical space. For five out of six grid modules in OF and four out of six in WW, the information content conveyed about position, in bits per spike, was higher for position on the torus than for position in physical space (Fig. 2c; R2, R3, Q1, Q2: all $P < 0.001$, $W > 1,932$ in OF and WW; R1: $P < 0.001$, $W = 4,010$ in OF, $P = 0.586$, $W = 2,129$ in WW; S1: $P = 1.000$ in OF and WW, $W = 620$ in OF, $W = 129$ in WW; Wilcoxon signed-rank test). We verified this difference by comparing the cross-validated prediction of two Poisson generalized linear model (GLM)-based encoding models of each cell's activity that included toroidal position (decoded as above) and 2D spatial position. For both environments (OF and WW), the toroidal covariate was closer to a perfectly fitted model of the data than was the physical position covariate in five out of six grid-cell modules (Fig. 2d; R1, R2, R3, Q1, Q2: $P < 0.001$, $W > 2,045$ in OF and WW; S1: $P < 0.001$, $W = 1,941$ in OF, $P = 1.000$, $W = 727$ in WW; Wilcoxon signed-rank test). Together, these differences point to toroidal structure as the primary feature of the population activity of grid cells, superior to that of the 2D coordinates of the animal's current position in the physical environment.

If grid cells operate on a toroidal manifold determined by intrinsic network features, this manifold may be expressed universally across environments, independently of sensory inputs. We tested this proposition by assessing, on the inferred tori, whether the locations of firing fields of different grid cells were maintained between OF and WW (Fig. 2b, [Supplementary Information](#)). To compare the toroidal parametrizations, we aligned the axes of the toroidal coordinates (Extended Data Fig. 5b). First, we compared, for each cell, the distance between the centres of mass of the toroidal rate maps in OF and WW (Fig. 2e,f, Extended Data Fig. 6Ba). This distance was substantially shorter (mean \pm s.e.m. of mean distances for all modules: 31.5 ± 6.3 degrees) than that of control data in which the order of the rate maps in one environment was shuffled (135.8 ± 1.7 degrees; maximum possible distance $\sqrt{2} \cdot 180 \approx 254.6$ degrees; data versus shuffled: $P < 0.001$ in all modules). Second, we calculated the pairwise Pearson correlations of binned toroidal rate maps across the two environments (Fig. 2g, Extended Data Fig. 6Ba). Consistent with the centre-of-mass comparison, the correlations between OF and WW were higher in observed data (mean \pm s.e.m. of mean r values for all modules: 0.79 ± 0.07) than in shuffled data ($r = 0.01 \pm 0.01$; $P < 0.001$ for all modules). Very similar results were obtained when applying the toroidal parametrization from the same environment (either OF or WW) to activity from both environments (Fig. 2f,g, 16.0 ± 3.4 degrees; $r = 0.95 \pm 0.02$; $P < 0.001$ for all modules and both mappings). Together, these findings suggest that physical space is mapped onto the same internal low-dimensional manifold irrespective of the specific environment.

Toroidal topology persists during sleep

If population activity is mapped onto the same toroidal manifold independently of sensory inputs, the toroidal topology should also be maintained during sleep. To test this idea, the rats rested in a high-walled, opaque box placed in the centre of the OF or WW track. Periods of rapid-eye-movement (REM) sleep and slow-wave sleep (SWS) were identified on the basis of the low-frequency rhythmic content of the aggregated multi-unit activity in combination with prolonged behavioural immobility (Extended Data Fig. 9).

Persistent cohomology analysis of the sleep population activity suggested toroidal topology in five of the six grid modules during REM and four out of six modules during SWS (modules R2, R3, Q1 and Q2 for both sleep stages and module R1 only in REM; Fig. 3a, Extended Data Fig. 6Ac,d). In the remaining module (S1), there were no long-lived bars in dimensions 1 or 2 (Extended Data Fig. 6Ac,d), indicating an absence of toroidal structure during sleep, perhaps because of an insufficient number of cells in this module (72 cells; Extended Data Fig. 4e). The barcode results were supported by the toroidal mapping, which revealed sharply tuned firing fields on the REM and SWS tori ($99.3 \pm 1.6\%$ and $99.1 \pm 1.8\%$, respectively, of the grid cells in

each module had higher information content than shuffled data, and in $95.3 \pm 7.2\%$ and $98.6 \pm 2.4\%$ of cells the toroidal tuning explained the activity better than a null model that assumes a constant firing rate; Fig. 3b, Extended Data Figs. 6C, 10, [Supplementary Information](#)). In addition, the spatial arrangements of toroidal firing locations of different cells were maintained between wake, REM and SWS states (Fig. 3c, Extended Data Fig. 6Bb, c). For between-condition pairs of rate maps, the mean distance (\pm s.e.m.) between the peak firing locations (OF versus REM 31.5 ± 15.4 degrees, OF versus SWS 29.8 ± 14.3 degrees) was well below the distribution of shuffled distances (Fig. 3d, Extended Data Fig. 6Bb, c; 135.8 ± 2.3 degrees in both REM and SWS, $P < 0.001$ for all 5 and 4 modules, respectively). Similarly, the mean correlations of pairs of toroidal rate maps (REM versus OF $r = 0.80 \pm 0.15$, SWS versus OF $r = 0.83 \pm 0.12$) were substantially larger than in shuffled versions of the data (Fig. 3e, Extended Data Fig. 6Bb, c; $r = 0.01 \pm 0.01$ in both REM and SWS, $P < 0.001$ for all 5 and 4 modules, respectively). Thus, the toroidal structure is maintained in both sleep conditions, despite the lack of external spatial inputs.

Fig. 3: Preservation of toroidal structure during sleep.

 **figure 3**

a, Barcodes indicating toroidal topology for grid-cell module R2 day 2 ($n = 152$ cells) during REM sleep and SWS (as in Fig. 1e, f). **b**, Toroidal rate maps showing preserved toroidal tuning for individual cells across environments and brain states (as in Fig. 2b; all cells shown in Extended Data Fig. 10). From left: rate map for OF in physical coordinates; and rate maps for OF, REM sleep and SWS in toroidal

coordinates. **c**, Distribution of toroidal field centres (as in Fig. 2e) in OF and sleep (n as in a). **d, e**, Left, cumulative distributions of distances between toroidal field centres (**d**) and Pearson correlation r values (**e**) of rate maps for all R2 grid cells, as in Fig. 2f,g, but comparing OF with REM or SWS. Right, mean value \pm s.e.m. for all modules. $n = 111$ (R1), 152 (R2), 165 (R3), 94 (Q1), 65 (Q2) and 72 (S1) cells. $n = 1,000$ shuffles.

[Source data](#)

Classes of grid cells

We next investigated why toroidal structure was not visible during REM in module S1 and during SWS in modules R1 and S1 (Fig. 4a, Extended Data Fig. 6Ad). Previous studies of medial entorhinal spiking activity have described cell populations with distinct burst-firing and theta-modulation characteristics^{34,35,36}; therefore, we asked whether a lack of toroidal structure was due to heterogeneity in the composition of the module. We quantified each cell's temporal modulation characteristics using the spike train temporal autocorrelogram from the OF session, and by applying clustering to the matrix of autocorrelograms we obtained three cell classes (Fig. 4b). Each class was distributed across multiple modules (Fig. 4d). Within each module, cells from the three classes showed overlapping grid spacing and orientation properties (Extended Data Fig. 8a). We named the classes 'bursty' (B), 'non-bursty' (N) and 'theta-modulated' (T), following the most prominent autocorrelogram feature of each class (Fig. 4e). We also examined the spike waveforms of the cells, and found that each class showed a characteristic spike width (Fig. 4f,g), suggesting that they differ in morphology or biophysical properties.

Fig. 4: Differential toroidal tuning of grid-cell subpopulations.

 [figure 4](#)

a, Barcode of all pure R1 grid cells (day 2, $n = 111$ cells) does not indicate toroidal structure during SWS. **b**, Matrix of cosine distances between pairs of spike-train autocorrelograms of grid cells in module R1. Rows and columns show 189 grid cells (pure and conjunctive) sorted by cluster identity. Three clusters were identified,

appearing as dark (that is, similar) squares along the matrix diagonal. On the basis of temporal firing patterns (**e**), they were named ‘bursty’ (B), ‘theta-modulated’ (T) and ‘non-bursty’ (N). **c**, Barcode of the ‘bursty’ class of R1 ($n = 69$ cells) indicates toroidal structure. Symbols as in **a**. Arrows point to the four most persistent features. **d**, Fractions of grid cells in each class, shown for each grid module. Left, pure grid cells only, right, conjunctive grid \times head-direction cells only. For n see Extended Data Fig. [2g](#). **e**, Average temporal autocorrelogram for cells in each class. Shaded area shows mean \pm s.e.m. (bursty $n = 523$, theta-modulated $n = 229$, non-bursty $n = 95$ cells). For each class, note short-latency peak (burst-firing) and long-latency peak (theta-modulation). **f**, Average spike template waveforms of cells from each class (n as in **e**). Shaded area indicates mean \pm s.e.m. **g**, Cell classes have different burst-firing characteristics, as expressed by latency of first autocorrelogram peak (x axis) and peak-to-peak spike width (y axis). Cells (dots) are colour-coded by class (n as in **e**) or by identity (pure or conjunctive, $n = 659$ or 188 cells, respectively). **h**, Example cells from each class (one row of plots per cell). Plots from left to right: OF firing rate map; head-direction (HD) tuning curve (black) compared to occupancy of head directions (light grey); temporal autocorrelogram; toroidal firing rate maps for OF, REM and SWS.

[Source data](#)

The firing rates of the cells during SWS exhibited marked correlation structure within—but not between—classes (Extended Data Fig. [8b](#)). Even though our classification strategy was not influenced by the cells’ directional tuning, class T contained 80% of all conjunctive grid cells and only 11% of all pure grid cells, supporting the idea that conjunctive grid cells are a distinct population. Accordingly, in modules R1 and S1, which contained the largest numbers of T cells, pairwise correlations of T cells’ spike trains were more strongly related to head-direction tuning than to toroidal tuning (Extended Data Fig. [8c](#)). When we subdivided module R1 into the three classes (Fig. [4b](#)), we found that during SWS toroidal topology was detectable only in B cells (Fig. [4c](#)). By decoding toroidal position from B cells, we were able to recover the selectivity of each cell with respect to toroidal position in module R1 (Fig. [4h](#)). The toroidal tuning locations were preserved between OF and SWS in each cell class in R1 (Extended Data Fig. [8d](#), B: distance of 26.4 ± 6.1 degrees and correlation of $r = 0.85 \pm 0.02$, T: 43.6 ± 3.9 degrees and $r = 0.74 \pm 0.02$, N: 29.9 ± 3.5 degrees and $r = 0.80 \pm 0.02$; mean values of shuffled versions of each class were between 135.4 ± 5.2 and 136.4 ± 6.2 degrees, and between $r = 0.00 \pm 0.07$ and $r = 0.02 \pm 0.03$; comparison between observed and shuffled $P < 0.001$ for all 3 classes and both measures). However, in R1 as well as all other modules, toroidal spatial information and explained deviance were highest for B cells and lower for N and T cells in OF, REM and SWS (Extended Data Fig. [8e](#)) (information content: $P < 10^{-56}$, $H > 255$; Kruskal–Wallis test; $P < 10^{-9}$, $Z > 6.4$; Dunn test with Bonferroni correction; explained

deviance: $P < 10^{-20}$, $H > 96$; Kruskal–Wallis test; $P < 10^{-12}$, $Z > 7.4$; Dunn test with Bonferroni correction, for OF, REM and SWS). Collectively, these results show that the B cell population (containing the majority of our grid cells) represents the torus most robustly across behavioural conditions. The weaker toroidal representation in T cells may partly be an effect of the higher dimensionality of the code carried by conjunctive grid \times direction cells. Indeed, running cohomology analysis on T cells from modules S1 and R1 (which contained the most T cells) revealed a circular feature that corresponded to the animal’s head direction (Extended Data Fig. 8f,g).

Discussion

Our findings, from many hundreds of simultaneously recorded grid cells, show that population activity in grid cells invariably spans a manifold with toroidal topology, with movement on the torus matching the animal’s trajectory in the environment. The toroidal representation was most stably encoded by the bursting subclass of grid cells. Toroidal topology was not simply inherited from the encoded variable, as 2D space is not characterized by toroidal topology, as opposed to pitch and azimuth of head orientation, which in bats together span a torus and thus naturally map onto a toroidal neural code³⁷. Using cohomological decoding, we were able to demonstrate, in each environment and in both sleep and awake states, that the toroidal coordinates of individual grid cells in individual grid modules were maintained, independently of external sensory inputs or environment-induced deformations of hexagonal symmetry in the rate maps^{29,30,31,32,33}. The uniform and consistent toroidal structure of the manifold suggests that distortions in grid patterns occur in the mapping between physical space and the toroidal grid code rather than in the grid code itself.

The invariance of the toroidal manifold across environments and brain states is informative about the mechanisms that underlie grid-cell activity. Although toroidal topology can be generated by both CAN^{1,8,9,10} and feedforward¹² mechanisms, the persistence of an invariant toroidal manifold under conditions that give rise to changes in the correlation structure of place-cell activity in the hippocampus^{6,7} is predicted only by CAN models. While the findings do not exclude co-existing feedforward mechanisms^{12,38}, they point to intrinsic network connectivity as the mechanism that underlies the rigid toroidal dynamics of the grid-cell system. What kind of network architecture keeps the activity on a toroidal manifold—whether it is geometrically organized^{1,8,9,10} or acquired from random connectivity by synaptic weight adjustments through learning^{39,40,41}—remains to be determined, as does the mode of connectivity with other CANs in the entorhinal–hippocampal system^{22,23}.

Methods

Rats

The data were collected from three experimentally naive male Long Evans rats (Rats Q, R and S, 300–500 g at time of implantation). The rats were group-housed with three to eight of their male littermates before surgery and were singly housed in large Plexiglas cages (45 × 44 × 30 cm) thereafter. They were kept on a 12-h light–12-h dark schedule, with strict control of humidity and temperature. All procedures were performed in accordance with the Norwegian Animal Welfare Act and the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes. Protocols were approved by the Norwegian Food Safety Authority (FOTS ID 18011 and 18013).

Electrode implantation and surgery

The rats were implanted with Neuropixels silicon probes^{25,26} targeting the MEC–parasubiculum (PaS) region. Two rats were implanted bilaterally with prototype Neuropixels ‘phase 3A’ single-shank probes and with one probe targeting MEC–PaS in each hemisphere; the third rat was implanted with a prototype Neuropixels 2.0 multi-shank probe in the left hemisphere. Probes were inserted at an angle of 25° from posterior to anterior in the sagittal plane. Implantation coordinates were AP 0.05–0.3 mm anterior to the sinus and 4.2–4.7 mm lateral to the midline. The probes were inserted to a depth of 4,200–6,000 µm. The implant was secured with dental cement. The detailed implantation procedure has been described elsewhere^{6,26}. After surgery, the rats were left to recover for approximately 3 h before beginning recording. Postoperative analgesia (meloxicam and buprenorphine) was administered during the surgical recovery period.

Recording procedures

The details of the Neuropixels hardware system and the procedures for freely moving recordings have been described previously. In brief, electrophysiological signals were amplified with a gain of 500 (for phase 3A probes) or 80 (for 2.0 probes), low-pass-filtered at 300 Hz (phase 3A) or 0.5 Hz (2.0), high-pass-filtered at 10 kHz, and then digitized at 30 kHz (all steps performed by the probe’s on-board circuitry). The digitized signals were multiplexed by an implant-mounted ‘headstage’ circuit board and were transmitted along a lightweight 5-m tether cable, made using either micro-coaxial (phase 3A) or twisted pair (2.0) wiring.

Three-dimensional motion capture (OptiTrack Flex 13 cameras and Motive recording software) was used to track the rat’s head position and orientation, by attaching a set of five retroreflective markers to implant during recordings. The 3D marker positions were projected onto the horizontal plane to yield the rat’s 2D position and head

direction. An Arduino microcontroller was used to generate digital pulses, which were sent to the Neuropixels acquisition system (via direct TTL input) and the OptiTrack system (via infra-red LEDs), to permit precise temporal alignment of the recorded data streams.

Behavioural procedures

Data were obtained from four recording sessions performed within the first 72 h after recovery from surgery. The recordings were performed while the rats engaged in three behavioural paradigms, each in a different arena within the same room. Abundant distal visual and sonic cues were available to the rat. On each day of recording, the rat remained continuously connected to the recording apparatus across the various behavioural sessions that were performed. Occasionally it was necessary to remove twists that had accumulated in the Neuropixels tether cable. In such cases, the ongoing behavioural task was paused while the experimenter gently turned the rat to remove the twists. During pre-surgical training, the rats were food-restricted, maintaining their weight at a minimum of 90% of their free-feeding body weight. Food was generally removed 12–18 h before each training session. Food restriction was not used at the time of recording.

Open-field foraging task

Rats foraged for randomly scattered food crumbs (corn puffs) in a square open-field (OF) arena of size 1.5×1.5 m, with black flooring and enclosed by walls of height 50 cm. A large white cue card was affixed to one of the arena walls (height same as the wall; width 41 cm; horizontal placement at the middle of the wall). At the time of the surgery, each rat was highly familiar with the environment and task (10–20 training sessions lasting at least 20 min each).

Wagon-wheel track foraging task

The wagon-wheel (WW) track task was designed to function as a 1D version of the 2D OF foraging task. The track's geometry comprised an elevated circular track with two perpendicular cross-linking arms spanning the circle's diameter. The track was 10 cm wide and was bounded on both sides by a 1-cm-high lip. Each section of the track was fitted with a reward point, placed halfway between the two nearest junctions, in the centre of the section. Each reward point consisted of an elevated well that could be remotely filled with chocolate milk via attached tubing. To encourage foraging behaviour, a pseudorandom subset of the wells (between one and four of the eight wells) was filled at a given time, and the rat was allowed to explore the full maze freely and continuously. Wells were refilled as necessary when the rat consumed rewards. Each rat was trained to high performance on the foraging task before the

surgery (collecting at least 30 rewards within a 30-minute session). Training to this level of performance took 5–10 half-hour sessions.

Natural sleep

For sleep sessions, the rat was placed in a black acrylic ‘sleep box’ with a 40 × 40-cm square base and 80-cm-high walls. The black coating of walls was transparent to infrared, which allowed the 3D motion capture to track the rat through the walls. The bottom of the sleep box was lined with towels, and the rat had free access to water. During recording sessions in the sleep box, the main room light was switched on and pink noise was played through the computer speakers to attenuate disturbing background sounds. Sleep sessions typically lasted 2–3 h, but were aborted prematurely if the rat seemed highly alert and unlikely to sleep.

Spike sorting and single-unit selection

Spike sorting was performed with KiloSort 2.5²⁶. In brief, the algorithm consists of three principal stages: (1) a raw-data alignment procedure that detects and corrects for shifts in the vertical position of the Neuropixels probe shank relative to the surrounding tissue; (2) an iterative template-matching procedure that uses low-rank, variable-amplitude waveform templates to extract and classify single-unit spikes; and (3) a curation procedure which detects appropriate template merging and splitting operations based on spike train auto- and cross-correlograms. Some customizations were made to the standard KiloSort 2.5 method to improve its performance on recordings from the MEC–PaS region, where there is a particularly high spatiotemporal overlap of spike waveforms owing to the high density of cells. Therefore, the maximum number of spikes extracted per batch in step 1 above was increased, as was the number of template-matching iterations in step 2. To improve the separation between cells with very similar-looking waveforms, the upper limit on template similarity was raised from 0.9 to 0.975 in step 2 and to 1.0 on step 3, while supervising manually all merge and split operations from step 3, using a custom-made GUI running in MATLAB. The manual supervision ensured that Kilosort 2.5 did not automatically merge pairs of units with a dip in the cross-correlogram, which in our data was often due to out-of-phase spatial tuning. The merge and split operations were repeated several times to ensure the best separation between single units.

Single units were discarded if more than 1% of their interspike interval distribution consisted of intervals less than 2 ms. In addition, units were excluded if they had a mean spike rate of less than 0.05 Hz or greater than 10 Hz (calculated across the full recording duration).

Single-unit spike waveforms

During spike sorting, Kilosort assigned each unit with a 2 ms spike waveform template on each recording channel. To calculate a representative single waveform for each unit, the peak-to-peak amplitude of the template was calculated on every channel, and the templates from the three highest-amplitude channels were averaged to generate the representative spike waveform. To calculate spike width, a unit's representative waveform was finely interpolated (from 61 to 1,000 points) using a cubic spline function. Spike width was defined as the time difference between the waveform's negative peak (to which the waveform was aligned by Kilosort), and the following positive peak.

Spatial position and direction tuning

During awake foraging sessions in the OF arena or wagon-wheel track, only time epochs in which the rat was moving at a speed above 2.5 cm s^{-1} were used for spatial or toroidal analyses. To generate 2D rate maps for the OF arena, position estimates were binned into a square grid of $3 \times 3\text{-cm}$ bins. The spike rate in each position bin was calculated as the number of spikes recorded in the bin, divided by the time the rat spent in the bin. To interpolate the values of unvisited bins, two auxiliary matrices were used, M_1 and M_2 , setting visited bins equal to the value of the original rate map in M_1 and to 1 in M_2 , and setting unvisited bins to zero in both. One iteration of the image-processing ‘closing’ operation was then performed (binary dilation followed by erosion, filling out a subset of the non-visited bins) on M_2 , using a disk-shaped structuring element, first padding the matrix border by one bin. Both matrices were then spatially smoothed with a Gaussian kernel of smoothing width 2.75 bins. Finally, the rate map was obtained by dividing M_1 by M_2 . Rate-map spatial autocorrelograms and grid scores were calculated as described previously²⁸. The selectivity of each cell’s position tuning was quantified by computing its spatial information content⁴², measured in bits per spike (see ‘Information content’).

Head-direction tuning curves were calculated by binning the head-direction estimates into 6° bins. The spike rate in each angular bin was calculated as the number of spikes recorded in the bin divided by the time that the rat spent in the bin. The resultant tuning curve was smoothed with a Gaussian kernel with $\sigma = 2$ bins, with the ends of the tuning curve wrapped together. The selectivity of head-direction tuning was quantified using the mean vector length (MVL) of the tuning curve. This was calculated according to:

$$\text{\$}\$\{\rm{rm\{M\}}\}\{\rm{rm\{V\}}\}\{\rm{rm\{L\}}\}\text{=}\frac{|\sum_{j=1}^M\{\bf{f}\}_j|}{(\sum_{j=1}^M\{\bf{f}\}_j)},\text{\$}\$$$

where vector \mathbf{f} represents the tuning curve values (firing rates), vector α represents the corresponding angles, M is the number of tuning curve values, and $|\cdot|$ represents the

absolute value of the enclosed term.

Grid module classification

A novel method was implemented to detect populations of cells corresponding to grid modules by finding clusters of cells that expressed similar spatially periodic activity in the open field (Extended Data Fig. 2). Contrary to previous clustering-based methods for grid modules³, this approach makes no assumptions about the specific geometry of the grid pattern, thus making it less susceptible to the detrimental effects of geometric distortions such as ellipticity^{3,30}.

For each MEC–PaS cell in a given recording, a coarse-resolution rate map of the OF session was constructed, using a grid of 10×10 -cm bins, with no smoothing across bins. The 2D autocorrelogram of this rate map was calculated, and the central peak was removed by excluding all bins located less than 30 cm from the autocorrelogram centre. Bins located more than 100 cm from the autocorrelogram centre were also excluded. The autocorrelograms for all cells were subsequently converted into column vectors, z -standardized, then concatenated to form a matrix with spatial bins as rows and cells as columns. The nonlinear dimensionality reduction algorithm UMAP^{43,44} was then applied to this matrix, yielding a two-dimensional point cloud in which each data point represented the autocorrelogram of one cell (Extended Data Fig. 2a–d; UMAP hyperparameters: ‘metric’=‘manhattan’, ‘n_neighbors’=5, ‘min_dist’=0.05, ‘init’=‘spectral’). In the resultant 2D point cloud, cells with small absolute differences between their autocorrelogram values were located near to one another. The point cloud was partitioned into clusters using the DBSCAN clustering algorithm (MATLAB function ‘dbscan’, minimum 30 points per cluster, eta = 0.6–1.0). In every recording, the largest cluster was mainly composed of cells that either lacked strong spatial selectivity or were spatially selective but without clear periodicity. All remaining clusters contained cells with high grid scores, and with similar grid spacing and orientation (Extended Data Fig. 2a–d); cluster membership was therefore used as the basis for grid module classification. In one recording (rat ‘R’ day 1), two clusters were identified that had similar average grid spacing and orientation (labelled as ‘R1a’ and ‘R1b’ in Extended Data Fig. 2a–d), suggesting that they represented the same grid module. R1b appeared to comprise cells with higher variability in the within-field firing rates of the spatial rate maps, accompanied by more irregularities in the autocorrelograms. These two clusters were merged together in subsequent analysis (in which the resultant cluster is called ‘R1’).

A subset of the cells that were assigned to grid module clusters by the above procedure were tuned to both location and head direction (conjunctive grid \times direction cells). These cells, which were defined as having a head-direction tuning curve with mean vector length above 0.3, were discarded from further analysis.

Classification of sleep states

SWS and REM periods were identified on the basis of a combination of behavioural and neural activity, following previously described approaches^{6,45,46}. First, sleep periods were defined as periods of sustained immobility (longer than 120 s with a locomotion speed of less than 1 cm s^{-1} and head angular speed of less than 6° s^{-1}). Qualifying periods were then subclassified into SWS and REM on the basis of the amplitude of delta- and theta-band rhythmic activity in the recorded MEC–PaS cells. Spike times for each cell were binned at a resolution of 10 ms and the resultant spike counts were binarized, such that ‘0’ indicated the absence of spikes and ‘1’ indicated one or more spikes. The binarized spike counts were then summed across all cells (Extended Data Fig. 9A). The rhythmicity of this aggregated firing rate with respect to delta (1–4 Hz) and theta (5–10 Hz) frequency bands was quantified by applying a zero-phase, fourth-order Butterworth band-pass filter, then calculating the amplitude from the absolute value of the Hilbert transform of the filtered signal, which was smoothed using a Gaussian kernel with $\sigma = 5 \text{ s}$ and then standardized (‘z-scored’). The ratio of the amplitudes of theta and delta activity was hence calculated (theta/delta ratio, ‘TDR’). Periods in which TDR remained above 5.0 for at least 20 s were classified as REM; periods in which TDR remained below 2.0 for at least 20 s were classified as SWS (Extended Data Fig. 9B).

Spectral analysis was performed on 10-ms-binned multi-unit activity using the multi-tapered Fourier transform, implemented by the Chronux toolbox (<http://chronux.org/>, function ‘mtspectrumeegc’). Non-overlapping 5-second windows were used, with a frequency bandwidth of 0.5 Hz and the maximum number of tapers.

Visualization of toroidal manifold

For each module of grid cells, spike times of co-recorded cells in the OF were binned for each cell at a resolution of 10 ms, and the binned spike counts were convolved with a Gaussian filter with $\sigma = 50 \text{ ms}$. Time bins in which the rat’s speed was below 2.5 cm s^{-1} were then discarded. To account for variability of average firing rates across cells, the smoothed firing rate of each cell was z-scored. For computational reasons, the time bins were downsampled, taking every 25th time bin (equating to 250-ms intervals between selected samples). Collectively, the downsampled firing rates of the full population of cells formed a matrix with time bins in rows and cells in columns. PCA was applied to this matrix (treating time bins as observations and cells as variables), and the first six principal components were retained (Extended Data Figs. 3Aa–c, 4a–d). UMAP^{43,47} was then run on these six principal components (with time bins as observations and principal components as variables). The hyperparameters for UMAP were: ‘n_dims’=3, ‘metric’=‘cosine’, ‘n_neighbours’=5000, ‘min_dist’=0.8 and ‘init’=‘spectral’.

For visualizing the toroidal manifold during WW, smoothed firing rates were first calculated by the same procedure described above for OF. Subsequently, to allow comparison of the toroidal manifold between OF and WW, the same PCA and UMAP transformations calculated for the OF data were re-applied to the WW data, by supplying the fitted OF UMAP transformation as the argument ‘template_file’ to the ‘run_umap’ function in the MATLAB implementation⁴⁷.

Preprocessing of population activity

Each topological analysis was based on the activity of a single module of grid cells, during a single experimental condition in one recording session. Topological analysis of multi-module and conjunctive grid \times direction cell activity was not considered as we expect such data to exhibit higher-dimensional topological structure requiring a higher number of cells²⁷. The experimental conditions were: open-field foraging (OF), wagon-wheel track foraging (WW), slow-wave sleep (SWS), and rapid eye-movement sleep (REM). Sleep epochs of the same type were collected from across the recording and concatenated for analysis purposes. Similarly, in one case (rat 'S'), two WW task sessions were concatenated to increase the sample size.

In total there were 27 combinations of module (Q1, Q2, R1, R2, R3, S1) and experimental condition (OF day 1, OF day 2, WW, REM, SWS).

Preprocessing of spike trains began by computing delta functions centred on the spike times (valued 1 at time of firing; 0 otherwise), and convolving these temporally with a Gaussian kernel with $\sigma = 50$ ms (OF, WW and REM) or 25 ms (SWS). Samples of the smoothed firing rates of all cells ('population activity vectors') were then computed at 50-ms intervals. The awake states were further refined by excluding vectors which originated from time periods when the rat's speed was below 2.5 cm s^{-1} .

Computing the persistent cohomology of a point cloud is computationally expensive and may be sensitive to outliers (for example, spurious points breaking the topology of the majority of points in the point cloud). For this reason, it is common to preprocess the data by downsampling and dimension-reducing the point cloud. The same preprocessing procedure was used for all datasets in the present study.

First, the data points were downsampled by keeping the 15,000 most active population activity vectors (as measured by the mean population firing rate). During SWS, this selection criterion had the consequence of automatically discarding population activity vectors during down-states, when neural activity is near-silent. As noise is inherently more prevalent and cosine distances less reliable in high-dimensional spaces ("the curse of dimensionality")⁴⁸, dimensionality-reduction and a normalization of distances were subsequently performed. The reduced point cloud was z -scored and projected to

its six first principal components, thus reducing noise while keeping much of the variance (see Extended Data Fig. 4a). This was supported by the lack of grid structure and the clear drop in explained deviance after six components (see Extended Data Fig. 4b,c). The explained deviance was computed by fitting a GLM model to each component individually, using the spatial coordinates as covariate, suggesting that the higher components are less spatially modulated and possibly better described by other (unknown) covariates. Consistent with this, the toroidal structure was most clearly detected in the barcodes when comparing the ratio of the lifetimes of the two most persistent H¹ bars versus the third longest-lived H¹ bar for the barcodes obtained when using different numbers of components in the analysis (see Extended Data Fig. 4d). These analyses both indicated that dimensionality reduction was required to firmly demonstrate the toroidal topology in the grid cells. The empirical findings are supported theoretically; see ‘Theoretical explanation of the six-dimensionality proposed by PCA’ in [Supplementary Methods](#).

To further simplify the low-dimensional point cloud, a different downsampling technique was introduced, based on a point-cloud density strategy motivated by a topological denoising technique introduced previously⁴⁹ and a fuzzy topological representation used in UMAP^{43,50}. Parts of the open-source implementation of the latter were copied in this computation. This approach consisted of assigning, for each point, a neighbourhood strength to its k nearest neighbours, and subsequently sampling points that represent the most tight-knit neighbourhoods of the point cloud in an iterative manner. First, we defined $\left(\{m\}_{i,i,j}\right)^{\prime} = \exp \left(-\frac{\{d\}_{i,i,j}}{\sigma_i}\right)$ where $\{d\}_{i,i,j}$ is the cosine distance between point x_i and its j th nearest neighbour and σ_i is chosen to make $\sum_{j=1}^k \{m\}_{i,i,j}^{\prime} = \log k$ using $k = 1,500$. The neighbourhood strength was then obtained by symmetrizing: $\{m\}_{i,i,j} = \{m\}_{i,i,j}^{\prime} + \{m\}_{j,i}^{\prime} - \{m\}_{i,i,j}^{\prime} \cdot \{m\}_{j,i}^{\prime}$. Finally, the point cloud was reduced to 1,200 points by iteratively drawing the i -th point as: $\max_{\tilde{I}} \sum_{j \in \tilde{I}} \{m\}_{j,i}^{\prime}$ where \tilde{I} denotes the indices of the points not already sampled. In other words, for each iteration, the sampled point is the one with the strongest average membership of the neighbourhoods of the remaining points.

To compute the persistent cohomology of the downsampled point cloud, the neighbourhood strengths were first computed for the reduced point cloud (using $k = 800$) and its negative logarithm was taken, obtaining a distance matrix. This matrix was then given as input to the Ripser implementation^{51,52} of persistent cohomology, returning a barcode. In short, the barcode gave an estimate of the topology of the fuzzy topological representation of the six principal components of the grid-cell population activity. Thus, in essence, the first step of UMAP was applied before describing the

resulting representation with persistent cohomology, instead of using it to project each point of the point cloud to a representation of user-specified dimensionality for visualization (Extended Data Fig. 3Ad,e). This gives a more direct and stable quantification of the global data structure, without having to choose an initialization⁵³ or optimize a lower-dimensional representation.

Persistent cohomology

Persistent cohomology, a tool in topological data analysis, was used to characterize the manifold assumed to underlie the data. This has clear ties with persistent homology and the main result (the barcode) is identical, thus the two terms are often used interchangeably. Persistent cohomology was chosen because the computation is (to our knowledge) faster and is required to obtain cocycle representatives, which are necessary to perform decoding (see ‘Cohomological decoding’). Persistent (co-)homology has previously been successful in analysing neural data, describing the ring topology of head direction cell activity^{22,23,24}, the spherical representation of population activity in primary visual cortex⁵⁴, and the activity of place cells^{55,56,57,58}.

The general outline of the algorithm is as follows. Each point in the cloud is replaced by a ball of infinitesimal radius, and the balls are gradually expanded in unison. Taking the union of balls at a given radius results in a space with holes of different dimensions. The range of radii for which each hole is detected is tracked; this is referred to as the ‘lifetime’ of the hole and is represented by the length of a bar. The totality of bars is referred to as the barcode.

The software package Ripser^{51,52} was used for all computations of persistent cohomology. Ripser computes the persistent cohomology of ‘Vietoris-Rips complexes’ (which approximate the union of balls for different radii), constructed based on the input distance matrix and a choice of coefficients (in our case, \mathbb{Z}_{47} -coefficients), and outputs the barcode and cocycle representatives for all bars. The prime 47 was chosen as homology and cohomology coincide in this case and as it is unlikely that this divides the torsion subgroup of the homology of the space. Torsion may indicate, for example, orientability of a manifold and in choosing 47 as our prime, we disregard all but 47-torsion. Testing with other primes (for example, 43) gave similar results (data not shown) and the Betti numbers stayed the same regardless of choice of prime.

To verify that the lifetimes of prominent bars in the barcodes were beyond chance, shuffled distributions were generated for the persistence lifetimes in each dimension. In each shuffling, the spike train of each cell was shifted independently in time by rolling the firing rate arrays a random length between 0 and the length of the session. The same preprocessing and persistence analysis were then performed on the shifted spike trains as for the unshuffled data. This was performed 1,000 times, and each time

a barcode was obtained. The barcodes were concatenated for all shuffles and the maximum lifetime was found for each dimension. This lifetime served as a significance criterion for the bar lifetimes. It is noted, however, that this is a heuristic and that statistics of barcodes are still not well established.

Cohomological decoding

As there are other spaces with similar barcodes as for a torus, the results identified by the barcode were further investigated, using the ‘cohomological decoding’ procedure introduced previously⁵⁹ to calculate a toroidal parametrization of the point clouds of population activity. This assigns to each point corresponding positions on each of the two circular features identified by the 1D bars with the longest lifetime, resulting in coordinates that further characterize the underlying shape of the data.

Cohomological decoding is motivated by the observation that the 1D cohomology (with integer coefficients) of a topological space X is equivalent to the set of homotopy-equivalent classes of continuous maps from X to the circle (S^1)⁶⁰; that is:

$$\text{H}^1(X; \mathbb{Z}) \cong [X, S^1].$$

This subsequently means that for each 1D bar existing at a given radius, there exists a corresponding continuous map from the Vietoris-Rips complex of that radius to the circle. Thus, we may first use persistent cohomology to detect which elements represent meaningful (long-lived) features of the data and choose a radius for which these features exist. As the vertices of the Vietoris-Rips complex are points in the point cloud, the circular values of the corresponding maps at the vertices describe circular coordinates of the data.

In the present case, persistent cohomology was first applied to the grid-cell population activity and X was identified as the Vietoris-Rips complex for which the two longest-lived one-dimensional bars in the barcode (representing each of the two circles of the torus) existed. To define the desired toroidal coordinates on a domain that was as large as possible, we chose the complex given at the scale of the birth plus 0.99 times the lifetime of the second longest-lived one-dimensional bar in the barcode^{22,59,61}. Next, the cocycle representatives (given by the persistent cohomology implementation of Ripser^{51,52}) of each of the chosen 1D bars defined \mathbb{Z}_{47} -values for each of the edges in the complex. These edge values were then lifted to integer coefficients and subsequently smoothed by minimizing the sum over all edges (using the scipy implementation ‘lsmr’). The values on the vertices (points) of each edge followed from the edge values and gave the circular parametrizations of the point cloud. The product of the two parametrizations thus provided a mapping from the neural activity

to the two-dimensional torus—that is, giving a toroidal coordinatization (decoding) of the data.

As persistent cohomology was computed for a reduced dataset of 1,200 points and therefore circular parametrizations were obtained only for this point cloud, each parametrization was interpolated to the population activity from the rest of the session(s). First, the 1,200 toroidal coordinates were weighted by the normalized ('z-scored') firing rates of the cells at those time points, obtaining a distribution of the coordinates for each grid cell. The decoded toroidal coordinates were then computed by finding the mass centre of the summed distributions, weighted by the population activity vector to be decoded. These activity vectors were calculated by first applying a Gaussian smoothing kernel of 15-ms standard deviation to delta functions centred on spike times, sampling at 10-ms intervals and then z-scoring the activity of each cell independently. Time intervals that contained no spikes from any cell were subsequently excluded. When decoding was used to assess or compare the tuning properties of single cells (for example, comparison of toroidal versus spatial description), the coordinates were computed using the weighted sum of the distributions of the other cells; that is, the contribution of the cell to be assessed or compared was removed. When comparing preservation of toroidal tuning across two sessions, coordinates were interpolated either using the toroidal parametrization in each session independently ('Separate') or using the same toroidal parametrization in both sessions ('Common').

Toroidal rate map visualization

For visualization, toroidal firing rate maps were calculated in the same way as the physical space covariate (see 'Spatial position and direction tuning'), first binning the toroidal surface into a square grid of $7.2^\circ \times 7.2^\circ$ bins and computing the average spike rate in each position bin. However, for toroidal maps, it was necessary to address the 60° angle between the toroidal axes before smoothing. After binning the toroidal coordinates, the rate map was 'straightened' by shifting the bins along the x axis ('horizontally') the length of $(y \bmod 2)/2$ bins, where y is the vertical enumeration of the given bin. Copies of the rate map were then tiled in a three-by-three square (similar to Extended Data Fig. 5d), before applying the closing and smoothing operations as for the spatial firing rate map. The single toroidal rate map was finally recovered by cutting out the centre tile, rotating it 90° and defining 15° shear angles along both the x and the y axis to correct for the 60° offset between them.

Comparison of spatial periodicity

Differences in grid periodicity between OF and WW environments were quantified for a given cell by comparing the grid scores in the two behavioural conditions. Two

alternative methods were used to generate the spatial autocorrelograms for this comparison: (1) comparing the autocorrelograms for OF and WW directly; and (2) comparing autocorrelograms for OF and WW after first equalizing the spatial coverage between the two conditions.

For method (1), rate maps were calculated as specified in the above section ‘Spatial position and direction tuning’, using the same grid of 3×3 -cm bins for both environments. This set of bins spanned the entirety of the OF arena and covered most of the WW track apart from some small regions at the outer extrema, which were discarded for the purpose of this analysis. For each of the two rate maps, the autocorrelogram was computed and the grid score was calculated.

Method (2) was similar to method (1), except that the cell’s OF rate map was converted into a ‘masked OF’ rate map, by removing all bins that were unvisited by the rat in the WW session. This effectively equalized the position coverage between the two conditions, and thus allowed for a more valid comparison.

Toroidal versus spatial description

The explanatory significance of the toroidal description was evaluated by comparing statistical measures of how well the toroidal coordinates explained neural activity on the torus and in physical space. For a fair comparison, it was important to avoid overfitting, which might occur if a toroidal parametrization of a point cloud is used to describe that same set of data points. Two precautions were taken to avoid such overfitting: first, the data were decoded using the toroidal parametrization from a different condition (an OF session for a WW recording and a WW session for an OF recording), and second, the cell for which the statistical measurement was made was omitted from the decoding.

The comparison of toroidal and environmental representations also accounted for tracking error in the physical position estimate, which mainly resulted from the approximately 4 cm vertical offset of the tracking device above the rat’s head. This causes a discrepancy when the angle α between the animal’s zenith and the axis of gravitation is different from 0° , measured as $4 \tan(\alpha)$ cm. The mean discrepancy in the recorded position data was measured to 1.5 cm. To account for this error of the position estimate, proportional Gaussian noise was added to the toroidal coordinates, using a standard deviation of $1.5 \text{ cm}/\Omega$, where Ω denotes the grid spacing of the particular grid-cell module, estimated from the mean period of the fitted cosine waves of the toroidal coordinates in the open field (see ‘Toroidal alignment’).

Information content

The information content (I) was calculated as previously described⁴², to quantify and compare the amount of information carried by single-cell activity about the location on the torus and physical space per spike. Both covariates were binned in a $M = 15 \times 15$ grid of square bins. For each bin j , the average firing rate \mathbf{f}_j (given in spikes per second), and the occupancy ratio, \mathbf{p}_j were computed. The information content for each grid cell was then given as:

$$\$I=\frac{1}{\bar{\mathbf{f}}}\sum_{j=1}^M \mathbf{f}_j \log \frac{\mathbf{f}_j}{\bar{\mathbf{f}}}, \$$$

where $\bar{\mathbf{f}}$ is the mean firing rate of the cell across the entire session.

Note that although the rate maps for physical space have multiple firing fields, whereas the toroidal rate maps have single firing fields, we expect the spatial information to be comparable, as the measure primarily depends on the ratio of bins with high firing activity. This number should be comparable as the firing field size (in bins) will be inversely related to the number of fields in the rate map, assuming that the discretization of the map captures the relevant firing rate variations. For example, given a similar binning of space, a larger OF environment will include more fields, but the number of bins per field will decrease correspondingly. The binning used should be sufficient to resolve the smallest fields, as the same discretization was used in classifying the grid cells in the recorded population.

Deviance explained

Deviance explained was computed to measure how well a Poisson GLM model fitted to the spike count was at representing the data, using either the toroidal coordinates or the tracked position as regressors. A similar set-up was used to that of a previous study⁶², with a smoothness prior for the GLM to avoid overfitting.

Both the toroidal and spatial coordinates were binned into a 15×15 grid of bins, and GLM design matrices were built with entries $X_i(t) = 1$ if the covariate at time t fell in the i -th bin and $X_i(t) = 0$ otherwise.

The Poisson probability of recording k spikes in time bin t is:

$$\$P(k|\mu(t),\beta)=\exp(-\mu(t))\frac{\mu(t)^k}{k!}, \$$$

where $\mu(t)=\exp(\sum_i \beta_i X_i(t))$ is the expected firing rate in time bin t . The parameters β of the Poisson GLM were optimized for each covariate by minimizing the cost function:

$$\$L(\beta | \mu(t), \gamma, k) = -\sum_{t} (\rm{rm}\{l\} \cdot \rm{rm}\{n\} \cdot P(k(t)|\mu(t), \beta)) + \frac{1}{2} \gamma \sum_{i,j \in N} ((\beta_i - \beta_j))^2$$

where N is the set of neighbour pairs. The first term is the negative log-likelihood of the spike count in the given time bin, whereas the second term puts a penalty on large differences in neighbouring parameters, enforcing smoothness in the covariate response of the predicted spike count.

The parameters, β , were initialized to zero and then modified to minimize the loss function by first running two iterations of gradient descent, before optimizing using the ‘l-bfgs-b’-algorithm (as implemented in the ‘scipy.optimize’-module) with ‘gtol’=1e-5 as the cut-off threshold, and finally running two more iterations of gradient descent. A three-fold cross validation procedure was used, repeatedly fitting the model to two-thirds of the data and testing on the held-out last third.

The smoothness hyperparameter γ was optimized a priori on each grid-cell module based on the summed likelihood, testing $\gamma \in (1, \sqrt{10}, 10, \sqrt{1,000})$, and found to be either 1 or $\sqrt{10}$ in all cases.

Similarly, after fitting a null model (using only the intercept term) and the saturated model (perfectly fitting each spike count), the deviance explained could be computed as:

$$\$1 - \frac{\{\{\rm{rm}\{ll\}\}\}_{\{\{\rm{rm}\{s\}\}\}} - \{\{\rm{rm}\{ll\}\}\}_{\{\{\rm{rm}\{p\}\}\}}}{\{\{\rm{rm}\{ll\}\}\}_{\{\{\rm{rm}\{s\}\}\}} - \{\{\rm{rm}\{ll\}\}\}_{\{0\}}}$$

where ll_p , ll_0 and ll_s denote the cross-validated log likelihood of the fitted model, the null model and the saturated model, respectively. This provides a normalized comparison describing the difference between the fitted model and the idealized model.

Toroidal alignment

To infer a geometric interpretation of the tori, as characterized via the cohomological decoding, and compare the toroidal parametrizations across modules and conditions, two cosine waves of the form $\cos(\omega t + k)$ were fitted to the OF mappings of the decoded circular coordinates (Extended Data Fig. 5a), where t is the centre 100^2 -bins of a $540^\circ \times 540^\circ$ -valued 150^2 -bin grid rotated θ degrees. The parameters (ω, k, θ) were optimized by minimizing the square difference between the cosine waves and the cosine of the mean of the circular coordinates in 100^2 bins of the physical environment (smoothed using a Gaussian kernel with 1-bin standard deviation). Estimates were first obtained by finding the minimum when testing all combinations in the following

intervals, each discretized in 10 steps: $\omega \in [1,6]$, $\phi \in [0, 360)$ and $\theta \in [0,180)$. The parameters of the cosine waves were further optimized using the ‘slsqp’-minimization algorithm (as implemented in the ‘scipy.optimize’-module using default hyperparameters). The period of each cosine wave was computed as $1.5 \text{ m}/\omega$, giving a spatial scale estimate of the grid-cell modules.

As circular coordinates have arbitrary origin and orientation (that is, clockwise or counterclockwise evolution) we needed to realign the directions of the circular coordinates to compare these across modules and sessions (see Extended Data Fig. 4b). The clockwise orientation of each circular coordinate was first determined by noting whether $(\omega t + k)$ or $360^\circ - (\omega t + k)$ best fit the spatial mapping of the circular means of the toroidal coordinates, and subsequently reoriented to obtain the same orientation for both coordinates. The coordinate for which $\cos(\theta)$ was largest (intuitively, the ‘x axis’) was then defined as the first coordinate (denoted ϕ_1 , with parameters $(\omega_1, k_1, \theta_1)$) and the other as the second coordinate (ϕ_2). Although (ϕ_1, ϕ_2) fully describe the toroidal location, the hexagonal torus allows for three axes, and the two axes obtained are thus oriented at either 60° or 120° relative to each other (see Extended Data Fig. 5b). The difference in directions was given by $\theta_1 - \theta_2$ and if this difference was greater than 90° , ϕ_2 was replaced with $\phi_2 + 60^\circ \cdot \phi_1$. Finally, the origin of the coordinates was aligned to a fixed reference, by subtracting the mean angular difference between the decoded coordinates and the corresponding coordinates obtained when using the toroidal parametrization of the reference OF session.

For visualization (Extended Data Fig. 5), it was furthermore necessary, in some cases, to rotate both vectors of the rhombi 30 degrees depending on whether one of the axes was directed outside of the box.

Preservation of toroidal tuning

Centre-to-centre distance and Pearson correlation were computed between toroidal tuning maps of different sessions to measure the degree of preservation between the toroidal descriptions.

First, the preferred toroidal firing location for each cell was computed as the centre of mass of the toroidal firing distribution:

```
 $$\langle T \rangle = \arctan \left( \frac{\sum_i y_i \sin \theta_i}{\sum_i y_i \cos \theta_i} \right) $$

```

where y_i denotes the mean spike count of the given cell in the i -th bin whose binned toroidal coordinates are given by θ_i . The distance between mass centres found in two

sessions (' S_1 ' and " S_2 ") was then defined as:

$$\$d = \sqrt{\arctan(2\{\sin(\{T\}_{\rm c})\}^{\{S\}_2} - \{T\}_{\rm c}^{\{S\}_1}, \cos(\{T\}_{\rm c})^{\{S\}_2} - \{T\}_{\rm c}^{\{S\}_1})^2)} \sqrt{2}$$

where $\|\cdot\|_2$ refers to the L_2 -norm.

Pearson correlation between two tuning maps was computed by flattening the smoothed 2D rate maps to 1D arrays and calculating the correlation coefficient, r , using the 'pearsonr'-function given in the 'scipy.stats'-library.

To determine how much the preservation of the toroidal representations across two sessions (measured with Pearson correlation and peak distance) differed from a random distribution, the indices of the cells in one of the sessions were randomly re-ordered before computing correlation and distance for the pair of conditions. This process was repeated 1,000 times, and the P value was calculated from the rank of the original r value or distance with respect to the shuffled distribution.

Classification of grid cells

Temporal autocorrelograms were computed, for each cell, by calculating a histogram of the temporal lags between every spike and all surrounding spikes within a 200 ms window, using 1 ms bins. The histogram was then divided by the value of the zero-lag bin, which was subsequently set to zero. The autocorrelogram was smoothed using a gaussian kernel with smoothing window 4 ms. Considering the autocorrelograms of all modules during OF foraging (day 2 for R1–3) as a point cloud, the cosine distances between all points were calculated, and hence each point's 80 nearest neighbours were found. This defined a graph in which each point described a vertex and the neighbour pairs gave rise to edges. A density estimate was then calculated as the exponential of the negative distances summed over each neighbour for each point. The graph and the density estimate were given as the input to the Gudhi implementation⁶³ of ToMATo⁶⁴. ToMATo uses a hill-climbing procedure to find modes of the density function and uses persistence to determine stable clusters. In the present case, the algorithm finds three long-lived clusters.

Minimum number of cells for torus detection

To address the question of how many cells are minimally needed to expect to see toroidal structure, random samples of $n = 10, 20, \dots, 140$ cells were taken from R2 ($n = 149$ cells) during OF foraging, and the same topological analysis was repeated as for the whole population. The cells were resampled 1,000 times for each number of cells

in the subsample. To determine whether toroidal structure was detected, a heuristic was introduced based on the circular parameterization given by the two most persistent 1D bars in the barcode mapped onto physical space. An estimate of the resulting planar representation of the torus was obtained by fitting planar cosine waves to each mapping (see ‘Toroidal alignment’). For the analysis to be determined ‘successful’ in detecting toroidal structure, we required: (i) the mean value of the least-squares fitting (across bins of the mapping) to be less than 0.25; (ii) the angle of the rhombus to be close to 60° (between 50° and 70°); and (iii) the side lengths to be within 25% of each other.

Toroidal peak detection

The number of peaks per toroidal rate map was detected to assert the number of grid cells whose toroidal rate map portrayed single fields. First, 1,000 points were sampled from the toroidal distribution given by the mean activity of each cell in 150×150 bins of the stacked toroidal surface (that is, as described in ‘Toroidal rate map visualization’, each 50×50 -binned toroidal rate map is first ‘straightened’ and subsequently stacked in 3×3 to address the toroidal boundaries) and then spatially smoothed using a Gaussian kernel with smoothing widths 0, 1, 2, ..., 10 bins with mode set to ‘constant’ in the ‘scipy.gaussian_filter’ function. Next, the points were clustered by computing a density estimate, using the Euclidean distance, and defining neighbours as points closer than 5 bins. Cluster labels were iteratively assigned to each point and all its neighbours in a downhill manner, instantiating a new cluster identity if the point was not already labelled. Finally, the centroids for each cluster were computed and counted as a peak depending on whether its position fell within the centre 50×50 bins of the stacked rate maps.

Simulated CAN models

To confirm the expected outcomes of topological analyses of grid cell CAN models, grid cells were simulated using two different, noiseless CAN models (Extended Data Fig. 7).

First, a 56×44 grid cell network was simulated based on the CAN model proposed previously⁹, but using solely lateral inhibition (for details see ref. 11) in the connectivity matrix, W . The animal movement was given as the first 1,000 s of the recorded trajectory of rat ‘R’ during OF session, originally sampled at 10 ms, and interpolated to 2-ms time steps. The speed, $v(t)$, and head direction $\theta(t)$ of the animal was calculated as the (unsmoothed) displacement in position for every time step. The activity, s , was updated as:

$$\$\$ \{ \{ \bf{s} \} \} _{i+1} = \{ \{ \bf{s} \} \} _i + \frac{1}{\tau} (- \{ \{ \bf{s} \} \} _i + \{ (I + \{ \{ \bf{s} \} \} _i) \cdot W + \alpha v(t) \cos(\theta(t) - \theta_0) \} \} _i)_{+}, \$\$$$

where $\{ \dots \}_+$ is the Heaviside function and $\{ \theta \}_{\sim}$ is the population vector of preferred head directions. The following parameters were used: $I = 1$, $\alpha = 0.15$, $l = 2$, $W_0 = -0.01$, $R = 20$ and $\tau = 10$, and let the activity pattern stabilize by first initializing to random and performing 2,000 updates, disregarding animal movement. For computational reasons, the activity was set to 0 if $s_i < 0.0001$. The simulation was subsequently downsampled keeping only every 5th time frame.

Next, a 20×20 grid-cell network was simulated, for a synthetically generated OF trajectory ('random walk'), based on the twisted torus model formulated in a previous study¹⁰. The parameter values and the code for computing both the grid cell network (choosing a single grid scale by defining the parameter 'grid_gain' = 0.04) and the random navigation (using 5,000 time steps) were given by the implementation by Santos Pata⁶⁵.

Idealized torus models

To compare the results of both the original and simulated grid cell networks with point clouds where the topology is known, a priori, to be toroidal, points were sampled from a square and a hexagonal torus. First, a 50×50 (angle) mesh grid (θ_1, θ_2) was created in the square $[0, 2\pi] \times [0, 2\pi]$ and slight Gaussian noise ($\epsilon = 0.1 \cdot N(0, 1)$) was added to each angle. The square torus was then constructed via the 4D Clifford torus parametrization: $(\cos(\theta_1), \sin(\theta_1), \cos(\theta_2), \sin(\theta_2))$. The hexagonal torus was constructed using the 6D embedding: $(\cos(\theta_1), \sin(\theta_1), \cos(a_1\theta_1 + \theta_2), \sin(a_1\theta_1 + \theta_2), \cos(a_2\theta_1 + \theta_2), \sin(a_2\theta_1 + \theta_2))$, where $a_1 = 1/\sqrt{3}$ and $a_2 = -1/\sqrt{3}$.

Histology and recording locations

Rats were given an overdose of sodium pentobarbital and were perfused intracardially with saline followed by 4% formaldehyde. The extracted brains were stored in formaldehyde and a cryostat was used to cut 30- μm sagittal sections, which were then Nissl-stained with cresyl violet. The probe shank traces were identified in photomicrographs, and a map of the probe shank was aligned to the histology by using two reference points that had known locations in both reference frames: (1) the tip of the probe shank; and (2) the intersection of the shank with the brain surface. In all cases, the shank traces were near-parallel to the cutting plane, therefore it was deemed sufficient to perform a flat 2D alignment in a single section where most of the shank

trace was visible. The aligned shank map was then used to calculate the anatomical locations of individual electrodes (Extended Data Fig. 1).

Data analysis and statistics

Data analyses were performed with custom-written scripts in Python and MATLAB. Open-source Python packages used were: umap (version 0.3.10), ripser (0.4.1), numba (0.48.0), scipy (1.4.1), numpy (1.18.1), scikit-learn (0.22.1), matplotlib (3.1.3), h5py (2.10.0) and gudhi (3.4.1.post1). Samples included all available cells that matched the classification criteria for the relevant cell type. Power analysis was not used to determine sample sizes. The study did not involve any experimental subject groups; therefore, random allocation and experimenter blinding did not apply and were not performed. All statistical tests were one-sided.

The most intensive computations were performed on resources provided by the NTNU IDUN/EPIC computing cluster⁶⁶.

Additional discussion

The demonstration that populations of grid cells operate on a toroidal manifold, which is preserved across environments and behavioural states, confirms a central prediction of CAN models. The present observations provide the first—to our knowledge—population-level visualization of a two-dimensional CAN manifold, though there is accumulating evidence for one-dimensional CANs in a number of neural systems. The most powerful support for the latter has been obtained in fruit flies, in which CAN-like dynamics can be visualized in a ring of serially connected orientation-tuned cells of the central complex^{67,68,69}. In mammals, analysis of data from dozens of simultaneously recorded head direction cells has shown that population activity in these cells faithfully traverses a conceptual ring^{22,23,24}, in accordance with ring-attractor models^{17,18,19}. Dynamics along low-dimensional manifolds with line, ring, or sheet topologies is also thought to underlie a wide range of other mammalian brain functions that operate on continuous scales, spanning from visual orientation tuning¹⁴ to neural operations underlying place-cell formation^{70,71,72}, as well as motor control⁷³, decision making and action selection^{74,75,76}, and certain forms of memory^{39,77,78,79,80}. The present analyses provide a visualization of 2D CAN dynamics in pure grid cells within a module and, together with the previous work, point to a widespread implementation of CAN dynamics in the brain. The existence of CAN structure to constrain activity to low-dimensional manifolds does not preclude additional mechanisms for pattern formation, however. Grid cell patterns may emerge also by feedforward mechanisms^{12,38,81,82,83,84,85,86}. Such mechanisms may operate in parallel with recurrent networks⁸⁷ and may even be the primary mechanism for grid-like firing

at early stages of development, before the full maturation of recurrent connectivity^{11,88,89,90}.

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 the current study are available at https://figshare.com/articles/dataset/Toroidal_topology_of_population_activity_in_grid_cells/16764508. [Source data](#) are provided with this paper.

Code availability

Code for reproducing the analyses in this article is available at https://figshare.com/articles/dataset/Toroidal_topology_of_population_activity_in_grid_cells/16764508.

References

1. 1.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. Path integration and the neural basis of the ‘cognitive map’. *Nat. Rev. Neurosci.* **7**, 663–678 (2006).

2. 2.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).

3. 3.

Stensola, H. et al. The entorhinal grid map is discretized. *Nature* **492**, 72–78 (2012).

4. 4.

Fyhn, M., Hafting, T., Treves, A., Moser, M.-B. & Moser, E. I. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* **446**, 190–194 (2007).

5. 5.

Yoon, K. et al. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nat. Neurosci.* **16**, 1077–1084 (2013).

6. 6.

Gardner, R. J., Lu, L., Wernle, T., Moser, M.-B. & Moser, E. I. Correlation structure of grid cells is preserved during sleep. *Nat. Neurosci.* **22**, 598–608 (2019).

7. 7.

Trettel, S. G., Trimper, J. B., Hwaun, E., Fiete, I. R. & Colgin, L. L. Grid cell co-activity patterns during sleep reflect spatial overlap of grid fields during active behaviors. *Nat. Neurosci.* **22**, 609–617 (2019).

8. 8.

Fuhs, M. C. & Touretzky, D. S. A spin glass model of path integration in rat medial entorhinal cortex. *J. Neurosci.* **26**, 4266–4276 (2006).

9. 9.

Burak, Y. & Fiete, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput. Biol.* **5**, e1000291 (2009).

10. 10.

Guanella, A., Kiper, D. & Verschure, P. A model of grid cells based on a twisted torus topology. *Int. J. Neural Syst.* **17**, 231–240 (2007).

11. 11.

Couey, J. J. et al. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nat. Neurosci.* **16**, 318–324 (2013).

12. 12.

Kropff, E. & Treves, A. The emergence of grid cells: intelligent design or just adaptation? *Hippocampus* **18**, 1256–1269 (2008).

13. 13.

Amari, S. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* **27**, 77–87 (1977).

14. 14.

Ben-Yishai, R., Bar-Or, R. L. & Sompolinsky, H. Theory of orientation tuning in visual cortex. *Proc. Natl Acad. Sci. USA* **92**, 3844–3848

(1995).

15. 15.

Seung, H. S. How the brain keeps the eyes still. *Proc. Natl Acad. Sci. USA* **93**, 13339–13344 (1996).

16. 16.

Taube, J. S., Muller, R. U. & Ranck, J. B. Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J. Neurosci.* **10**, 420–435 (1990).

17. 17.

Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S. & McNaughton, B. L. A model of the neural basis of the rat's sense of direction. *Adv. Neural Inf. Process. Syst.* **7**, 173–180 (1995).

18. 18.

Redish, A. D., Elga, A. N. & Touretzky, D. S. A coupled attractor model of the rodent head direction system. *Network* **7**, 671–685 (1996).

19. 19.

Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* **16**, 2112–2126 (1996).

20. 20.

Yoganarasimha, D., Yu, X. & Knierim, J. J. Head direction cell representations maintain internal coherence during conflicting proximal and distal cue rotations: comparison with hippocampal place cells. *J. Neurosci.* **26**, 622–631 (2006).

21. 21.

Peyrache, A., Lacroix, M. M., Petersen, P. C. & Buzsáki, G. Internally organized mechanisms of the head direction sense. *Nat. Neurosci.* **18**, 569–575 (2015).

22. 22.

Rybakken, E., Baas, N. & Dunn, B. Decoding of neural data using cohomological feature extraction. *Neural Comput.* **31**, 68–93 (2019).

23. 23.

Chaudhuri, R., Gercek, B., Pandey, B., Peyrache, A. & Fiete, I. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nat. Neurosci.* **22**, 1512–1520 (2019).

24. 24.

Rubin, A. et al. Revealing neural correlates of behavior without behavioral measurements. *Nat. Commun.* **10**, 4745 (2019).

25. 25.

Jun, J. J. et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).

26. 26.

Steinmetz, N. A. et al. Neuropixels 2.0: a miniaturized high-density probe for stable, long-term brain recordings. *Science* **372**, eabf4588 (2021).

27. 27.

Kang, L., Xu, B. & Morozov, D. Evaluating state space discovery by persistent cohomology in the spatial representation system. *Front. Comput. Neurosci.* **15**, 616748 (2021).

28. 28.

Sargolini, F. et al. Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* **312**, 758–762 (2006).

29. 29.

Barry, C., Hayman, R., Burgess, N. & Jeffery, K. J. Experience-dependent rescaling of entorhinal grids. *Nat. Neurosci.* **10**, 682–684 (2007).

30. 30.

Stensola, T., Stensola, H., Moser, M.-B. & Moser, E. I. Shearing-induced asymmetry in entorhinal grid cells. *Nature* **518**, 207–212 (2015).

31. 31.

Krupic, J., Bauza, M., Burton, S., Barry, C. & O’Keefe, J. Grid cell symmetry is shaped by environmental geometry. *Nature* **518**, 232–235 (2015).

32. 32.

Butler, W. N., Hardcastle, K. & Giocomo, L. M. Remembered reward locations restructure entorhinal spatial maps. *Science* **363**, 1447–1452 (2019).

33. 33.

Boccara, C. N., Nardin, M., Stella, F., O’Neill, J. & Csicsvari, J. The entorhinal cognitive map is attracted to goals. *Science* **363**, 1443–1447 (2019).

34. 34.

Latuske, P., Toader, O. & Allen, K. Interspike intervals reveal functionally distinct cell populations in the medial entorhinal cortex. *J. Neurosci.* **35**, 10963–10976 (2015).

35. 35.

Newman, E. L. & Hasselmo, M. E. Grid cell firing properties vary as a function of theta phase locking preferences in the rat medial entorhinal cortex. *Front. Syst. Neurosci.* **8**, 193 (2014).

36. 36.

Csordás, D. É., Fischer, C., Nagele, J., Stemmler, M. & Herz, A. V. M. Spike afterpotentials shape the *in vivo* burst activity of principal cells in medial entorhinal cortex. *J. Neurosci.* **40**, 4512–4524 (2020).

37. 37.

Finkelstein, A. et al. Three-dimensional head-direction coding in the bat brain. *Nature* **517**, 159–164 (2015).

38. 38.

Ginosar, G. et al. Locally ordered representation of 3D space in the entorhinal cortex. *Nature* **596**, 404–409 (2021).

39. 39.

Sussillo, D. & Barak, O. Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Comput.* **25**, 626–649 (2013).

40. 40.

Sorscher, B., Mel, G. C., Ocko, S. A., Giocomo, L. & Ganguli, S. A unified theory for the computational and mechanistic origins of grid cells. Preprint at <https://doi.org/10.1101/2020.12.29.424583> (2020).

41. 41.

Darshan, R. & Rivkind, A. Learning to represent continuous variables in heterogeneous neural networks. Preprint at <https://doi.org/10.1101/2021.06.01.446635> (2021).

42. 42.

Skaggs, W., Mcnaughton, B. & Gothard, K. An information-theoretic approach to deciphering the hippocampal code. *Adv. Neural Inf. Process. Syst.* **5**, 1030–1037 (1992).

43. 43.

McInnes, L., Healy, J. & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2018).

44. 44.

McInnes, L. UMAP <https://github.com/lmcinnes/umap>.

45. 45.

Lee, A. & Wilson, M. A. Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* **36**, 1183–1194 (2002).

46. 46.

Montgomery, S. M., Sirota, A. & Buzsáki, G. Theta and gamma coordination of hippocampal networks during waking and rapid eye movement sleep. *J. Neurosci.* **28**, 6731–6741 (2008).

47. 47.

Meehan, C., Ebrahimian, J., Moore, W. & Meehan, S. *Uniform Manifold Approximation and Projection (UMAP)* <https://www.mathworks.com/matlabcentral/fileexchange/71902> (MATLAB, 2021).

48. 48.

Bellman, R. *Dynamic Programming* (Princeton Univ. Press, 1957).

49. 49.

Kloke, J. & Carlsson, G. Topological de-noising: strengthening the topological signal. Preprint at <https://arxiv.org/abs/0910.5947> (2009).

50. 50.

Spivak, D. I. Metric realization of fuzzy simplicial sets
<https://www.semanticscholar.org/paper/metric-realization-of-fuzzy-simplicial-sets-spivak/a73fb9d562a3850611d2615ac22c3a8687fa745e>
(Semantic Scholar, 2009).

51. 51.

Tralie, C., Saul, N. & Bar-On, R. Ripser. py: a lean persistent homology library for python. *J. Open Source Softw.* **3**, 925 (2018).

52. 52.

Bauer, U. Ripser: efficient computation of Vietoris-Rips persistence barcodes. *J Appl. Comput. Topol.* **5**, 391–423 (2021).

53. 53.

Kobak, D. & Linderman, G. C. Initialization is critical for preserving global data structure in both t-SNE and UMAP. *Nat. Biotechnol.* **39**, 156–157 (2021).

54. 54.

Singh, G. et al. Topological analysis of population activity in visual cortex. *J. Vis.* **8**, 1–18 (2008).

55. 55.

Giusti, C., Pastalkova, E., Curto, C. & Itskov, V. Clique topology reveals intrinsic geometric structure in neural correlations. *Proc. Natl Acad. Sci. USA* **112**, 13455–13460 (2015).

56. 56.

Dabaghian, Y., Mémoli, F., Frank, L. & Carlsson, G. A topological paradigm for hippocampal spatial map formation using persistent homology. *PLoS Comput. Biol.* **8**, e1002581 (2012).

57. 57.

Spreemann, G., Dunn, B., Botnan, M. B. & Baas, N. A. Using persistent homology to reveal hidden covariates in systems governed by the kinetic Ising model. *Phys. Rev. E* **97**, 032313 (2018).

58. 58.

Baas, N. A. On the concept of space in neuroscience. *Curr. Opin. Syst. Biol.* **1**, 32–37 (2017).

59. 59.

De Silva, V., Morozov, D. & Vejdemo-Johansson, M. Persistent cohomology and circular coordinates. *Discrete Comput. Geom.* **45**, 737–759 (2011).

60. 60.

Hatcher, A. *Algebraic Topology* (Cambridge University Press, 2002).

61. 61.

Perea, J. A. in *Topological Data Analysis: The Abel Symposium 2018* (eds Baas, N. A. et al.) 435–458 (Springer, 2020).

62. 62.

Ledergerber, D. et al. Task-dependent mixed selectivity in the subiculum. *Cell Rep.* **35**, 109175 (2021).

63. 63.

The GUDHI Project. *GUDHI User and Reference Manual* (GUDHI Editorial Board, 2021).

64. 64.

Chazal, F., Guibas, L. J., Oudot, S. Y. & Skraba, P. Persistence-based clustering in Riemannian manifolds. *J. ACM* **60**, 41 (2013).

65. 65.

Santos Pata, D. *Grid Cells*

<https://github.com/DiogoSantosPata/gridcells> (2020).

66. 66.

Själander, M., Jahre, M., Tufte, G. & Reissmann, N. EPIC: an energy-efficient, high-performance GPGPU computing research infrastructure. Preprint at <https://arxiv.org/abs/1912.05848> (2020).

67. 67.

Seelig, J. D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521**, 186–191 (2015).

68. 68.

Kim, S. S., Rouault, H., Druckmann, S. & Jayaraman, V. Ring attractor dynamics in the *Drosophila* central brain. *Science* **356**, 849–853 (2017).

69. 69.

Green, J. et al. A neural circuit architecture for angular integration in *Drosophila*. *Nature* **546**, 101–106 (2017).

70. 70.

McNaughton, B. L. et al. Deciphering the hippocampal polyglot: the hippocampus as a path integration system. *J. Exp. Biol.* **199**, 173–185 (1996).

71. 71.

Samsonovich, A. & McNaughton, B. L. Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.* **17**, 5900–5920 (1997).

72. 72.

Tsodyks, M. & Sejnowski, T. J. Associative memory and hippocampal place cells. *Int. J. Neural Syst.* **6**, 81–86 (1995).

73. 73.

Aksay, E. et al. Functional dissection of circuitry in a neural integrator. *Nat. Neurosci.* **10**, 494–504 (2007).

74. 74.

Wang, X.-J. Decision making in recurrent neuronal circuits. *Neuron* **60**, 215–234 (2008).

75. 75.

Machens, C. K., Romo, R. & Brody, C. D. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* **307**, 1121–1124 (2005).

76. 76.

Lukashin, A. V. & Georgopoulos, A. P. A dynamical neural network model for motor cortical activity during movement: population coding of movement trajectories. *Biol. Cybern.* **69**, 517–524 (1993).

77. 77.

Romani, S. & Tsodyks, M. Continuous attractors with morphed/correlated maps. *PLoS Comput. Biol.* **6**, e1000869 (2010).

78. 78.

Compte, A., Brunel, N., Goldman-Rakic, P. S. & Wang, X. J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* **10**, 910–923 (2000).

79. 79.

Wimmer, K., Nykamp, D. Q., Constantinidis, C. & Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* **17**, 431–439 (2014).

80. 80.

Koyluoglu, O. O., Pertzov, Y., Manohar, S., Husain, M. & Fiete, I. R. Fundamental bound on the persistence and capacity of short-term memory stored as graded persistent activity. *eLife* **6**, e22225 (2017).

81. 81.

Stepanyuk, A. Self-organization of grid fields under supervision of place cells in a neuron model with associative plasticity. *Biol. Inspired Cogn. Archit.* **13**, 48–62 (2015).

82. 82.

Dordek, Y., Soudry, D., Meir, R. & Derdikman, D. Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *eLife* **5**, e10094 (2016).

83. 83.

Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nat. Neurosci.* **20**, 1643–1653 (2017).

84. 84.

D’Albis, T. & Kempter, R. A single-cell spiking model for the origin of grid-cell patterns. *PLoS Comput. Biol.* **13**, e1005782 (2017).

85. 85.

Monsalve-Mercado, M. M. & Leibold, C. Hippocampal spike-timing correlations lead to hexagonal grid fields. *Phys. Rev. Lett.* **119**, 038101 (2017).

86. 86.

Weber, S. N. & Sprekeler, H. Learning place cells, grid cells and invariances with excitatory and inhibitory plasticity. *eLife* **7**, e34560 (2018).

87. 87.

Si, B., Kropff, E. & Treves, A. Grid alignment in entorhinal cortex. *Biol. Cybern.* **106**, 483–506 (2012).

88. 88.

Langston, R. F. et al. Development of the spatial representation system in the rat. *Science* **328**, 1576–1580 (2010).

89. 89.

Wills, T. J., Cacucci, F., Burgess, N. & O’Keefe, J. Development of the hippocampal cognitive map in preweanling rats. *Science* **328**, 1573–1576 (2010).

90. 90.

Donato, F., Jacobsen, R. I., Moser, M.-B. & Moser, E. I. Stellate cells drive maturation of the entorhinal–hippocampal circuit. *Science* **355**, eaai8178 (2017).

91. 91.

Gray, C. M., Maldonado, P. E., Wilson, M. & McNaughton, B. Tetrodes markedly improve the reliability and yield of multiple single-

unit isolation from multi-unit recordings in cat striate cortex. *J. Neurosci. Methods* **63**, 43–54 (1995).

92. 92.

Hardcastle, K., Maheswaranathan, N., Ganguli, S. & Giocomo, L. M. A multiplexed, heterogeneous, and adaptive code for navigation in medial entorhinal cortex. *Neuron* **94**, 375–387 (2017).

Acknowledgements

We thank M. P. Witter for help with evaluation of recording locations, and A. M. Amundsgård, K. Haugen, K. Jenssen, E. Kråkvik, I. Ulsaker-Janke and H. Waade for technical assistance. The work was supported by a Synergy Grant to E.I.M. and Y.B. from the European Research Council ('KILONEURONS', grant agreement no. 951319); an RCN FRIPRO grant to E.I.M. (grant no. 286225); a Centre of Excellence grant to M.-B.M. and E.I.M. and a National Infrastructure grant to E.I.M. and M.-B.M. from the Research Council of Norway (Centre of Neural Computation, grant number 223262; NORBRAIN, grant number 295721); the Kavli Foundation (M.-B.M. and E.I.M.); the Department of Mathematical Sciences at the Norwegian University of Science and Technology (B.A.D., E.H. and N.A.B.); a direct contribution to M.-B.M. and E.I.M. from the Ministry of Education and Research of Norway; and grants to Y.B. from the Israel Science Foundation (grant no. 1745/18) and the Gatsby Charitable Foundation. Some of the computations were performed on resources provided by the NTNU IDUN/EPIC computing cluster.

Author information

Author notes

1. These authors contributed equally: Richard J. Gardner, Erik Hermansen

Affiliations

1. Kavli Institute for Systems Neuroscience and Centre for Neural Computation, Norwegian University of Science and Technology, Trondheim, Norway

Richard J. Gardner, Benjamin A. Dunn, May-Britt Moser & Edvard I. Moser

2. Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway

Erik Hermansen, Nils A. Baas & Benjamin A. Dunn

3. HHMI Janelia Research Campus, Ashburn, VA, USA

Marius Pachitariu

4. Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

Yoram Burak

5. Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem, Israel

Yoram Burak

Contributions

R.J.G., M.-B.M. and E.I.M. designed experiments. R.J.G. performed experiments. N.A.B., E.H., B.A.D., R.J.G., Y.B. and E.I.M. conceptualized and proposed analyses. E.H. and R.J.G. developed and performed the analyses. M.P. shared unpublished Kilosort software. R.J.G., E.H., B.A.D., Y.B., M.-B.M. and E.I.M. interpreted data. E.H. and R.J.G. visualized data. R.J.G., E.H., B.A.D., Y.B. and E.I.M. wrote the paper, with periodic input from all authors. E.I.M., M.-B.M., B.A.D. and N.A.B. supervised the project. E.I.M., M.-B.M. and Y.B. obtained funding.

Corresponding authors

Correspondence to [Richard J. Gardner](#), [Nils A. Baas](#), [Benjamin A. Dunn](#) or [Edvard I. Moser](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Carina Curto, Arseny Finkelstein 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 Nissl-stained sagittal brain sections showing recording locations for rats Q, R and S.](#)

Red arrows indicate the dorsoventral range of the probe's active recording sites (corresponding to the yellow stripe in the inset). Stippled lines indicate borders between brain regions (MEC, medial entorhinal cortex; PaS, parasubiculum, PrS, presubiculum; PoR, postrhinal cortex). Layers are indicated for MEC (MECII, MECIII). Animal name, hemisphere (L, left; R, right) and shank number (for Rat 'S') are indicated in text above each section. Insets show, for each section, the number of grid cells recorded at each depth on the probe shank (histogram bin sizes 100 µm for Rats 'Q' and 'R', 75 µm for Rat 'S'; total numbers of cells are given in Extended Data Fig. 2g). Only the implanted portion of the probe shank is shown. Counts are

colour-coded according to module identity. Module R1 is subdivided into the two UMAP clusters R1a and R1b (as shown in Extended Data Fig. 2), shown here as two stacked histograms. The yellow stripe on the probe shank indicates the range of active recording sites. The indicated locations of units are subject to measurement error, because the anatomical registration of probe shanks can only be approximately estimated, and furthermore because units may be detected on electrodes up to 50 µm away⁹¹. Note that several modules spanned across hemispheres (see Extended Data Fig. 2g). The cell counts shown for Rat 'R' are from Recording Day 1. The same set of recording sites was used for both recording sessions, and therefore the anatomical distributions of recorded cells were similar between the two sessions.

[Source data](#)

[Extended Data Fig. 2 Grid module identification and properties.](#)

a–d, Clustering of grid modules (**a**, Rat 'Q'; **b**, Rat 'R', day 1; **c**, Rat 'R', day 2; **d**, Rat 'S'). For all experiments, coarse spatial autocorrelograms were first calculated from all cells' OF firing rate maps (n cells as shown in **g**). UMAP was then used to reduce the M -dimensional autocorrelograms (where $M = 668$ spatial bins) to a two-dimensional point cloud, where each point represented the autocorrelogram of a single cell, and distances between points represented the similarity between autocorrelograms. Left scatterplot in **a–d**: 2D point cloud, with points colour-coded according to cluster ID. Clusters were identified by applying the density-based clustering algorithm DBSCAN to the 2D point cloud. In every recording, the largest cluster (in grey, labelled “main”) comprised mainly non-grid cells, and the remaining smaller clusters (coloured) represented different modules of grid cells. The black crosses (“noise”) are identified as outlier data points. The well-isolated clusters formed by grid cells support the notion that these cells are a distinct functional class, in contrast to the claim that grid-like characteristics are expressed by MEC cells to different extents⁹². Right pair of scatterplots in **a–d**: Combinations of three grid parameters (grid score, grid spacing and grid orientation) for co-recorded cells from each recording. Each dot corresponds to one autocorrelogram (one cell). Dots are coloured

by cluster ID as in **a**. **e**, Comparison of grid-cell spatial periodicity in the open-field arena (OF) and on the wagon-wheel track (WW). Top: firing rate map and corresponding autocorrelogram for an example grid cell in OF (left) and WW (right). For the purposes of this comparison, the same position bins were applied to both environments, resulting in cropping of the outermost parts of WW. Colour coding as indicated by scale bar; peak rates 16.1 Hz (OF) and 15.8 Hz (WW); range of autocorrelation values: -0.56 to 0.83 and -0.58 to 0.71 , respectively. Note the more irregular appearance of the autocorrelogram for WW. Bottom: scatter plots showing grid scores of all grid cells in OF (*x* axis) and WW (*y* axis). Colours refer to the module assignment in **a**. Note the bias for points to lie in the lower-right quadrant, reflecting generally higher grid scores in OF than in WW. **f**, As for **e**, but controlling for differences in behavioural coverage of OF and WW environments. It is possible that the lower WW grid scores in **e** were a product of sparser behavioural coverage of the WW environment (animals visited only positions on the track). To control for this possibility, we created “masked OF” (MOF) rate maps by removing spatial bins from the original OF rate map which were not visited by the animal in WW. In all modules, grid scores in the “masked” OF condition were higher than in WW (grid score mean \pm S.E.M. across all cells: OF: 0.677 ± 0.017 , WW: 0.360 ± 0.017 , $N=618$ cells, *P* values for the 6 modules ranged from 1.26×10^{-14} to 0.03 , *Z*-values ranged from 2.12 to 7.71, Wilcoxon signed-rank test). Top row shows the same example cell as in **e** after leaving the same subset of position bins in OF as in WW. Bottom row shows comparison of grid scores for MOF and WW. As in **e**, grid scores are lower for WW, indicating that grid periodicity is reduced in WW even when differences in spatial coverage are accounted for. **g**, Table showing total number of cells and number of pure grid cells and conjunctive grid \times direction cells. **h**, Number of cells (as in **g**) broken down on recording sessions, with session lengths in minutes indicated for open field (OF), wagon wheel (WW), slow-wave sleep (SWS) and REM sleep.

[Source data](#)

[**Extended Data Fig. 3 Preprocessing steps for visualization and detection of toroidal topology.**](#)

A, Flow diagram showing method for extracting low-dimensional embeddings of neural activity. The animal foraged in an OF arena while spikes from 149 grid cells shown in Fig. 1a were recorded (**Aa**; cells are ordered arbitrarily). A 5-second example behavioural trajectory is highlighted, with colour indicating elapsed time. The spike trains were binned in time (N bins) and then smoothed and normalized, yielding a matrix of N -dimensional population activity vectors (**Ab**). After temporally downsampling and z-scoring the neural activity, PCA was applied to the N -dimensional neural activity, yielding a six-dimensional linear embedding (**Ac**). This preserved the grid structure in the activity (Extended Data Fig. 4b, c), while mitigating drawbacks associated with high-dimensional spaces (the “curse of dimensionality”)⁴⁸. The six principal components were then passed through a second, nonlinear, dimensionality reduction step by UMAP, which generated a three-dimensional nonlinear embedding (**Ae(i)**) allowing the toroidal structure to be visualized. UMAP consists of two steps: first, a fuzzy topological graph representation is constructed (i.e. a “Uniform Manifold Approximation” - UMA) using a distance metric in the high-dimensional space (**Ad**); second, to obtain the lower-dimensional projection (P), the coordinates of corresponding points in fewer dimensions are optimized to have a similar fuzzy topological representation. In the persistence analysis, we applied persistent cohomology to the fuzzy topological representation of the high-dimensional point cloud (**Ae(ii)**) and subsequently used cohomological decoding to obtain a two-dimensional projection of the original N -dimensional point cloud (**Ae(iii)**; right, showing a 5-second snippet; left, embedded in 3D, points are coloured by each angular coordinate, whose direction is indicated by a red arrow). **B**, Cohomology can help differentiate topological spaces such as the union of three discs (upper left), a circle (upper right), a sphere (lower left) and a torus (lower right) by counting the number of topological holes (β) in different dimensions. A disc has a 0D hole (a connected component); a circle additionally has a 1D hole; a (hollow) sphere is a connected component and has a 2D hole (a cavity); a torus is a connected component with two 1D holes (illustrated with red circles) and one 2D hole (a cavity in the interior of the torus). **C**, Persistent cohomology tracks the lifetime of topological holes in spaces associated with point clouds. Top: The radius of balls centred at each data point in the point cloud is continuously increased (left to right). The union of the balls forms a space with possible holes. The

lifetime of a hole during expansion of the radius is defined as the radial interval from when the hole first appears until it is filled in. Note the short lifetime of the hole marked with a red circle and the long lifetime of the hole indicated with a yellow circle. Second and third row: The lifetime of each hole of dimension zero (H^0) and one (H^1) in the example in the top row is indicated by the length of a bar (in green) in the barcode diagram. Two 1D holes are detected: the first bar, corresponding to the red hole in the top row, is short and regarded as noise, and the second, corresponding to the yellow hole, is substantially longer and captures the prominent topology of the point cloud.

Source data

Extended Data Fig. 4 Analysis of principal components, number of cells and number of toroidal peaks.

a, Variance explained by the first 15 principal components (PCs) after applying PCA to the n -dimensional neural activity, shown for each module. Note that during OF, a particularly large amount of variance is explained by the first 6 PCs, followed by a sharp drop in the 7th PC, in all modules. A drop in variance explained is also seen after the 6th PC in REM and SWS.

b, The first six PCs contain a grid-like representation at the population level. Each panel shows the mean value of one PC as a function of the animal's position in the OF. PC value is colour-coded as indicated by the scale bar. The 8 first PCs are arranged in descending order of explained variance (columns, from left to right), and are shown for each module (in rows). Note the presence of grid-like structure, which is particularly strong in the first six PCs, irrespective of the grid spacing. These six grid-like PCs correspond to the set with the highest explained variance in **a**. z -scored PC values are indicated by the scale bar (see [Supplementary Methods](#) for theoretical explanation of the six-dimensionality).

c, Line plots showing the goodness-of-fit of a Gaussian GLM model based on the position in the spatial environment (OF) fitted to each principal component (components as in **a**). This is measured (as in Fig. [2d](#)) as the explained deviance of the model showing that the six first components are better explained by space than the subsequent components for each module.

d, Line plots showing the lifetime of the two longest-lived H^1 -bars (longest-lived – “1st”, black;

second longest-lived – “2nd”, blue) divided by the lifetime of the third longest-lived H¹-bar as a function of number of principal components kept in the persistence analysis of R1 day 1 OF ($n = 93$ cells). This heuristic measures how clearly the two longest-lived H¹-bars (expected to be long for a torus) separates from the third (expected to be short), thus indicating how clearly the barcode displays toroidal topology. This is clearly the case when using 6 principal components in this dataset. **e**, The percentage of subsamples of R2 (resampled randomly 1,000 times per number of cells; total $n = 149$ cells) for which toroidal structure was detected in the parameterization given by the two most persistent 1D bars in the barcode (as in Extended Data Fig. 5). Note that approximately 60 cells were needed for the probability of detecting toroidal structure exceed 50%. **f**, Effect of varying spatial smoothing on the number of peaks in toroidal rate maps. The y axis displays the percentage of single-peaked (black) and multi-peaked (blue) toroidal rate maps of all grid cells ($n = 2,727$ cells) pooled across modules and behaviour conditions. The vertical dashed line marks the smoothing width used in Extended Data Fig. 10, and the horizontal dashed line marks 100%. Note that cells with single peaks quickly describe the majority of the pooled cells.

[Source data](#)

[Extended Data Fig. 5 Mapping of decoded circular coordinates onto the open field allows geometrical interpretation of toroidal structure.](#)

a, Top row: Toroidal coordinates given by cohomological decoding from activity of grid module R2 during OF foraging, mapped onto the recording box. In each plot, colour indicates the mean value of the cosine of each of the two circular coordinates. The mappings of both coordinates show 2D striped patterns, with similar periods but distinct angles. Bottom row: A cosine wave is fitted to each coordinate to obtain the direction of the toroidal axes. The period and angle of the cosine wave in the plane may be represented by spatial vectors, **v** and **w**, with corresponding length and orientation. Note the clear transversality of the two circles, expressed in the directions of the two vectors, further confirming the toroidal identification of the data. **b**, The periods and angles of the cosine waves in **a** reflect the

scale and orientation of the grid module. Taking the origin of the vectors in **a** to be alike, we see that the vectors span a parallelogram with approximately equal side lengths (0.67m and 0.72m) and an angle of 60 degrees, suggesting a rhomboidal tile representing the toroidal structure (top left). When repeated across the environment, the tile depicts the hexagonal grid pattern of the grid-cell module, confirming that the product of the two decoded circles defines a hexagonal (“twisted”) torus. As the orientation of the circular coordinates is arbitrary, the directions of the axes may be any of the following: reversely oriented (blue arrows), a different 60-degree pair of axes (green), or have a relative angle of 120 degrees (yellow). **c**, Rhombi of each module for each OF session (n cells as in Extended Data Fig. 2g), given by the cosine wave fitted to the toroidal coordinates (as in **b**). The toroidal parametrizations were obtained independently in different behavioural conditions (colour-coded), then used to decode the module’s activity during OF foraging, and subsequently mapped as a function of the rat’s position in the environment (see **f**). Positions of downsampled spikes from example cells of each module are shown in greyscale to illustrate grid scale and orientation. The consistent angle and side lengths suggest the geometry of the rhombus is retained across brain states and environments, with a constant scale relationship between modules. **d**, Mean value of a single neuron in rhomboidal coordinates displays a single bump (as in Fig. 2a), which, when repeated and arranged to tessellate a 2D surface, reveals a grid-like pattern in the activity of the grid cell, akin to its spatial firing. **e**, Table of side lengths and angles of the cosine waves that form the rhombi in **c**, shown for each grid module and each condition (n cells as in Extended Data Fig. 2g). **f**, Visualization of the cohomological decoding of toroidal coordinates as a function of physical space (one visualization for each grid module during each condition, with the toroidal parametrizations aligned to the same axes before creating the rate maps; n cells as in Extended Data Fig. 2g). All barcodes which indicated toroidal structure exhibited periodic stripes in the OF, with phase and orientation corresponding to the two-dimensional periodicity of the grid pattern of the respective module. SWS* refers to the decoding when considering only “bursty” (B) cells of R1 as given by the correlation clustering method described in Fig 4b.

[Source data](#)

Extended Data Fig. 6 Barcodes and toroidal tuning statistics for grid modules or recording sessions not included in Figs. 2–4.

Data are shown for six grid-cell modules: R1, R3, Q1, Q2, S1 and R2 (n cells as in Extended Data Fig. 2g). Toroidal structure is clearly present across environments and behavioural states. **Aa–Ad**, Barcode diagrams (as in Fig 1e,f) showing the results of the persistent cohomology analysis on open-field (OF), wagon-wheel track (WW) or sleep (REM or SWS) data. **Ba–Bc**, Preservation of toroidal field centres between conditions: OF vs WW (1), OF vs REM (2) OF vs SWS (3). Top row in each panel: Distribution of grid cells' receptive field centres on the inferred torus for OF and WW as well as sleep states, similar to Fig 2e. Each dot signifies the field centre of an individual grid cell. Grey lines connect field centres of the same cell across conditions. Note the proximity of red-black pairs (after separate alignment for the two recording sessions of each panel). Middle and bottom rows: Cumulative distributions showing stability of grid cells' toroidal tuning between brain states, as in Fig. 2f,g. Distributions show peak field distance (middle) and Pearson correlation of pairs of toroidal rate maps (bottom). Labelling as in Fig. 2e–g. **C**, Top: Histograms of the information content carried by individual cells' activity about position on the inferred torus during REM (left) and SWS (right). Counts (fractions of the cell sample) are shown as a function of information content (in bins of 0.28 bits/spike) for all grid modules (colour-coded). The vertical dashed line (close to zero) shows mean information content for shuffled distributions ($n = 1,000$ shuffles). The majority of cells have a higher information content. Bottom: Explained deviance of a GLM model fitted to the spike count with toroidal coordinates during REM (left) and SWS (right) as regressor. Distributions show counts (fractions of the cell sample) as a function of explained deviance, in bins of 0.035, for all grid modules. Values larger than 0 indicate that the fitted model explains the data better than a null model that assumes a constant firing rate.

[Source data](#)

Extended Data Fig. 7 Barcodes and decoding of simulated firing activity for two grid-cell CAN models (with no noise),

and for two point clouds randomly sampled on a hexagonal and a square torus.

a, Persistent cohomology analysis of a simulated grid-cell network based on the CAN model from Couey et al (2013)¹¹ during OF foraging. Left: Colour-coded firing rates for a single time frame of the 56×44 grid cells, shown at their respective positions on the neural sheet. Middle: Barcode of the simulated data. Arrows point to one 0D, two 1D and one 2D bar with long lifetimes, indicating toroidal structure. Right: Each coordinate of the toroidal parametrization of the two longest lived 1D features is mapped onto the spatial trajectory, colour-coded by its cosine value (as in Extended Data Fig. 5a, f). The resulting striped patterns of the two maps are oriented approximately 60 degrees relative to each other, as expected from a hexagonal torus network structure (see d). **b**, Analysis of a random sample of 100 grid cells (of a total of 400 cells) of a simulated grid cell network, using the twisted torus CAN model formulated by Guanella et al (2007)¹⁰. Left: Firing rates of the cells in the network at a single time frame. The model generates a single bump of activity based on both inhibitory and excitatory, asymmetric connections representing a twisted torus. Barcode (middle) and cohomological decoding of toroidal position (right) are shown as in a. The barcode shows four prominent bars: one 0D bar, two 1D bars and one 2D bar, similar to that of a torus. Note that the pair of stripes in toroidal coordinates are oriented 60 degrees relative to each other. **c, d**, To verify the expected barcodes and decoding of a torus and compare with both real and synthetic grid cell data, we performed the same topological analysis on point clouds sampled from two idealized toroidal parametrizations ($n = 2,500$ points): a 4D description of a square torus (c) and a 6D embedding of a hexagonal torus (d). Left: Representing the firing of a cell as a Gaussian function centred at a single toroidal coordinate on the toroidal sheet results in a square (c) and hexagonal (d) firing pattern, when arranged to tessellate a 2D surface. Middle: The expected barcode of a torus (one 0D, two 1D, and one 2D bar clearly longer than the other bars) is seen in both cases. Right: each sampled angle is coloured according to the decoded toroidal coordinates. Note the difference in the relative angle of the pair of stripes between the square and the hexagonal torus.

[Source data](#)

Extended Data Fig. 8 Subpopulations of grid cells with different temporal spiking statistics have different degrees of toroidal selectivity.

a, Geometry of grid-cell pattern of all six modules with classes of grid cells (B, bursty; T, theta-modulated; N, non-bursty; as defined in Fig. 4). Each plot shows the locations of the innermost six peaks of the spatial autocorrelogram for every grid cell in one module. Each dot indicates the position of one peak from one cell (total of 6 dots per cell); dots are coloured by the cell's class. The grey crosshair indicates the centre of the autocorrelogram. **b**, Correlation matrix showing pairwise correlation of firing rates for all grid cells belonging to S1 (left; $n = 73$ cells) and R1 (right – same data as for autocorrelogram distance matrix in Fig. 4b; $n = 111$ cells). Correlation is colour-coded according to the scale bar, with minimum and maximum defined as the 1st and 99th percentile, respectively, of the pairwise correlation distribution for each module. Rows and column (cells) are ordered according to class, as assigned by the clustering analysis shown in Fig. 4. Each cluster displays strong inner correlation structure for both modules during SWS. Cluster boundaries are indicated on the x axis of the correlation matrix. **c**, Summary of pairwise correlations of SWS activity for grid cells in modules R1 and S1, shown according to cell class. In each matrix plot, rows and columns indicate cell classes, and each element represents all pairs of grid cells from the classes corresponding to the row and column. Matrix elements are colour-coded to represent (top) the median of the spike train Pearson correlation r value across all cell pairs, (middle) Spearman rank correlation between cell pairs' grid (toroidal) phase offsets and their spike train Pearson correlation r values, (bottom) same as middle, but for head-direction phase instead of grid phase. Number of cell pairs were as follows: module R1, B-B 2346, B-T 6348, B-N 1932, T-T 4186, T-N 2576, N-N 378; module S1 B-B 378, B-T 1680, B-N 1456, T-T 1770, T-N 3120, N-N 1326. Note that, in agreement with the topological analyses, the correlation between cell pairs' grid phases and their spike-time correlations are weaker for theta-modulated cells than non-bursty and particularly bursty cells. This drop is explained by an increase in the correlation with head direction, suggesting, as expected in conjunctive cells, that head direction accounts for much of the variation in these cells, unlike the other classes. Furthermore, the median spike

correlation for pairs of theta-modulated and non-bursty cells is higher than for bursty cells, indicating a stronger positive correlation bias, consistent with more global fluctuations of activity in these populations. **d**, Cumulative distributions showing distance between toroidal field centres (upper) and Pearson correlation r values (lower) for toroidal rate maps of grid cells in each class as in Fig. 2f,g, but here comparing awake behaviour in OF with SWS, n cells = 523(B), 229(T) and 95(N) cells for OF and 495(B), 169(T), 43(N) cells for REM and SWS. $n = 1,000$ shuffles. **e**, Cumulative distributions showing toroidal explained deviance (left) and information content (right) for all grid cells in each class – bursty (B), theta-modulated (T) and non-bursty (N) – and for each of three conditions – OF, REM and SWS. Cells are from all modules. n cells as in **d**. **f**, Barcode of T-class grid cells from modules R1 (left; $n = 92$ cells) and S1 (right; $n = 60$ cells) during SWS reveals a single prominent long-lived H^1 bar (indicated by black arrow). **g**, Cohomological decoding of the longest-lived H^1 bar in each barcode in **f** reveals strong correlation with recorded head direction. Recorded head direction (black) and decoded direction (blue) are shown as a function of time (total snippet length 10 s).

[Source data](#)

[Extended Data Fig. 9 Classification of sleep and wake states based on behavioural and neural activity during rest sessions.](#)

A, Example traces of MEC multi-unit activity (upper; coloured lines), and rasters of spike times of 444 grid cells (lower; black dots) recorded from rat 'R' during OF foraging, REM sleep and slow-wave sleep (SWS). Cells are ranked from top to bottom by the number of spikes fired during the example time window. Note the presence of regular theta waves (5–10 Hz) during OF and REM, and presence of slower, more irregular fluctuations between active "up-states" and silent "down-states" during SWS. Middle: times of population activity vectors (calculated in 10 ms time bins) which were selected for persistent cohomology analysis, for each module (R1-R3). Each dot indicates a vector which was included in the initial downsampled set of 15,000 vectors with the highest mean firing rate across cells in the module. Vertical ticks indicate the subset of these vectors which were retained after using a density-based method to reduce the data to a representative point

cloud. Note that during SWS, all of the selected population activity vectors occurred during up-states. **B**, Classification of sleep/wake states based on behavioural and neural activity during rest sessions. Each of the three horizontal blocks shows a recording from one animal. Rat 'R' day 1 did not contain a rest session and is not shown on this figure. **Ba**, Detection of REM and SWS sleep epochs in the rest session. The plots show the time courses of the three variables used for detecting REM and SWS epochs. Top panel of each block: animal locomotion speed; middle panel: the animal's head angular speed; bottom panel: the ratio of the amplitude of theta (5–10 Hz) and delta (1–4 Hz) frequency bands in the multi-unit spiking activity (theta/delta ratio, TDR). **Bb**, Log-power spectra of MEC multi-unit activity during each sleep/wake state. The line and shaded area indicate the mean and 95% bootstrap confidence intervals, calculated across time windows (confidence intervals are narrow). Note the pronounced peak corresponding to the theta band (5–10 Hz) during OF and REM, and the higher power in the delta band (1–4 Hz) during SWS. **Bc**, Histograms showing distributions of firing rates for all grid cells during each sleep/wake state (number of grid cells: rat 'Q' 159, rat 'R' 428, rat 'S' 72). **C**, Table showing total time and median bout length of recorded sleep for each animal.

[Source data](#)

[Extended Data Fig. 10 Tuning to coordinates in space and on the inferred torus for all grid cells of module R2 \(separated into pure and conjunctive categories\) on recording day 2.](#)

Plots show all 152 cells in module R2, a subset of which is shown in Fig. [3b](#). Plots from left to right: OF firing rate map, head-direction tuning curve (black) compared to occupancy of head directions (light grey), temporal autocorrelogram, toroidal firing rate maps for OF, REM and SWS. The full set of plots, for all remaining grid cells of all recordings, is shown in [Supplementary Information](#).

[Source data](#)

Supplementary information

Supplementary Information

This file contains Supplementary Data; Supplementary Methods and Supplementary References for references cited only in Supplementary Methods.

Reporting Summary

Supplementary Video 1

3D UMAP visualization of toroidal manifold (related to Fig. 1). The 3D point cloud shows a UMAP embedding of the activity of 149 grid cells from module R2 in the open-field arena, as shown in Fig. 1b, c, d. Each dot represents the population activity state at one point in time. Dots are coloured by the value of the first principal component of the population activity.

Peer Review File

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

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

Gardner, R.J., Hermansen, E., Pachitariu, M. *et al.* Toroidal topology of population activity in grid cells. *Nature* **602**, 123–128 (2022).
<https://doi.org/10.1038/s41586-021-04268-7>

- Received: 24 February 2021
- Accepted: 19 November 2021
- Published: 12 January 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04268-7>

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/s41586-021-04268-7>

- Article
- [Published: 26 January 2022](#)

Brahma safeguards canalization of cardiac mesoderm differentiation

- [Swetansu K. Hota](#) ORCID: orcid.org/0000-0003-3105-9729^{1,2,3},
- [Kavitha S. Rao](#)^{1,2},
- [Andrew P. Blair](#)¹,
- [Ali Khalilimaybodi](#) ORCID: orcid.org/0000-0001-9318-8433⁴,
- [Kevin M. Hu](#) ORCID: orcid.org/0000-0001-5513-7854^{1,2},
- [Reuben Thomas](#)¹,
- [Kevin So](#) ORCID: orcid.org/0000-0002-5193-7384^{1,2},
- [Vasumathi Kameswaran](#) ORCID: orcid.org/0000-0003-2552-6183^{1,2},
- [Jiewei Xu](#)^{1,5,6},
- [Benjamin J. Polacco](#) ORCID: orcid.org/0000-0003-1570-9234^{1,5,6},
- [Ravi V. Desai](#) ORCID: orcid.org/0000-0002-0336-0883¹,
- [Nilanjana Chatterjee](#) ORCID: orcid.org/0000-0002-1203-5222⁷,
- [Austin Hsu](#)^{1,2},
- [Jonathon M. Muncie](#) ORCID: orcid.org/0000-0002-7405-6213¹,
- [Aaron M. Blotnick](#) ORCID: orcid.org/0000-0001-8646-6853^{1,2},
- [Sarah A. B. Winchester](#)^{1,2},
- [Leor S. Weinberger](#)^{1,8,9},
- [Ruth Hüttenhain](#)^{1,5,6},
- [Irfan S. Kathiriya](#) ORCID: orcid.org/0000-0003-1207-1490^{1,2,10},
- [Nevan J. Krogan](#)^{1,5,6},
- [Jeffrey J. Saucerman](#)⁴ &
- [Benoit G. Bruneau](#) ORCID: orcid.org/0000-0002-0804-7597^{1,2,3,11}

Nature volume 602, pages 129–134 (2022)

- 5144 Accesses
- 432 Altmetric
- [Metrics details](#)

Subjects

- [Chromatin remodelling](#)
- [Differentiation](#)
- [Embryonic stem cells](#)
- [Epigenomics](#)

Abstract

Differentiation proceeds along a continuum of increasingly fate-restricted intermediates, referred to as canalization^{1,2}. Canalization is essential for stabilizing cell fate, but the mechanisms that underlie robust canalization are unclear. Here we show that the BRG1/BRM-associated factor (BAF) chromatin-remodelling complex ATPase gene *Brm* safeguards cell identity during directed cardiogenesis of mouse embryonic stem cells. Despite the establishment of a well-differentiated precardiac mesoderm, *Brm*^{-/-} cells predominantly became neural precursors, violating germ layer assignment. Trajectory inference showed a sudden acquisition of a non-mesodermal identity in *Brm*^{-/-} cells. Mechanistically, the loss of *Brm* prevented de novo accessibility of primed cardiac enhancers while increasing the expression of neurogenic factor POU3F1, preventing the binding of the neural suppressor REST and shifting the composition of BRG1 complexes. The identity switch caused by the *Brm* mutation was overcome by increasing BMP4 levels during mesoderm induction. Mathematical modelling supports these observations and demonstrates that *Brm* deletion affects cell fate trajectory by modifying saddle–node bifurcations². In the mouse embryo, *Brm* deletion exacerbated mesoderm-deleted *Brg1*-mutant phenotypes, severely compromising cardiogenesis, and reveals an *in vivo* role for *Brm*. Our results show that *Brm* is a compensable safeguard of the fidelity of

mesoderm chromatin states, and support a model in which developmental canalization is not a rigid irreversible path, but a highly plastic trajectory.

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: BRM activates cardiac gene expression programs and prevents acquisition of neural fate during directed cardiomyocyte differentiation.

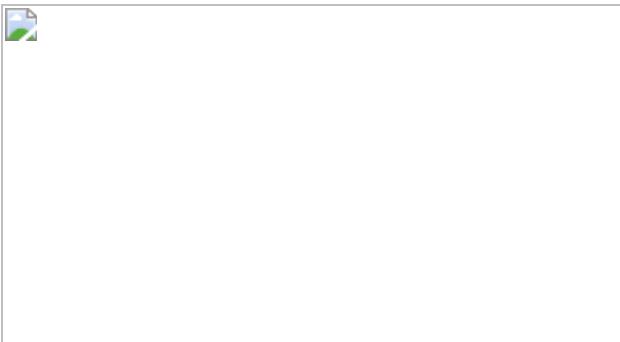


Fig. 2: BRM modulates regulatory chromatin accessibility near cardiac and neural genes and modulates POU3F1 and REST to prevent neurogenesis during cardiac differentiation.

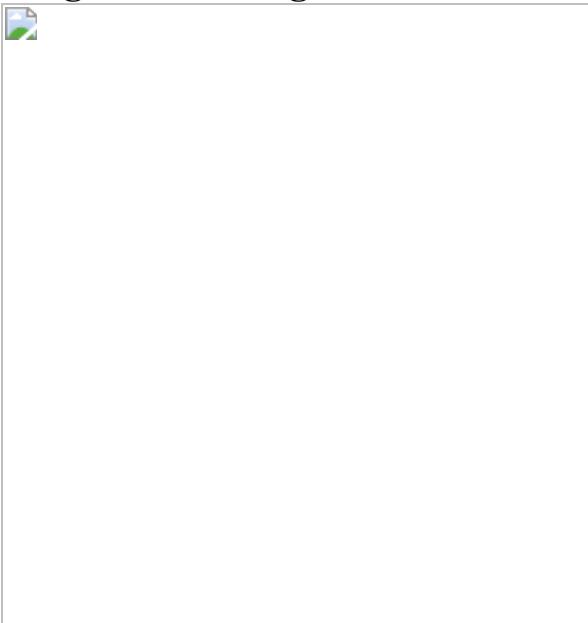
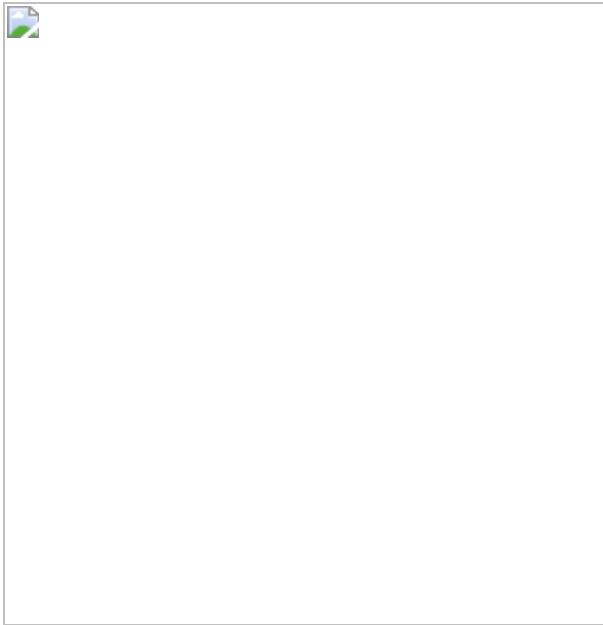


Fig. 3: Loss of *Brm* is compensable in vitro and in vivo.



Data availability

Bulk and single-cell RNA-seq, ATAC-seq and ChIP-seq datasets have been deposited in Gene Expression Omnibus under accession number [GSE150186](#) and mass spectrometry proteomics data to the ProteomeXchange Consortium with the dataset identifier [PXD026638](#).

Code availability

Open source GitHub repository codes are provided for single-cell data analysis (<https://github.com/swhota/Brm-scripts>), mathematical modelling of Brahma with definitions of all model state variables, parameters, parameter values and interactions (https://github.com/mkm1712/Brahma_model), logic-based differential equations generation using Netflux (<https://github.com/saucermanlab/Netflux>), peaks from ChIP-seq and ATAC-seq datasets (https://github.com/gladstone-institutes/Hota_et_al_2021_Brm_safeguards_canalization_cardiac_diff) and the pipeline for automated ChIP-seq and ATAC-seq data processing (<https://github.com/gladstone-institutes/MonkeyPipeline>).

References

1. 1.

Waddington, C. H. *The Strategy of the Genes, a Discussion of Some Aspects of Theoretical Biology* 20 (George Allen & Unwin Ltd, 1957).

2. 2.

Ferrell, J. E., Jr Bistability, bifurcations, and Waddington's epigenetic landscape. *Curr. Biol.* **22**, R458–R466 (2012).

3. 3.

Reyes, J. C. et al. Altered control of cellular proliferation in the absence of mammalian Brahma (SNF2 α). *EMBO J.* **17**, 6979–6991 (1998).

4. 4.

Van Houdt, J. K. J. et al. Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nat. Genet.* **44**, 445–449 (2012).

5. 5.

Tsurusaki, Y. et al. Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat. Genet.* **44**, 376–378 (2012).

6. 6.

Kadoch, C. Diverse compositions and functions of chromatin remodeling machines in cancer. *Sci. Transl. Med.* **11**, eaay1018 (2019).

7. 7.

Hoffman, G. R. et al. Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc. Natl Acad. Sci. USA* **111**, 3128–3133 (2014).

8. 8.

Smith-Roe, S. L. & Bultman, S. J. Combined gene dosage requirement for SWI/SNF catalytic subunits during early mammalian development. *Mamm. Genome* **24**, 21–29 (2013).

9. 9.

Bultman, S. J. et al. BRG1 and BRM SWI/SNF ATPases redundantly maintain cardiomyocyte homeostasis by regulating cardiomyocyte mitophagy and mitochondrial dynamics in vivo. *Cardiovasc. Pathol.* **25**, 258–269 (2016).

10. 10.

Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59–59 (2019).

11. 11.

Farrell, J. A. et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).

12. 12.

Gouti, M. et al. A gene regulatory network balances neural and mesoderm specification during vertebrate trunk development. *Dev. Cell* **41**, 243–261 (2017).

13. 13.

Thomson, M. et al. Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145**, 875–889 (2011).

14. 14.

Hota, S. K. et al. Dynamic BAF chromatin remodeling complex subunit inclusion promotes temporally distinct gene expression programs in cardiogenesis. *Development* **146**, dev174086 (2019).

15. 15.

Takeuchi, J. K. et al. Chromatin remodelling complex dosage modulates transcription factor function in heart development. *Nat. Commun.* **2**, 187 (2011).

16. 16.

Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

17. 17.

Weber, C. M. et al. mSWI/SNF promotes polycomb repression both directly and through genome-wide redistribution. *Nat. Struct. Mol. Biol.* **28**, 501–511 (2021).

18. 18.

Zhu, Q. et al. The transcription factor Pou3f1 promotes neural fate commitment via activation of neural lineage genes and inhibition of external signaling pathways. *eLife* **3**, e02224 (2014).

19. 19.

Battaglioli, E. et al. REST repression of neuronal genes requires components of the hSWI.SNF complex. *J. Biol. Chem.* **277**, 41038–41045 (2002).

20. 20.

Kattman, S. J. et al. Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell* **8**, 228–240 (2011).

21. 21.

Paulsen, M., Legewie, S., Eils, R., Karaulanov, E. & Niehrs, C. Negative feedback in the bone morphogenetic protein 4 (BMP4) synexpression group governs its dynamic signaling range and canalizes development. *Proc. Natl Acad. Sci. USA* **108**, 10202–10207 (2011).

22. 22.

Arias, A. M. & Hayward, P. Filtering transcriptional noise during development: concepts and mechanisms. *Nat. Rev. Genet.* **7**, 34–44 (2006).

23. 23.

Bier, E. & De Robertis, E. M. BMP gradients: a paradigm for morphogen-mediated developmental patterning. *Science* **348**, aaa5838 (2015).

24. 24.

Kraeutler, M. J., Soltis, A. R. & Saucerman, J. J. Modeling cardiac β -adrenergic signaling with normalized-Hill differential equations: comparison with a biochemical model. *BMC Syst. Biol.* **4**, 157–12 (2010).

25. 25.

Lessard, J. et al. An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55**, 201–215 (2007).

26. 26.

Lamba, D. A., Hayes, S., Karl, M. O. & Reh, T. Baf60c is a component of the neural progenitor-specific BAF complex in developing retina. *Dev. Dyn.* **237**, 3016–3023 (2008).

27. 27.

Zuryn, S. et al. Sequential histone-modifying activities determine the robustness of transdifferentiation. *Science* **345**, 826–829 (2014).

28. 28.

Molina-García, L. et al. Direct glia-to-neuron transdifferentiation gives rise to a pair of male-specific neurons that ensure nimble male mating. *eLife* **9**, e48361 (2020).

29. 29.

Jiang, Z. et al. Knockdown of *Brm* and *Baf170*, components of chromatin remodeling complex, facilitates reprogramming of somatic cells. *Stem Cells Dev.* **24**, 2328–2336 (2015).

30. 30.

Treutlein, B. et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature* **534**, 391–395 (2016).

31. 31.

Wamstad, J. A. et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* **151**, 206–220 (2012).

32. 32.

Alexander, J. M. et al. Brg1 modulates enhancer activation in mesoderm lineage commitment. *Development* **142**, 1418–1430 (2015).

33. 33.

Ho, L. et al. An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc. Natl Acad. Sci. USA* **106**, 5181–5186 (2009).

34. 34.

Conti, L. et al. Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.* **3**, e283 (2005).

35. 35.

Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).

36. 36.

Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).

37. 37.

Abmayr, S. M., Yao, T., Parmely, T. & Workman, J. L. Preparation of nuclear and cytoplasmic extracts from mammalian cells. *Curr. Protoc. Pharmacol.* **75**, 12.1.1–12.1.10 (2006).

38. 38.

Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

39. 39.

Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

40. 40.

Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

41. 41.

Zambon, A. C. et al. GO-Elite: a flexible solution for pathway and ontology over-representation. *Bioinformatics* **28**, 2209–2210 (2012).

42. 42.

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

43. 43.

McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2018).

44. 44.

McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337 (2019).

45. 45.

Lambiotte, R., Delvenne, J. C. & Barahona, M. Laplacian dynamics and multiscale modular structure in networks. Preprint at <https://arxiv.org/abs/0812.1770> (2008).

46. 46.

Teschendorff, A. E. & Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nat. Commun.* **8**, 15599 (2017).

47. 47.

Eling, N., Richard, A. C., Richardson, S., Marioni, J. C. & Vallejos, C. A. Correcting the mean-variance dependency for differential variability testing using single-cell RNA sequencing data. *Cell Syst.* **7**, 284–294.e12 (2018).

48. 48.

Wang, J., Zhang, K., Xu, L. & Wang, E. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc. Natl Acad. Sci. USA* **108**, 8257–8262 (2011).

49. 49.

Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565 (1942).

50. 50.

Bhattacharya, S., Zhang, Q. & Andersen, M. E. A deterministic map of Waddington’s epigenetic landscape for cell fate specification. *BMC Syst. Biol.* **5**, 85–12 (2011).

51. 51.

O’Geen, H., Echipare, L. & Farnham, P. J. Using ChIP-seq technology to generate high-resolution profiles of histone modifications. *Methods Mol. Biol.* **791**, 265–286 (2011).

52. 52.

Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

53. 53.

Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137–R139 (2008).

54. 54.

Xing, H., Mo, Y., Liao, W. & Zhang, M. Q. Genome-wide localization of protein-DNA binding and histone modification by a bayesian change-point method with ChIP-seq data. *PLoS Comput. Biol.* **8**, e1002613 (2012).

55. 55.

Neph, S. et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).

56. 56.

McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).

57. 57.

Afgan, E. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, W537–W544 (2018).

58. 58.

McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

59. 59.

Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

60. 60.

Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide

protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

61. 61.

Choi, M. et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–2526 (2014).

62. 62.

Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

63. 63.

Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S: Statistics and Computing* 4th edn (2002).

64. 64.

Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2018).

65. 65.

Rhee, H. S. et al. Expression of terminal effector genes in mammalian neurons is maintained by a dynamic relay of transient enhancers. *Neuron* **92**, 1252–1265 (2016).

Acknowledgements

We thank N. Carli, Y. Hao, M. Bernardi and J. McGuire (Gladstone Genomics Core) for RNA-seq and 10x Genomics library preparation; staff at the UCSF Center for Applied Technologies for sequencing; E. Nora for help with *Brm-AID* strain construction; J. Zhang (Gladstone Transgenic Core) for mouse knockout generation; staff at the Gladstone Stem Cell Core

for cell culture; R. Wang for ChIP-seq; K. Choudhary for ATAC-seq analysis; K. Claiborn for editorial assistance; and G. Maki for graphics. This work was supported by grants from the NIH/NHLBI (P01HL089707 and P01HL146366 to B.G.B. and N.J.K.; Bench to Bassinet Program UM1HL098179; and R01HL114948 to B.G.B. and R01HL137755 to J.J.S.); and postdoctoral fellowships from the American Heart Association (13POST17290043), Tobacco Related Disease Research Program (22FT-0079), NIH training grant (2T32-HL007731-26) and career development award (861914) from the American Heart Association to S.K.H. I.S.K. was supported by funds from the Society for Pediatric Anesthesia, Hellman Family Fund, UCSF REAC Award and the UCSF Department of Anesthesia and Perioperative Care. This work was also supported by an NIH/NCRR grant (C06 RR018928) to the J. David Gladstone Institutes, The Roddenberry Foundation and The Younger Family Fund (to B.G.B.).

Author information

Affiliations

1. Gladstone Institutes, San Francisco, CA, USA

Swetansu K. Hota, Kavitha S. Rao, Andrew P. Blair, Kevin M. Hu, Reuben Thomas, Kevin So, Vasumathi Kameswaran, Jiewei Xu, Benjamin J. Polacco, Ravi V. Desai, Austin Hsu, Jonathon M. Muncie, Aaron M. Blotnick, Sarah A. B. Winchester, Leor S. Weinberger, Ruth Hüttenhain, Irfan S. Kathiriya, Nevan J. Krogan & Benoit G. Bruneau

2. Roddenberry Center for Stem Cell Biology and Medicine at Gladstone, San Francisco, CA, USA

Swetansu K. Hota, Kavitha S. Rao, Kevin M. Hu, Kevin So, Vasumathi Kameswaran, Austin Hsu, Aaron M. Blotnick, Sarah A. B. Winchester, Irfan S. Kathiriya & Benoit G. Bruneau

3. Cardiovascular Research Institute, University of California, San Francisco, CA, USA

Swetansu K. Hota & Benoit G. Bruneau

4. Department of Biomedical Engineering, University of Virginia,
Charlottesville, VA, USA

Ali Khalilimeybodi & Jeffrey J. Saucerman

5. Department of Cellular and Molecular Pharmacology, University of
California, San Francisco, CA, USA

Jiewei Xu, Benjamin J. Polacco, Ruth Hüttenhain & Nevan J. Krogan

6. Quantitative Biosciences Institute, University of California, San
Francisco, CA, USA

Jiewei Xu, Benjamin J. Polacco, Ruth Hüttenhain & Nevan J. Krogan

7. Department of Medicine, University of California, San Francisco, CA,
USA

Nilanjana Chatterjee

8. Department of Pharmaceutical Chemistry, University of California,
San Francisco, CA, USA

Leor S. Weinberger

9. Department of Biochemistry and Biophysics, University of California,
San Francisco, CA, USA

Leor S. Weinberger

10. Department of Anesthesia and Perioperative Care, University of
California, San Francisco, CA, USA

Irfan S. Kathiriya

11. Department of Pediatrics, University of California, San Francisco, San
Francisco, CA, USA

Benoit G. Bruneau

Contributions

Project design and direction: B.G.B. and S.K.H. ES cell engineering, in vitro differentiation, protein purification, gene expression, scRNA-seq, ATAC-seq, ChIP-seq and data analysis: S.K.H. Additional scRNA-seq analysis: A.P.B. and K.S.R. under the direction of B.G.B. and I.S.K. Mathematical modelling: A.K. under the direction of J.J.S. Genotyping and echocardiography for pregnant female identification: S.A.B.W. Mouse embryo dissection: I.S.K. and S.K.H. Embryo staining: K.M.H. Embryo imaging: S.K.H. and J.M.M. Mass spectrometry: J.X. and B.J.P. under the direction of R.H. and N.J.K. Additional ChIP-seq analysis: V.K. Additional ATAC-seq and mass spectrometry data analysis: R.T. Cell culture: A.M.B. and K.S. Single-cell entropy analysis: R.V.D. under the direction of L.S.W. Western blot analysis: N.C., A.H. and K.S. Manuscript writing: S.K.H. and B.G.B. with contributions from all of the authors.

Corresponding authors

Correspondence to [Swetansu K. Hota](#) or [Benoit G. Bruneau](#).

Ethics declarations

Competing interests

B.G.B. is a co-founder and shareholder of Tenaya Therapeutics and consults for and has equity in Silvercreek Pharmaceuticals. The work presented here is not related to the interests of Tenaya Therapeutics or Silvercreek Pharmaceuticals. The Krogan laboratory has received research support from Vir Biotechnology and F. Hoffmann-La Roche. N.K. has consulting agreements with the Icahn School of Medicine at Mount Sinai, New York, Maze Therapeutics and Interline Therapeutics, is a shareholder of Tenaya Therapeutics and has received stocks from Maze Therapeutics and Interline Therapeutics. The other authors declare no competing interests.

Peer review information

Nature thanks Gerald Crabtree, Brian Hendrich 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 Loss of BRM leads to expression of neural genes in cardiac differentiation and has minimal effect in neural differentiation.](#)

a, *Brm* mRNA expression during cardiac differentiation from Wamstad et al.³¹. **b**, Violin plots of *Brm* expression of single-cell data from this study. **c**, Western blot of WT and BRM KO cells at D10 of cardiac differentiation. **d**, Bulk RNA-seq analysis of WT and BRM KO cells at D4, D5.3 and D10 stages of differentiation. Counts per million (CPM) average of three biological replicates were plotted as a ratio of KO over WT. Gene Ontology (GO) biological process enrichment was determined by GOElite. **e**, Dots plots showing expression of indicated genes from D10 WT and *Brm*^{-/-} scRNA-seq data. **f**, Scheme of neural precursor differentiation from ES cells and TUBB3 immunostaining of WT and *Brm*^{-/-} cells differentiated to neural precursor (D13) cells. Scale bars are 200 μ m.

[Extended Data Fig. 2 BRM prevents acquisition of neural fate after pre-cardiac mesoderm formation.](#)

a–d, scRNA-seq data of cardiac differentiation projected on UMAP space showing gene expression feature plots (**a**), dot plots of quantitative bulk

changes in gene expression between WT and *Brm*^{-/-} cells for early developmental, cardiac mesoderm, cardiac precursors, cardiomyocytes, genes enriched in *Brm*^{-/-} cells, and a select set of genes involved in neuroectoderm development (**b**), PAGA connectivity lines for WT (**c**) or *Brm*^{-/-} (**d**) at D4, D6 and D10 stages of differentiation. **e**, Feature plots of developmental trajectory analysis using URD for selected cardiac and neural genes. **f, g**, Pluripotency is unaffected in BRM KO cells. **f**, Immunostaining of WT and *Brm*^{-/-} ES cells with indicated pluripotency markers. Scale bars are 2 μm, magnification 63x. **g**, scRNA-seq of WT and *Brm*^{-/-} cells in ES cell cluster together. **h**, Integration of scRNA-seq data from D0 ES cells with D4, D6 and D10 scRNA-seq datasets.

Extended Data Fig. 3 Loss of BRG1 early in differentiation leads to formation of non-cardiac cell types.

a, Comparison of *Brg1* and *Brm* expression during cardiac differentiation³¹. **b**, Scheme of cardiac differentiation showing timing of induction with 4-hydroxy tamoxifen (4-OHT) or the control tetrahydrofuran (THF) and scRNA-seq. THF or 4-OHT was treated for 2 days to achieve complete *Brg1* deletion¹⁴. **c–e**, UMAPs of scRNA-seq data at D4 and D10 of differentiation of WT and conditional BRG1 KO genotypes (**c**), clusters with inferred cell types (**d**) and feature plots of expression of indicated genes (**e**). **f**, Dot plots comparing gene expression quantification of WT and conditional BRG1 KO at D4 and D10 of differentiation. **g**, Cardiac troponin T (cTnT) and TUBB3 immunostaining at D10 for WT and BRG1 cKO cells deleted at D4 of differentiation. Scale bars are 200 μm. **h**, Integration of scRNA-seq data of *Brg1* cKO and *Brm* KO at D10 stage of differentiation.

Extended Data Fig. 4 BRM is required during cardiac mesoderm formation.

a, Mean difference plots of ATAC-seq data plotting average log fold change between WT and *Brm*^{-/-} cells and average log CPM (3 biological replicates each) at D0 and D2 of differentiation. Statistically significant (FDR <0.05) peaks showing log2 fold change >1, unchanged, and <1 are shown in red, black and blue respectively. **b–c**, ATAC-seq browser tracks showing WT

and BRM KO chromatin accessibility at D4, D6 and D10 of cardiac differentiation along with H3K27ac active enhancer marks near cardiac genes (**b**) and indicated neural gene loci, along with neural precursor H3K27ac marks⁶⁵ (**c**). **d–e**, BRM-mediated open and closed chromatin regions compared with cardiac and neural progenitor enhancers. Closed and open chromatin in *Brm*^{-/-} at D6 (**d**) and at D10 (**e**) are compared with respective cardiac and neural progenitor enhancers. **f**, Motifs enriched at the open chromatin regions in WT and BRM KO cells at D4, D6, D10 differentiation stages. BRM activity is essential before D4 of differentiation. **g**, Auxin inducible degron mouse ES strain of BRM (*Brm-AID*) differentiated to cardiomyocytes at D10 and treated without (lane 1) or with auxin analog indole acetic acid (IAA) for indicated period of time shows rapid BRM degradation by western blot. **h–i**, Schematic of cardiac differentiation showing time of IAA treatment and beating at D10. Cells treated with IAA for indicated length of time (**h**) or a period of two days at a time (**i**) were analysed by immunostaining of cardiac troponin T at D10. Scale bars are 200 μm.

Extended Data Fig. 5 BRM loss leads to reduced H3K27ac marks near cardiac genes and increased H3K27ac marks near neural genes.

a, Differential enrichment of H3K27me3 marks in WT and *Brm*^{-/-} cells during cardiac differentiation displayed in the form of a heat map. **b**, Clusters b, c, and d were re-clustered and shown in a separate heat map (right). GREAT analysis of significant (Benjamini-Hochberg adjusted p-value (FDR) <0.01) GO biological processes (within 1Mb) enrichment for the clusters are on the right with representative genes shown. **c**, Heat map of significantly affected (FDR < 0.05, fold change 2) H3K27ac peaks due to loss of BRM at D4, D6 and D10 of differentiation. GREAT GO biological processes enriched (within 1mb) are shown to the right of the clusters. **d**, Number of regions significantly affected in *Brm*^{-/-} cells at D4, D6 and D10 of differentiation are plotted over WT. **e–g**, GO biological processes enriched for genes (within 1mb) near sites that gained (upper panels) or reduced (lower panels) H3K27ac marks in *Brm*^{-/-} cells at D4 (**e**), D6 (**f**) and D10 (**g**) of differentiation. **h–j**, Motifs enriched at the differentially

enriched sites in $Brm^{-/-}$ cells are shown at D4 (**h**), D6 (**i**) and D10 (**j**) stages of cardiac differentiation respectively. **k**, Western blot of indicated proteins in WT or BRM KO cells during D0, D2, D4, D6 and D10 of cardiac differentiation.

Extended Data Fig. 6 BRM regulates REST binding during cardiac differentiation.

a–d, Genome browser (IGV) tracks showing BRM-3xFLAG ChIP-seq over indicated loci (**a**) and heat maps of BRM-3xFLAG ChIP-seq over identified BRM binding sites at D4 (**b**), D6 (**c**) and D10 (**d**) of differentiation. **e–f**, GO biological processes enriched (within 100kb) (**e**) and motifs enriched (**f**) in BRM binding sites at the indicated differentiation stages. **g**, Western blot of REST expression in WT or BRM KO cells during D0, D2, D4, D6 and D10 of cardiac differentiation **h–i**, Genome browser (IGV) tracks of Brm-3x FLAG ChIP-seq near neural related genes over indicated genomic loci and REST ChIP-seq in WT and $Brm^{-/-}$ cells at D4 (**h**) and D6 (**i**) of cardiac differentiation.

Extended Data Fig. 7 BMP4 restores WT-like chromatin accessibility in $Brm^{-/-}$ cells.

a, Scheme of cardiac differentiation showing timing of IAA and BMP4 addition. Cardiac troponin T (cTnT) immunostaining of an auxin inducible degron strain of BRM (*Brm-AID*) at D10 of differentiation induced with two different BMP4 concentrations with or without IAA present throughout the differentiation. **b**, Immunostaining with cTnT shows that *Brg1* loss is not rescued by addition of increasing the amount of BMP4. Scale bars are 200 μ m. **c–e**, Heat maps showing differential enrichment of ATAC-seq peaks of WT and BRM KO cells at D4 (**c**), D6(**d**) and D10 (**e**) of cardiac differentiation with normal (1x) and high (4x) BMP4 concentrations. Boxed regions show restoration of WT-like chromatin in KO cells at high BMP4 condition. Vertical lanes show replicate data. **f–g**, Browser tracks show chromatin accessibility in WT and $Brm^{-/-}$ cells along with H3K27ac marks in cardiomyocytes and neural precursor cells (purple track) near indicated cardiac genes (**f**) and neural genes (**g**).

Extended Data Fig. 8 BMP4 restore WT-like gene expression in *Brm*^{-/-} cells and increases gene expression noise in D4 cells.

a, Dot plots showing quantitative changes in gene expression between WT and *Brm*^{-/-} cells induced with normal (1x) or high (4x) BMP4 concentrations at D4, D6 and D10 stages of differentiation for early developmental, cardiac mesoderm, precursor, and myocyte genes enriched in BRM KO cells. **b–d**, Transcriptional trajectory analysis of WT and BRM KO cells showing the genotype representation in normal BMP4 concentration (**b**), normal BMP4 for WT and 4x BMP4 concentration for BRM KO cells (**c**) and URD feature plots of expression of *Nkx2–5*, and *Actcl* (**d**). **e**, Western blots showing BMP receptor, Smad1 and phospho-SMAD expression during D0 to D4 of cardiac differentiation, **f–g**, Scatter plots of scRNA-seq data showing mean gene expression and variance from mean gene expression at D4 stage of differentiation for WT (**f**) and *Brm*^{-/-} cells (**g**) in low and high BMP4 conditions. **h–i**, Signalling entropy calculated similarly for WT (**h**) and *Brm*^{-/-} cells (**i**) with low and high BMP4 conditions.

Extended Data Fig. 9 Computational model using logic-based differential equations supports BRM's role in cardiac and neural cell fate.

a, The model interaction graph including signalling components and transcription factors critical for cardiac differentiation. **b–d**, The model outputs determine the cell fate (**b**) and temporal variations in fractional cell population during cardiac differentiation for WT (**c**) and *Brm*^{-/-} (**d**) cells. **e–h**, Model-predicted fractional activities of cardiac and neural transcription factors GATA4 (**e**), and FGF8 (**f**), as well as mediators of BRM POU3F1 (**g**) and REST (**h**) during cardiac differentiation. **i–j**, Model-predicted variations of quasi-potential landscape and subsequent path of WT (**i**) and *Brm*^{-/-} (**j**) cells induced with different levels of BMP4 from normal (3.2 ng ml⁻¹) to high (12.8 ng ml⁻¹) during cardiac differentiation. **k**, Model simulation shows that *Brm*^{-/-} cells (solid line) induced with high BMP4 at D3 (dotted line) would follow a path similar to that induced with D2

(dashed line) as computed from the GATA4 (red) and FGF8 (black) fractional activities, forming cardiomyocytes. Green line show fate variables with neural fate at 1 and cardiac fate at 0 and predicts D4 as the time of fate divergence. **I**, Phase portrait plots of bifurcation analysis of WT (upper panels) and BRM KO (lower panels) during indicated differentiation days. As differentiation progresses, WT cells undergo two sequential saddle-node bifurcations ($V \rightarrow VRV^*$ and $VRV^* \rightarrow V^*$) completing a hysteresis, while BRM KO cells undergo a saddle node bifurcation ($V \rightarrow VRV^*$) that reverses with a delay in differentiation timing ($VRV^* \rightarrow V$) with a dampened hysteresis. V = valley, R = ridge and V^* = valley different from V .

Extended Data Fig. 10 BRG1 compensates for BRM loss in vivo.

a, Anti-FLAG affinity purification of BRG1- complex followed by mass spectrometry. BRG1 (bait protein) normalized peptide intensity ratios of $Brm^{-/-}$ ($Brg1-3xFLAG; Brm^{-/-}$) over WT ($Brg1-3x FLAG$) are plotted at five different stages of differentiation (left panel) and $Brm^{-/-}$ cells at high BMP4 over normal BMP4 at MES, cardiac precursor (CP) and cardiomyocyte (CM) stages of differentiation (right panel). **b**, The exon-intron organization of *Smarca2* (encodes BRM) and the site of guide RNA that targets exon 3. The mouse strain from this transfection had a 4 bp deletion leading to premature stop codon. **c**, Western blot with anti-BRM antibody showing loss of BRM protein in $Brm^{-/-}$ mouse brain whole cell extract. α -tubulin is used as a loading control. **d**, Heterozygous Brm mouse mating resulted in pups and embryos at expected mendelian ratios. **e–f**, Western blot with antibody against BRG1 shows partial BRG1 compensation in absence of BRM in adult mouse brain (upper panel) and heart (lower panel) with quantifications shown to the right (e), but no compensation in the in vitro cardiac differentiation system (f) **g**, E 8.5 mouse embryos stained with MEF2c or cardiac troponin T (cTnT) for the indicated genotypes. Scale bars are 200 μ m.

Supplementary information

Supplementary Information

Supplementary Fig. 1: raw immunoblots of Fig. 3, Extended Data Figs. 1, 4–6, 8 and 10. Supplementary Fig. 2: the gating strategy for FACS analysis using isotype control IgG or cardiac troponin T.

Reporting Summary

Peer Review File

Supplementary Table 1

Differential gene expression between WT and homozygous *Brm*-KO cells at D4 of cardiac differentiation.

Supplementary Table 2

Marker genes enriched in the individual clusters of Fig. 1j consisting of both WT and *Brm*-KO cells at D4, D6 and D10 of cardiac differentiation.

Supplementary Table 3

Differential gene expression between WT and homozygous *Brm*-KO embryonic stem cells.

Supplementary Table 4

Differentially enriched ATAC-seq peaks between WT and homozygous *Brm*-KO cells at D4, D6 and D10 of cardiac differentiation.

Supplementary Table 5

BRM–3×Flag tag ChIP–seq peaks at D4, D6 and D10 of cardiac differentiation.

Supplementary Table 6

Marker genes enriched in the individual clusters of Fig. 3c (bottom) consisting of both WT and homozygous *Brm*-KO cells induced with normal and high BMP4 concentrations at D4, D6 and D10 of cardiac differentiation.

Supplementary Table 7

Quality control parameters for scRNA-seq data processing pipelines showing various quality control and cut-off parameters.

Supplementary Video 1

Video of WT cells showing beating CMs at D10 of differentiation.

Supplementary Video 2

Video of homozygous *Brm*-KO clone 1 cells at D10 of differentiation.

Supplementary Video 3

Video of homozygous *Brm*-KO clone 2 cells at D10 of differentiation.

Supplementary Video 4

Video of homozygous *Brm*-KO clone 3 cells at D10 of differentiation.

Supplementary Video 5

Video of heterozygous *Brm*-KO clone 1 cells at D10 of differentiation showing beating CMs.

Supplementary Video 6

Video of heterozygous *Brm*-KO clone 2 cells at D10 of differentiation showing beating CMs.

Supplementary Video 7

Quasi-Waddington diagram video of WT cells at normal BMP4 concentration from D0 to D10 of cardiac differentiation.

Supplementary Video 8

Quasi-Waddington diagram video of homozygous *Brm*-KO cells at normal BMP4 concentration from D0 to D10 of cardiac differentiation.

Supplementary Video 9

Quasi-Waddington diagram video of WT cells at high BMP4 concentration from D0 to D10 of cardiac differentiation.

Supplementary Video 10

Quasi-Waddington diagram video of homozygous *Brm*-KO cells at high BMP4 concentration from D0 to D10 of cardiac differentiation.

Supplementary Video 11

Bifurcation analysis of the model showing cell phase portraits from D0 to D10 for WT cells at normal BMP4 concentration.

Supplementary Video 12

Bifurcation analysis of the model showing cell phase portraits from D0 to D10 for homozygous *Brm*-KO cells at normal BMP4 concentration.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Hota, S.K., Rao, K.S., Blair, A.P. *et al.* Brahma safeguards canalization of cardiac mesoderm differentiation. *Nature* **602**, 129–134 (2022).
<https://doi.org/10.1038/s41586-021-04336-y>

- Received: 15 May 2020
- Accepted: 08 December 2021
- Published: 26 January 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04336-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/s41586-021-04336-y>.

- Article
- Open Access
- [Published: 05 January 2022](#)

Emergence of methicillin resistance predates the clinical use of antibiotics

- [Jesper Larsen](#) [ORCID: orcid.org/0000-0003-0582-0457](#)¹_{nal},
- [Claire L. Raisen](#)²_{nal},
- [Xiaoliang Ba](#) [ORCID: orcid.org/0000-0002-3882-3585](#)²,
- [Nicholas J. Sadgrove](#)³,
- [Guillermo F. Padilla-González](#) [ORCID: orcid.org/0000-0002-8300-6891](#)³,
- [Monique S. J. Simmonds](#)³,
- [Igor Loncaric](#)⁴,
- [Heidrun Kerschner](#)⁵,
- [Petra Apfaltrer](#)⁵,
- [Rainer Hartl](#)⁵,
- [Ariane Deplano](#)⁶,
- [Stien Vandendriessche](#)⁶_{nAff46},
- [Barbora Černá Bolfíková](#) [ORCID: orcid.org/0000-0001-8059-4889](#)⁷,
- [Pavel Hulva](#)^{8,9},
- [Maiken C. Arendrup](#)¹,
- [Rasmus K. Hare](#)¹,
- [Céline Barnadas](#)^{1,10},
- [Marc Stegger](#)¹,
- [Raphael N. Sieber](#)¹,
- [Robert L. Skov](#) [ORCID: orcid.org/0000-0002-6079-5381](#)¹¹,
- [Andreas Petersen](#)¹,
- [Øystein Angen](#)¹,
- [Sophie L. Rasmussen](#) [ORCID: orcid.org/0000-0002-2975-678X](#)^{12,13},
- [Carmen Espinosa-Gongora](#) [ORCID: orcid.org/0000-0002-9536-0548](#)¹⁴,
- [Frank M. Aarestrup](#) [ORCID: orcid.org/0000-0002-7116-2723](#)¹⁵,
- [Laura J. Lindholm](#)¹⁶,
- [Suvi M. Nykäsenoja](#)¹⁷,
- [Frederic Laurent](#)¹⁸,

- [Karsten Becker](#) [ORCID: orcid.org/0000-0002-6391-1341¹⁹](#),
- [Birgit Walther²⁰](#)^{nAff47},
- [Corinna Kehrenberg²¹](#),
- [Christiane Cuny²²](#),
- [Franziska Layer](#) [ORCID: orcid.org/0000-0002-4613-6478²²](#),
- [Guido Werner²²](#),
- [Wolfgang Witte²²](#),
- [Ivonne Stamm²³](#),
- [Paolo Moroni](#) [ORCID: orcid.org/0000-0002-0974-3084²⁴](#)^{nAff48},
- [Hannah J. Jørgensen](#) [ORCID: orcid.org/0000-0002-1788-9219²⁵](#),
- [Hermínia de Lencastre](#) [ORCID: orcid.org/0000-0001-6816-8932^{26,27}](#),
- [Emilia Cercenado](#) [ORCID: orcid.org/0000-0002-5279-3773²⁸](#),
- [Fernando García-Garrote²⁸](#)^{nAff49},
- [Stefan Börjesson](#) [ORCID: orcid.org/0000-0003-2219-2659²⁹](#)^{nAff50},
- [Sara Hæggman³⁰](#),
- [Vincent Perreten³¹](#),
- [Christopher J. Teale³²](#),
- [Andrew S. Waller](#) [ORCID: orcid.org/0000-0002-7111-9549³³](#)^{nAff51}^{nAff52},
- [Bruno Pichon³⁴](#),
- [Martin D. Curran³⁵](#),
- [Matthew J. Ellington³⁵](#)^{nAff53},
- [John J. Welch³⁶](#),
- [Sharon J. Peacock](#) [ORCID: orcid.org/0000-0002-1718-2782³⁷](#),
- [David J. Seilly²](#),
- [Fiona J. E. Morgan²](#)^{nAff54},
- [Julian Parkhill](#) [ORCID: orcid.org/0000-0002-7069-5958²](#),
- [Nazreen F. Hadjirin²](#),
- [Jodi A. Lindsay³⁸](#),
- [Matthew T. G. Holden](#) [ORCID: orcid.org/0000-0002-4958-2166³⁹](#),
- [Giles F. Edwards⁴⁰](#),
- [Geoffrey Foster](#) [ORCID: orcid.org/0000-0002-5527-758X⁴¹](#),
- [Gavin K. Paterson](#) [ORCID: orcid.org/0000-0002-1880-0095⁴²](#),
- [Xavier Didelot](#) [ORCID: orcid.org/0000-0003-1885-500X⁴³](#),
- [Mark A. Holmes](#) [ORCID: orcid.org/0000-0002-5454-1625²](#)^{na2},
- [Ewan M. Harrison](#) [ORCID: orcid.org/0000-0003-2720-0507^{37,44,45}](#)^{na2} &
- [Anders R. Larsen¹](#)^{na2}

Nature volume 602, pages 135–141 (2022)

- 43k Accesses
- 1 Citations
- 1204 Altmetric
- [Metrics details](#)

Subjects

- [Antimicrobial resistance](#)
- [Bacterial evolution](#)
- [Infectious-disease epidemiology](#)
- [Natural product synthesis](#)
- [Population dynamics](#)

Abstract

The discovery of antibiotics more than 80 years ago has led to considerable improvements in human and animal health. Although antibiotic resistance in environmental bacteria is ancient, resistance in human pathogens is thought to be a modern phenomenon that is driven by the clinical use of antibiotics¹. Here we show that particular lineages of methicillin-resistant *Staphylococcus aureus*—a notorious human pathogen—appeared in European hedgehogs in the pre-antibiotic era. Subsequently, these lineages spread within the local hedgehog populations and between hedgehogs and secondary hosts, including livestock and humans. We also demonstrate that the hedgehog dermatophyte *Trichophyton erinacei* produces two β-lactam antibiotics that provide a natural selective environment in which methicillin-resistant *S. aureus* isolates have an advantage over susceptible isolates. Together, these results suggest that methicillin resistance emerged in the pre-antibiotic era as a co-evolutionary adaptation of *S. aureus* to the colonization of dermatophyte-infected hedgehogs. The evolution of clinically relevant antibiotic-resistance genes in wild animals and the connectivity of natural, agricultural and human ecosystems demonstrate that the use of a One Health approach is critical for our understanding and management of antibiotic resistance, which is one of the biggest threats to global health, food security and development.

[Download PDF](#)

Main

Methicillin-resistant *S. aureus* (MRSA) is one of the most common antibiotic-resistant bacterial pathogens, causing approximately 171,000 invasive infections each year in Europe alone². MRSA was first identified in 1960 shortly after the introduction of methicillin (celbenin) as a treatment option against penicillin-resistant *S. aureus* clones³, but was possibly selected for by the clinical use of penicillin over the previous 20 years⁴. Methicillin resistance has subsequently emerged in many *S. aureus* clones around the world, both in hospital and community settings as well as in livestock such as pigs and cattle^{5,6}. This has serious implications for the treatment of severe infections and the World Health Organization now considers MRSA to be an important threat to human health⁷.

Methicillin resistance in *S. aureus* is mediated by the *mecA* and *mecC* genes, which encode the enzymes penicillin-binding protein 2a (PBP2a) and PBP2c, respectively. *mecA* and *mecC* confer resistance to almost all β-lactam antibiotics, including penicillinase-labile penicillins (such as penicillin G), penicillinase-stable penicillins (such as methicillin) and cephalosporins (such as cefoxitin).

Hedgehog surveys from Denmark and Sweden demonstrated a surprisingly high prevalence of MRSA carrying *mecC* (*mecC*-MRSA)^{8,9}, raising the possibility that the evolution of these bacteria was driven by natural selection in wildlife, as opposed to clinical use of antibiotics. Historically, *mecC*-MRSA was first discovered in dairy cows and subsequently in humans¹⁰, suggesting that the use of antibiotics in livestock was providing a selective advantage and that human infections were the result of zoonotic transmission. Studies from many different European countries revealed that *mecC*-MRSA is also present in other domesticated animals such as sheep, goats and horses as well as in a broad range of wild animals, albeit at low frequencies¹¹.

Our hypothesis that the evolution of *mecC*-MRSA was driven by natural selection is supported by studies from northwestern Europe and New Zealand that showed that hedgehogs are frequently colonized with the dermatophyte *T. erinacei*, which produces a penicillinase-labile penicillin-like substance that was recently identified as penicillin G^{12,13,14,15,16,17,18,19}. To test our hypothesis, we examined the distribution of *mecC*-MRSA and other *S. aureus* isolates in hedgehogs in ten European countries and New Zealand. We sequenced 244 *S. aureus* isolates from hedgehogs and 913 *S. aureus* isolates from other sources to infer the evolutionary histories, host dynamics, geographical dispersal patterns and zoonotic potential of the major *mecC*-MRSA clones in Europe. The potential mechanisms for the natural selection of *mecC*-MRSA by *T. erinacei* were assessed by analysing the genome of *T. erinacei* for β-lactam biosynthetic genes and by screening *T. erinacei* for the production of β-lactams and antibiotic activity against a panel of *S. aureus* strains.

The distribution of *mecC*-MRSA in hedgehogs

We first examined the geographical distribution and population structure of *mecC*-MRSA in European hedgehogs, which inhabit large parts of Europe as a result of postglacial expansion from Pleistocene refugia²⁰. European hedgehogs have also become widespread in New Zealand after a series of introductions from the UK between 1869 and 1892 (ref. ²¹). We analysed 828 samples from the nasal area, skin and feet of 276 hedgehogs originating from 16 wildlife rescue centres in 10 European countries and 2 wildlife rescue centres in New Zealand (Fig. 1 and Extended Data Fig. 1). *mecC*-MRSA was present in 101 of the 172 hedgehogs (222 out of 516 samples) from England and Wales (66%, 81 out of 123), Czech Republic (50%, 6 out of 12), Denmark (50%, 11 out of 22), Portugal (29%, 2 out of 7) and New Zealand (6%, 1 out of 17), therefore extending the known geographical distribution of *mecC*-MRSA in hedgehogs (Fig. 1 and Extended Data Fig. 1). By contrast, all 104 hedgehogs (312 samples) from Greece, Romania, Italy, France and Spain tested negative for *mecC*-MRSA. Whole-genome sequencing showed that the 222 *mecC*-MRSA isolates belonged to 6 clonal complexes, CC130 (75%), CC1943 (15%), CC2616 (6%), CC425 (3%), CC49 (1%) and CC599 (1%), of which CC130 had the most widespread distribution across western and central Europe (Fig. 1 and Extended Data Fig. 1). We screened all of the MRSA-negative hedgehog samples from our study ($n = 606$) for the presence of methicillin-susceptible *S. aureus* (MSSA) isolates belonging to the same clonal complexes as the *mecC*-MRSA isolates (Extended Data Fig. 1). This led to the identification of 22 MSSA isolates, including 13 CC49 isolates from Spain ($n = 9$), Denmark ($n = 3$) and Portugal ($n = 1$), and 9 CC130 isolates from England ($n = 8$) and Spain ($n = 1$).

Fig. 1: Distribution of *mecC*-MRSA clones in European and New Zealand hedgehog samples.

 **figure 1**

The analysis included 828 samples from the nasal area, skin and feet of 276 hedgehogs originating from 16 wildlife rescue centres in 10 European countries and 2 wildlife rescue centres in New Zealand. The red dots indicate the sampling locations. The pie charts are connected to the sampling locations by a red line. The area of the pie chart is proportional to the number of samples from that location. The introduction of European hedgehogs into New Zealand from the UK between 1869 and 1892 is shown. A detailed description of the results is provided in Extended Data Fig. 1. Maps

were provided by Eurostat under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence; the administrative boundaries are copyright of EuroGeographics.

Source data

The *mecC* gene encoding PBP2c is located immediately upstream of a *blaZ* gene (hereafter, *blaZ*_{LGA251}) on a chromosomally integrated mobile genetic element known as a type XI staphylococcal cassette chromosome *mec* (SCC*mec*). PBP2c and the *blaZ*_{LGA251}-encoded penicillinase are orthologues of the PBP2a enzyme and penicillinase produced by other *S. aureus* clones, although they share only 63% and 65% amino acid identities with each other, respectively¹⁰. Penicillinases have a narrower spectrum than PBP2a and PBP2c and provide resistance only to penicillin G and other penicillinase-labile subclasses of penicillin. As expected, *blaZ*_{LGA251} was present in the 222 *mecC*-MRSA isolates but absent in the 22 MSSA isolates. However, 14 of the MSSA isolates carried the *blaZ* gene found in other *S. aureus* clones (Supplementary Table 1).

Production of β-lactams by *T. erinacei*

The abundance of *mecC*-MRSA in hedgehogs led us to speculate that antibiotic production by *T. erinacei* provides a selective environment in which *mecC*-MRSA isolates have an advantage over susceptible isolates. Genome sequencing and analysis of the *T. erinacei* type strain IMI 101051 (ATCC 28443) identified orthologues of *pcbAB*, *pcbC* and *penDE*, which are responsible for key steps in penicillin G production by *Penicillium chrysogenum*, as well as the *Acremonium chrysogenum* early cephalosporin C biosynthetic genes *cefD1* and *cefD2*, which are involved in the conversion of isopenicillin N into penicillin N (Fig. 2 and Extended Data Table 1). By contrast, *T. erinacei* IMI 101051 lacked the *A. chrysogenum* late cephalosporin C biosynthetic genes *cefEF* and *cefG*. *P. chrysogenum* also carries *cefD1* and *cefD2* but is nevertheless incapable of producing cephalosporins due to the lack of *cefEF* and *cefG*^{22,23}.

Fig. 2: Penicillin biosynthetic genes and antibiotic activity of *T. erinacei* IMI 101051.

 **figure 2**

a, Schematic of the key steps in the biosynthesis of penicillin G and cephalosporin C. The presence (green) or absence (red) of *T. erinacei* penicillin G and cephalosporin C biosynthetic genes is indicated. **b**, *T. erinacei* inhibition zones against a collection of *S. aureus* control strains (black) and two *mecC*-MRSA wild-type strains belonging to CC130 (green) and CC425 (blue) and their isogenic mutants. Two-tailed paired Student's *t*-tests were used to compare inhibition zones of each mutant to the corresponding wild-type strain. Data are mean \pm s.d.; $n = 4$ biologically independent fungal culture extracts. A detailed description of the results is provided in Extended Data Fig. 4.

[Source data](#)

We processed four distinct culture broths of *T. erinacei* IMI 101051 for metabolic profiling using liquid chromatography–mass spectrometry (LC–MS) and molecular networking analysis. This led to the identification of two β -lactams, penicillin G and 6-(5-hydroxy-*n*-valeramido)-penicillanic acid (KPN), both of which belong to the penicillin class of antibiotics (Extended Data Figs. 2 and 3 and Supplementary Fig. 1).

KPN has to date been found only in culture broths of fungal strains belonging to the genus *Paecilomyces*²⁴ and differs from penicillin G by having a unique side chain (Extended Data Figs. 2 and 3 and Supplementary Fig. 1). The biosynthetic pathway of KPN is currently unknown.

Selection of *mecC*-MRSA by *T. erinacei*

Four culture broths of *T. erinacei* IMI 101051 were screened for antibiotic activity against a collection of *S. aureus* control strains. All of the culture broths produced large inhibition zones against two penicillin-susceptible *S. aureus* strains—ATCC 9144 (Oxford *S. aureus*) and ATCC 25923—but much smaller zones against the penicillinase-producing *S. aureus* strain ATCC 29213 and the *mecA*-positive *S. aureus* strain ATCC 43300 (Fig. 2 and Extended Data Fig. 4). The role of *mecC* and *blaZ*_{LGA251} was assessed by screening the culture broths for antibiotic activity against two *mecC*-MRSA wild-type strains belonging to CC130 (02.5099.D) and CC425 (LGA251) and their isogenic mutants. The mutants with deleted *mecC* ($\Delta m\text{ecC}$), *blaZ*_{LGA251} ($\Delta blaZ$ _{LGA251}), and *mecC* and *blaZ*_{LGA251} ($\Delta m\text{ecC}-blaZ$ _{LGA251}) produced significantly larger inhibition zones compared with the corresponding wild-type strains, although the zones of the $\Delta blaZ$ _{LGA251} and $\Delta m\text{ecC}-blaZ$ _{LGA251} mutants were larger compared with the zones of the $\Delta m\text{ecC}$ mutants (Fig. 2 and Extended Data Fig. 4). These results indicate that *mecC* and *blaZ*_{LGA251} both contribute to the reduced susceptibility of *mecC*-MRSA to penicillin G and KPN present in culture broths of *T. erinacei* IMI 101051.

Evolutionary history of *mecC*-MRSA

We sought to infer the evolutionary histories of *S. aureus* CC130, CC425 and CC1943, which constitute the most successful *mecC*-MRSA clones in Europe^{10,11,25}. For this purpose, we collected and sequenced 786 *mecC*-MRSA and 127 MSSA CC130, CC425 and CC1943 isolates selected to represent the known geographical distribution (mainly western and central Europe) and host repertoire (mainly humans, cattle, sheep, goats and wild animals) of each clone (Supplementary Table 1). We used core-genome single-nucleotide polymorphism (SNP) diversity and isolation dates to infer time-scaled phylogenies of these isolates and the 205 *mecC*-MRSA and 9 MSSA CC130, CC425 and CC1943 isolates collected from hedgehogs (Supplementary Table 1). The sequencing data were processed for pan-genome analysis to identify antibiotic-resistance genes (ARGs) and mobile genetic elements that encode human- and ruminant-specific immune modulators that are involved in host switching events, including a phage-encoded immune evasion cluster-1 (IEC-1) enabling *S. aureus* to evade the human innate immune response and a staphylococcal pathogenicity island

(SaPI)-encoded *vwb* gene (*vwb*_{SaPI}), which encodes a von Willebrand factor-binding protein with coagulase activity against ruminant plasma²⁶.

We also sought to infer a time-scaled phylogeny of the 991 type XI SCCmec elements containing the *mecC* and *blaZ*_{LGA251} genes but the correlation between root-to-tip distances and isolation dates was too weak with a coefficient of determination $R^2 = -0.05$ (Extended Data Fig. 5). Instead, we used the topology of the type XI SCCmec phylogeny to identify monophyletic *mecC*-MRSA lineages harbouring orthologous type XI SCCmec elements. The type XI SCCmec elements could be traced back to seven nodes that were connected to each other on a long backbone. Each of the backbone nodes and its orthologous descendants received the same letter designation to reflect their genetic relationship (A to G) (Fig. 3 and Supplementary Fig. 2). Manual mapping of the tips onto the CC130, CC425 and CC1943 phylogenies, and vice versa, enabled us to assign the *mecC*-MRSA isolates to 16 monophyletic lineages harbouring orthologous type XI SCCmec elements (Fig. 3 and Supplementary Figs. 2–5). The 129 *mecC*-MRSA CC1943 isolates could be divided into three lineages (C1 to C3), which probably originated in the early-to-late 1800s, long before the first β -lactam—penicillin G—became widely available as a therapeutic option in the 1940s (Fig. 3). The 786 *mecC*-MRSA CC130 isolates and 76 *mecC*-MRSA CC425 isolates belonged to 10 and 3 lineages (A1 to A10 and B1 to B3, respectively) (Fig. 3). Several of these lineages also originated in the pre-antibiotic era (Fig. 3). Most *mecC*-MRSA isolates lacked *vwb*_{SaPI} (96%, 949 out of 991) and IEC-1 (100%, 990 out of 991) and were genotypically susceptible to non- β -lactam antibiotics (Supplementary Figs. 3–5). The largest *mecC*-MRSA CC425 lineage (CC425:B3) had a unique evolutionary trajectory with signs of adaptation to ruminants (Supplementary Fig. 4). The basal *mecC*-MRSA CC425:B3 isolates probably originated in England during the early 1940s and shared epidemiological and genetic characteristics with the other *mecC*-MRSA lineages: they were associated with multiple hosts, including hedgehogs, cattle and humans, and lacked *vwb*_{SaPI} and IEC-1. By contrast, their descendants (CC425:B3.1) harboured *vwb*_{SaPI}, were restricted to cattle and humans in southwest England and probably diverged during the 1960s (date of the most recent common ancestor (MRCA), 1965; 95% confidence interval (CI), 1926–1986). Our analysis revealed that some of the *mecC*-MRSA lineages (CC130:A2, CC425:B3 and CC1943:C1) carried unique variants of the type XI SCCmec element, supporting that they have evolved through vertical inheritance from the MRCA of each *mecC*-MRSA lineage, whereas others shared the same type XI SCCmec variant (Fig. 3 and Supplementary Figs. 2–5). The latter findings could be the result of either purifying (negative) selection, convergent evolution, homologous recombination between different *mecC*-MRSA lineages or horizontal gene transfer.

Fig. 3: Timeline of *mecC*-MRSA CC130, CC425 and CC1943 evolution in Europe.

 **figure 3**

Manual mapping of the tips on the type XI SCC*mec* phylogeny onto the CC130, CC425 and CC1943 phylogenies, and vice versa, enabled us to assign the *mecC*-MRSA isolates to 16 monophyletic lineages containing orthologous type XI SCC*mec* elements (A–G). The trees are redrawn from Supplementary Figs. 2–5 to illustrate the branching order of the different type XI SCC*mec* variants and *mecC*-MRSA lineages. Branch lengths are not drawn to scale. The presence and absence of hedgehog isolates in a given lineage are shown as black and white boxes, respectively. A detailed description of the geographical distribution and host range of major *mecC*-MRSA CC130, CC425 and CC1943 lineages is provided in Extended Data Fig. 7. The estimated date of the most recent common ancestor and 95% confidence interval of each *mecC*-MRSA lineage are illustrated by filled circles and horizontal lines,

respectively. The introduction of penicillin G and methicillin as therapeutic options is indicated by red lines.

Source data

To better understand the potential role of horizontal gene transfer in the evolution of the three early *mecC*-MRSA CC1943 lineages, we determined the smallest number of sublineages that were present at a given time point, which is also the smallest number of acquisition events that could explain the presence of the same type XI SCC*mec* variant in all sublineages (Extended Data Fig. 6). The analysis showed that *mecC*-MRSA CC1943 consisted of 15 sublineages in 1940, just before penicillin G became available as a therapeutic option. It has previously been estimated that SCC*mec* was acquired in a single horizontal gene transfer event in three of the major hospital- and community-associated MRSA clones^{4,27,28}, although multiple introductions have also been reported²⁹. Thus, it is more plausible to assume that the type XI SCC*mec* element was present in the MRCA of each *mecC*-MRSA CC1943 lineage rather than assuming horizontal gene transfer into each of the 15 sublineages within a few years at the beginning of the antibiotic era. Vertical inheritance of the type XI SCC*mec* element is also consistent with the apparent absence of admixture between the different *mecC*-MRSA CC1943 lineages despite the fact that they are often found in hedgehogs within the same geographical area (Extended Data Fig. 7). Notably, the type XI SCC*mec* variants found in *mecC*-MRSA CC1943:C2 (E) and CC1943:C3 (F) were each other's neighbours on the type XI SCC*mec* phylogeny, and it is therefore possible that the type XI SCC*mec* element was acquired even earlier (date of the MRCA, 1737; 95% CI, 1562–1824).

MSSA isolates comprised 8% (67 out of 851), 47% (68 out of 144) and 0.8% (1 out of 130) of the CC130, CC425 and CC1943 isolates, respectively. The vast majority of the MSSA CC130 and MSSA CC425 isolates, but not the single MSSA CC1943 isolate, were basal to the corresponding *mecC*-MRSA lineages (Supplementary Figs. 3–5). Most of the basal MSSA CC130 isolates originated from sheep and goats in Italy, France, Spain and Norway and harboured *vwb*_{SaPI}. Moreover, some of the isolates carried ARGs against antibiotics that are used to treat infections in sheep and goats, including the two tetracycline-resistance genes *tet*(K) and *tet*(L), the *blaZ* gene found in other *S. aureus* clones, the chloramphenicol resistance gene *cat* and the macrolide resistance gene *erm*(C). The earliest branching CC425 lineages were epidemiologically and genetically diverse with respect to host range and the presence/absence of the type XI SCC*mec* element, *vwb*_{SaPI} and ARGs, although most originated from wild animals in Spain and lacked the aforementioned genetic determinants. Together, these findings suggest that CC130 and CC425 emerged from distinct ruminant and wildlife reservoirs in Europe and that methicillin resistance is an acquired phenotype within these clones.

Population dynamics of *mecC*-MRSA

Hedgehogs constitute a large reservoir of *mecC*-MRSA clones, as demonstrated here and elsewhere^{8,9}, whereas *mecC*-MRSA isolates are present at much lower frequencies in humans, domesticated animals and other wild animals. Hedgehog isolates were present in 9 out of the 16 *mecC*-MRSA lineages, including the 8 largest (≥ 25 isolates) and 3 earliest (200–130 years ago) lineages (Fig. 3). The 2 largest *mecC*-MRSA CC130 lineages (CC130:A9 and CC130:A10) encompassed 67% (232 out of 344), 65% (339 out of 520) and 43% (66 out of 153) of all *mecC*-MRSA isolates from hedgehogs, humans and other sources, respectively, and had the broadest geographical ranges in western and central Europe (Extended Data Fig. 7 and Supplementary Figs. 3–5). Several of the major *mecC*-MRSA CC130 and CC1943 lineages (such as CC130:A9, CC130:A10, CC1943:C1, CC1943:C2 and CC1943:C3) contained isolates that were separated by wide expanses of seawater, reflecting numerous long-distance dispersal events between British and Danish islands and mainland Europe within the past 200 years (Extended Data Fig. 7 and Supplementary Figs. 3–5). By contrast, analysis of the fine-scale population structure of *mecC*-MRSA isolates from hedgehogs revealed a detailed pattern of diversifications over short distances, resulting in substantial concordance between genetic clusters and geography at the local level (Supplementary Figs. 3–5). The observed clustering is consistent with the limited dispersal capacity of hedgehogs and the effects of habitat fragmentation.

Assuming that hedgehogs act as local reservoirs (sources) of *mecC*-MRSA in Denmark, theoretical considerations predict that secondary transmission to other hosts (sinks) would lead to similar patterns of genetic variation within these secondary hosts cohabiting the same geographical region. To examine this theory, we determined the local population structures of two broad collections of Danish *mecC*-MRSA isolates from hedgehogs and humans covering the geographical ranges of two of the three Danish hedgehog subpopulations, namely Jutland (a peninsula connected to continental Europe) and major islands³⁰. Hedgehog and human isolates from the geographical range of the remaining hedgehog subpopulation (the small island of Bornholm) were excluded from the analysis due to their small sample size (nine isolates). Most of the CC130:A10 isolates from Jutland formed a distinct sublineage (CC130:A10.1) together with isolates from other parts of Europe (Supplementary Fig. 3). As a consequence, CC130:A10.1 isolates were treated as a separate group in the analysis. The results revealed distinct patterns of regional dispersal with little overlap between Jutland and the major islands and a notable correlation between the population structures of hedgehog and human isolates at the regional level ($P = 0.0149$, two-sided Wilcoxon matched-pairs signed-rank test) (Fig. 4 and Extended Data Fig. 8). These findings support the hypothesis that human isolates originate from local hedgehog reservoirs, although it should be noted that the data presented here do

not provide evidence for directionality or rule out the involvement of other animal reservoirs (such as livestock) as part of the transmission chains.

Fig. 4: Population structures of Danish *mecC*-MRSA isolates from hedgehogs and humans.

 figure 4

a, The map of Denmark shows the geographical ranges of two of the three hedgehog subpopulations in Jutland and on the major islands. **b**, The geographical distribution of major *mecC*-MRSA CC130, CC425 and CC1943 lineages in two broad collections of *mecC*-MRSA isolates recovered from hedgehogs ($n = 141$) and humans ($n = 327$) in Jutland and on the major islands. Hedgehog and human isolates from the remaining hedgehog subpopulation on the small island of Bornholm (not shown) were excluded from the analysis due to their small sample size ($n = 9$). A detailed map of the sampling locations is provided in Extended Data Fig. 8. Maps were provided by Eurostat under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence; the administrative boundaries are copyright of EuroGeographics.

Source data

Discussion

This research shows that hedgehogs are a natural reservoir of zoonotic *mecC*-MRSA lineages that predate the antibiotic era, which is inconsistent with the commonly accepted view that widespread resistance in clinical pathogens is a modern phenomenon that is driven by our use of antibiotics in human and veterinary medicine.

Data on the prevalence of *mecC*-MRSA in humans and different animal species indicate that hedgehogs are the most likely primary host in some countries. For example, in Denmark, the prevalence in hedgehogs is considerably higher than in cattle (veal calves and bulk tank milk), sheep and goats (61% versus 0.0–1.1%)^{8,31,32}, and the number of human cases is relatively low (3–36 cases per year)²⁵. This is further supported by our finding that *mecC*-MRSA generally lacks genetic markers of human and ruminant adaptation, with the notable exception of the CC425:B3.1 lineage that has undergone a host jump from hedgehogs to cattle in southwest England. Before this study, dairy cows were considered to be the most likely reservoir of *mecC*-MRSA and a major source of zoonotic infections in humans. This hypothesis was supported by the fact that β-lactams are commonly used to treat bovine mastitis, as shown by sales data from Denmark and the UK^{33,34}. However, our findings strongly suggest that most *mecC*-MRSA lineages originate from hedgehogs, although dairy cows and other domesticated animals probably act as intermediate hosts and vectors in zoonotic transmission from hedgehogs to humans, as previously demonstrated³⁵.

Here we show that the *T. erinacei* type strain IMI 101051 from New Zealand produces two β-lactams—penicillin G and KPN—and that *mecC* and *blaZ*_{LGA251} both contribute to reduced susceptibility of *mecC*-MRSA to these antibiotics. Previous studies have established that *T. erinacei* is widespread among hedgehogs in New Zealand and northwestern Europe and that isolates from both continents produce a penicillinase-labile penicillin-like substance^{12,13,14,15,16,17,18,19}, and a recent study

characterized penicillin G from *T. erinacei* in Sweden (the presence of other β -lactams was not investigated)¹⁹. This suggests that penicillin-producing *T. erinacei* isolates were circulating in European hedgehogs long before they were introduced into New Zealand in the late 1800s and that methicillin resistance first emerged in Europe as a co-evolutionary adaptation of *S. aureus* to colonization of hedgehogs. By contrast, it cannot be ruled out that clinical use of antibiotics in humans and livestock has contributed to the evolution of some of the younger *mecC*-MRSA CC130 and CC425 lineages, although only one of these lineages (CC425:B3) showed signs of adaptation to either of these hosts. Our findings indicate that seven of the younger *mecC*-MRSA lineages (CC130:A3, CC130:A4, CC130:A5, CC130:A6, CC130:A8, CC130:A9 and CC425:B1) have acquired their type XI SCC*mec* variants from the early *mecC*-MRSA CC1943:C2 and CC1943:C3 lineages (Fig. 3 and Supplementary Figs. 2–5). The type XI SCC*mec* element has also been found at low frequencies in coagulase-negative *Staphylococcus* species from wild animals and livestock but the evolutionary links between these potential donors and *mecC*-MRSA remain to be investigated³⁶.

Our analyses suggest that most *mecC*-MRSA transmission events within hedgehog populations and between hedgehogs and secondary hosts are highly localized. The finding that some human *mecC*-MRSA isolates probably originate from local hedgehog reservoirs indicates that *mecC*-MRSA has been a cause of sporadic infections in humans for the past 200 years, more than a century before MRSA was first identified in patients in 1960 (ref. 3). The host interactions that lead to zoonotic transmission probably include direct contact with hedgehogs or contact with secondary animal hosts such as dairy cows, as previously shown for *T. erinacei* (the cause of ‘hedgehog ringworm’ in humans)¹³. We also identified several long-distance dispersal events between British and Danish islands and mainland Europe. The connections that bridge geographically isolated hedgehog populations are poorly understood but might involve oversea movements of humans and livestock. Furthermore, a recent report of *mecC*-MRSA in white storks raises the possibility that migratory birds could be efficient long-distance carriers³⁷.

β -Lactams target PBPs that catalyse carboxypeptidase and transpeptidase reactions during bacterial cell wall synthesis, whereby they inhibit cross-linking of neighbouring peptidoglycan strands. The primary mechanisms of β -lactam resistance in *S. aureus* are enzymatic cleavage of the amide bond in the β -lactam ring of penicillinase-labile penicillins by *blaZ*-encoded penicillinases and the production of *mecA*- or *mecC*-encoded PBP2a and PBP2c, respectively, with a decreased affinity for a broad spectrum of β -lactams. Our findings support that both mechanisms contribute to protection against penicillin production by *T. erinacei* in hedgehogs, although it should be noted that the $\Delta m e c C$ mutants produced smaller inhibition zones than the $\Delta b l a Z_{L G A 251}$ mutants. This might be due to the fact that PBP2c has a relatively high binding affinity for penicillins compared with PBP2a³⁸. It is also possible that PBP2c

provides additional ecological benefits, such as protection against cephalosporin-producing fungi and bacteria that occur naturally in all environments³⁹.

mecC and *blaZ*_{LGA251} have also been found on a pseudo-SCC*mec* element (Ψ SCC*mec*_{P5085}) in *Staphylococcus edaphicus*, a soil-dwelling bacterial species isolated from Antarctica⁴⁰. In contrast to the type XI SCC*mec* element, Ψ SCC*mec*_{P5085} lacks the cassette chromosome recombinase (*ccr*) genes that are responsible for movement (excision and integration) of SCC*mec*⁴⁰. Several studies have identified ARGs in ancient and modern samples and bacteria from natural environment such as soil^{41,42,43}. Yet, the environmental resistome shows limited potential for horizontal gene transfer and, for this reason, the contribution of environmental ARGs to resistance in human pathogens has so far been controversial⁴⁴. It seems reasonable to assume that the microbiota of wild animals have greater exposure to the environmental resistome than the human microbiota and are therefore more likely to acquire environmental ARGs. Thus, wild animals might represent a hitherto unrecognized conduit through which environmental ARGs can be transferred to clinical pathogens.

We acknowledge some limitations of our study. Although we broadly sampled hedgehogs across Europe and New Zealand, our study represents only a small part of the geographical range and a small number of hedgehog samples in most countries. Thus, the distribution and diversity of *mecC*-MRSA in Europe and New Zealand might be larger than documented here. We cannot account for the potential effect of transmission within the different wildlife rescue centres. To better understand the transmission dynamics, we examined the frequency of potential transmission events of *mecC*-MRSA CC130 within the different facilities using a range of maximum pairwise SNP distance thresholds to define a cluster (Extended Data Fig. 9). Using a conservative cut-off of 25 SNPs (transmission age < 6 months)⁴⁵, 25% (683 out of 2,783) of the *mecC*-MRSA CC130 isolate pairs collected within the same facility belonged to a potential transmission cluster. It is therefore possible that the prevalence of *mecC*-MRSA in hedgehogs kept in such facilities is higher than in the wild, although it is also probable that some of these potential transmission events occurred in the wild before admission to the facility. Notably, a previous study found that there is only a slightly lower prevalence of *mecC*-MRSA in hedgehogs that died in the wild compared with hedgehogs that died while staying in a wildlife rescue centre⁸, which is consistent with our finding that most hedgehogs acquired *mecC*-MRSA outside the facility. We were unable to test the hedgehogs for carriage of *T. erinacei*, because ethical constraints precluded us from collecting appropriate tissues (skin scrapings, hair and spines) for fungal culture, which leaves some important questions unanswered. For example, it remains unclear whether penicillin-producing *T. erinacei* isolates are present throughout Europe and whether there is a link between their distribution and the geographical range of *mecC*-MRSA.

In conclusion, we describe the ecological and evolutionary mechanisms that led to the emergence of methicillin resistance in the pre-antibiotic era, possibly as a co-evolutionary adaptation of *S. aureus* to colonization of dermatophyte-infected hedgehogs. These results underscore the importance of taking a broad One Health perspective on antibiotic resistance that recognises the role of natural selection in wild animals and the connectivity of natural, agricultural and human ecosystems in the evolution and spread of antibiotic-resistant pathogens.

Methods

Hedgehog survey

The aim was to collect hedgehog samples covering the geographical range of hedgehogs in Europe and New Zealand. Personnel at 16 wildlife rescue centres in ten European countries and two wildlife rescue centres in New Zealand were instructed to obtain samples from the nasal area, skin and feet of hedgehogs kept in separate enclosures using FLOQSwabs (Copan). Swabs were stored in liquid Amies medium at ambient temperature and sent to the National Reference Laboratory for Antimicrobial Resistance at Statens Serum Institut in Denmark or the Department of Veterinary Medicine at University of Cambridge in the UK immediately after sample collection. For each swab, a loopful (10 µl) of liquid Amies was inoculated into 5 ml Mueller–Hinton broth (Oxoid) supplemented with 6.5% NaCl and incubated overnight at 35–37 °C. A loopful (10 µl) of enrichment broth was then streaked on a Brilliance MRSA 2 (Oxoid) agar plate followed by incubation at 35–37 °C for 24 h. One presumptive MRSA colony from each plate was subcultured on a blood agar plate at 35–37 °C for 24 h and archived at –80 °C. We also screened all MRSA-negative samples for the presence of MSSA isolates by streaking a loopful (10 µl) of enrichment broth on a SaSelect (Bio-Rad) and Brilliance Staph 24 (Oxoid) agar plate followed by incubation at 35–37 °C for 24 h. One presumptive *S. aureus* colony from each plate was subcultured on a blood agar plate at 35–37 °C for 24 h and archived at –80 °C.

The hedgehog survey was conducted by staff members of the wildlife rescue centres who collected swab samples in connection with routine checks. In accordance with the Animal Welfare Act 1999 administered by the New Zealand Ministry for Primary Industries and Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes, no ethical approval was required as sample collection did not cause pain, suffering, distress or lasting harm equivalent to, or higher than, that caused by the introduction of a needle in accordance with good veterinary practice or deprived the animal of usual care. Ethical review was undertaken at the Department of Veterinary Medicine, University of Cambridge (CR76).

Bacterial isolates and whole-genome sequencing

A list of the 1,157 isolates used in this study is provided in Supplementary Table 1, including all of the *mecC*-MRSA ($n = 222$) and MSSA ($n = 22$) isolates identified in the hedgehog survey described above and an European collection of *mecC*-MRSA ($n = 786$) and MSSA ($n = 127$) isolates belonging to CC130, CC425 and CC1943, which constitute the most successful *mecC*-MRSA clones in Europe^{10,11,25}. Most of the Danish isolates originated from two nationwide collections of *mecC*-MRSA and MSSA isolates recovered from hedgehogs and humans. The Danish collection of hedgehog isolates included 114 *mecC*-MRSA isolates collected from 188 hedgehogs in a previous study⁸. We re-examined the 74 MRSA-negative hedgehog samples for the presence of MSSA isolates using the method described above, which led to the inclusion of two MSSA isolates belonging to CC130 ($n = 1$) and CC425 ($n = 1$) as well as six *mecC*-MRSA isolates belonging to CC130 ($n = 3$) and CC1943 ($n = 3$) that were missed in the original screening. The Danish collection of human isolates comprised 334 *mecC*-MRSA and 2 MSSA isolates that were collected from colonized or infected individuals between 1975 and 2016 as part of the national MRSA and *S. aureus* bacteraemia surveillance programmes. The remaining CC130, CC425 and CC1943 isolates were selected to represent the known geographical distribution (mainly western and central Europe) and host repertoire (mainly humans, cattle, sheep, goats and wild animals) of each clone. Whole-genome sequencing of the 1,157 isolates was performed on different Illumina platforms at the Wellcome Sanger Institute or at Statens Serum Institut. Short-read sequence data are available in the European Nucleotide Archive/NCBI Sequence Read Archive under BioProject IDs [PRJEB15105](#), [PRJEB21015](#), [PRJEB2655](#), [PRJEB2755](#), [PRJEB2756](#), [PRJEB28206](#), [PRJEB3174](#), [PRJEB32898](#), [PRJNA596428](#) and [PRJEB43456](#) and the genome accession numbers are provided in Supplementary Table 1.

Sequence analyses

Draft genomes were de novo assembled using SPAdes (v.3.15)⁴⁶. Multilocus sequence typing (MLST) was performed by comparing the draft genomes with the *S. aureus* MLST database⁴⁷. We used the *scn* gene (which encodes staphylococcal complement inhibitor-A and an indicator of the IEC-1 element⁴⁸) in *S. aureus* strain Newman (GenBank: [NC_009641](#)) and the *vwb*_{SaPI} gene in SaPIbov4 (GenBank: [HM211303](#)) as queries in BLASTN searches against the draft genomes, setting length match to 0.9 and similarity match to 0.7. These parameters were chosen to account for allelic diversity of the *vwb* genes located on SaPIs (sharing 76–100% nucleotide identities with each other)⁴⁹ while effectively excluding *scn* homologues located outside the IEC-1 element (sharing 48–61% nucleotide identities with *scn*)⁵⁰. Contigs with hits for *vwb* were analysed using PHASTER⁵¹ to ascertain that they were located on SaPIs.

ARGs were detected by mapping sequence reads against the ResFinder database⁵² using the *k*-mer alignment (KMA) tool (v.1.3)⁵³, setting both length match and similarity match to 0.9.

Phylogenetic analyses of *S. aureus* CC130, CC425 and CC1943

All 991 *mecC*-MRSA and 136 MSSA isolates belonging to CC130, CC425 and CC1943 were included. Mapping of sequence reads and SNP calling were performed using NASP (v.1.0)⁵⁴ as follows: (1) sequence reads were mapped against the reference genome of *mecC*-MRSA CC425 isolate LGA251 (GenBank: [NC_017349](#)) with the Burrows–Wheeler Alignment tool⁵⁵; (2) SNP calling was achieved using the GATK Unified Genotyper^{56,57}, setting depth of coverage and unambiguously base calls to $\geq 10\times$ and $\geq 90\%$, respectively, and ignoring insertions and deletions; and (3) SNPs contained in repeats were excluded using NUCmer^{58,59}.

Unrooted maximum-likelihood phylogenetic trees were built from core-genome SNP alignments with PhyML (v.3.0)^{60,61} under the HKY85 substitution model after applying NNI moves to improve the BIONJ starting tree. Putative recombinogenic regions were detected and a recombination-corrected phylogeny was built with ClonalFrameML (v.1.12)⁶². Time-resolved phylogenies, in which the date of each node is estimated, were constructed with BactDating (v.1.0)⁶³ using an additive uncorrelated relaxed clock model⁶⁴. Convergence and mixing of the Markov Chain Monte Carlo chains were determined using the R package coda⁶⁵. The time-resolved phylogenies were rooted as inferred with BactDating (v.1.0)⁶³ to maximize the posterior probability of the tree. Treedater (v.0.5)⁶⁶ was also applied for the same purpose and found to give similar estimates for the dating of nodes (Supplementary Figs. [6–11](#) and Source Data of Fig. 3).

The phylogenetic relationships between the type XI SCC*mec* elements were investigated in a separate analysis. A SNP alignment was generated from precalled SNPs from the 991 *mecC*-MRSA isolates. SNPs located outside the type XI SCC*mec* element (corresponding to nucleotide positions 34,403 through 63,839 in *mecC*-MRSA CC425 isolate LGA251) were manually removed, and the remaining SNPs were used to construct an unrooted maximum-likelihood phylogenetic tree using PhyML (v.3.0)^{60,61} under the HKY85 substitution model after applying NNI moves to improve the BIONJ starting tree. Putative recombinogenic regions were detected and a recombination-corrected phylogeny was built using ClonalFrameML (v.1.12)⁶². The tips were manually mapped onto the CC130, CC425 and CC1943 phylogenies, and vice versa, to identify monophyletic *mecC*-MRSA lineages harbouring orthologous type XI SCC*mec* elements.

Genome sequencing and analysis of the *T. erinacei* type strain IMI 101051

DNA extracted from the *T. erinacei* type strain IMI 101051 was used to prepare a DNA library in accordance with the Nextera XT DNA Library Prep Guide (Illumina) and sequenced on a MiSeq platform (Illumina) with 2×251 bp using a MiSeq Reagent Kit v2 (Illumina). Short-read sequencing data were submitted to the European Nucleotide Archive under BioProject [PRJEB43453](#). A draft genome was de novo assembled using SPAdes (v.3.15)⁴⁶. We used the translated protein sequences of the *P. chrysogenum* *pcbAB*, *pcbC* and *penDE* genes and the *A. chrysogenum* *cefDI*, *cefD2*, *cefEF* and *cefG* genes (Extended Data Table 1) as queries in TBLASTN searches against the draft genome. Hits were used as queries in BLASTP searches against the UniProtKB/Swiss-Prot protein database. Bidirectional best hits with a TBLASTN and BLASTP value of less than 1×10^{-100} in both directions were considered as orthologous gene pairs.

In vitro antibiotic production by the *T. erinacei* type strain IMI 101051

In vitro antibiotic production by the *T. erinacei* type strain IMI 101051 was determined using a method modified from Smith and Marples¹⁵. In brief, the strain was grown on a Sabouraud dextrose agar plate (Oxoid) for 7 days at 30 °C. The culture was removed using distilled water and a 100-µl suspension was placed in 5 ml Sabouraud dextrose broth (SDB) (Sigma-Aldrich) followed by incubation at 30 °C with shaking at 200 rpm for 3 days. The culture was placed on a Miracloth mesh (Calbiochem) in a 13-cm petri dish containing 35 ml SDB. The plate was incubated at 30 °C for 7 days, after which the mesh containing the fungal mass was placed in a 2-l conical flask containing 75 ml SDB supplemented with 2% glucose. The flask was incubated at 30 °C with shaking at 200 rpm for 7 days. The broth was replaced with fresh 5-strength SDB, and the flask was incubated at 30 °C for another 6 days. The culture medium was transferred into two 50-ml Falcon tubes containing 3 ml Diaion HP-20 resin (Supelco) slurried in 3 ml distilled water. The tubes were placed on a rotator for 1 h, after which the beads were allowed to settle. The supernatant was discarded and antibiotics were eluted from the resin with 9 ml acetone. The eluate was placed on a rotator for 30 min and centrifuged briefly to pellet the resin. The supernatant was transferred to 2-ml Eppendorf tubes, which were placed in a rotary evaporator at 30 °C. This process was repeated until all of the eluate had evaporated to dryness. To prepare bacterial inocula, *S. aureus* strains were individually grown overnight on blood agar plates (Oxoid) at 37 °C. Colonies were suspended in phosphate-buffered saline (PBS) to a 0.5 McFarland standard, diluted 1:10 in PBS, and streaked evenly on the surface of a Iso-Sensitest agar plate (Oxoid) using a sterile cotton-wool swab. Dried fungal culture extracts were suspended in 200 µl distilled water. Sterile paper discs were impregnated with 10 µl solution and placed on the agar

plates. In vitro antibiotic production was assessed by measuring the inhibition zones after overnight incubation at 35 °C.

Metabolic profiling by LC–MS

Dried fungal culture extracts were suspended in 50% methanol to a concentration of 100 mg ml⁻¹, diluted 1:10 in pure methanol and analysed using LC–MS. Metabolic profiles were obtained on a Vanquish UHPLC system (Thermo Fisher Scientific) coupled to a Vanquish diode array detector (Thermo Fisher Scientific) and an Orbitrap Fusion Tribrid high-resolution tandem mass spectrometer (Thermo Fisher Scientific). Chromatographic separation of 5 µl fungal culture extracts was performed on a Luna C18 column (3 µm, 3 × 150 mm) (Phenomenex) using a linear mobile phase gradient from 0% methanol, 90% water and 10% acetonitrile containing 1% (v/v) formic acid to 90% methanol, 0% water and 10% acetonitrile containing 1% (v/v) formic acid over 20 min at a flow rate of 400 µl min⁻¹. Ultraviolet data were recorded between 210 nm and 550 nm. For comparative purposes, a pure standard of penicillin G (Sigma-Aldrich) was prepared at a final concentration of 9.4 µg ml⁻¹ and analysed using LC–MS.

Mass spectra were obtained in the positive and negative ionization modes using the full scan and data-dependent MS² and MS³ acquisition modes. Full scan total ion current chromatograms were obtained over the range of 125–1,800 *m/z* using a spray voltage of +3.5 kV and -2.5 kV for the positive and negative ionization modes, respectively. Three different scan events were recorded for the data-dependent acquisition modes as follows: (1) MS² of the most intense ion in the full scan acquisition mode; (2) MS³ of the most intense ion in the MS² spectra; and (3) MS³ of the second-most intense ion in the MS² spectra. Additional parameters for the MS included the following: full scan resolution set to 60,000 (full-width at half-maximum, FWHM), capillary temperature set to 350 °C, ion transfer tube temperature set to 325 °C, RF lens set to 50%, automatic gain control target set to 4.0×10^5 (full scan) or 1.0×10^4 (MS² and MS³), intensity threshold set to 1.0×10^4 , collision-induced dissociation energy set to 35 eV, activation *Q* set to 0.25 and isolation window set to 4 *m/z*. Nitrogen was used as the drying, nebulizer and fragmentation gas.

Molecular networking analysis of LC–MS data

A molecular network was created using the Global Natural Products Social Molecular Networking (GNPS) workflow⁶⁷. Chromatographic raw data in the positive ionization mode were transformed to the mzXML format using the ProteoWizard command-line utility msConvert⁶⁸. The data were then filtered by removing all MS² fragment ions within ±17 Da of the precursor *m/z* value. The precursor ion mass and MS² fragment

ion tolerances were set to 2.0 Da and 0.5 Da, respectively, to enable more comprehensive comparisons with the GNPS database. MS² spectra were filtered by choosing the top six fragment ions in the ±50-Da window throughout the spectrum. A network was then created considering a cosine score above 0.7 and more than six matched peaks to link different nodes through edges. Edges between two nodes were kept in the network if each of the nodes appeared in each other's respective top ten most similar nodes. Finally, the maximum size of a molecular family was set to 100. The spectra in the network were then searched against spectral libraries in GNPS, filtering the library spectra in the same manner as the input data. All matches between network spectra and library spectra were required to have a score above 0.7 and at least six matched peaks.

Statistics

Statistical analyses were performed using GraphPad Prism (v.8.3) (GraphPad Software). We used two-tailed paired Student's *t*-tests to compare inhibition zones of each mutant to the corresponding *mecC*-MRSA wild-type strain and a two-tailed Wilcoxon matched-pairs signed rank test to compare population structures of *mecC*-MRSA at the lineage level.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

LC–MS data in mzXML format and molecular networking results are available at the MassIVE repository under identifier [MSV000087335](#). *S. aureus* short-read sequence data have been deposited in the European Nucleotide Archive/NCBI Sequence Read Archive under BioProject IDs [PRJEB15105](#), [PRJEB21015](#), [PRJEB2655](#), [PRJEB2755](#), [PRJEB2756](#), [PRJEB28206](#), [PRJEB3174](#), [PRJEB32898](#), [PRJNA596428](#) and [PRJEB43456](#) and the genome accession numbers are provided in Supplementary Table 1. *T. erinacei* type strain IMI 101051 short-read sequence data have been deposited in the European Nucleotide Archive under BioProject [PRJEB43453](#). Tree and genealogy files in Newick format are provided in Supplementary Data 1–10. [Source data](#) are provided with this paper.

References

1. 1.

Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **74**, 417–433 (2010).

2. 2.

European Centre for Disease Prevention and Control, European Medicines Agencies. *The Bacterial Challenge: Time to React. A Call to Narrow the Gap Between Multidrug-Resistant Bacteria in the EU and the Development of New Antibacterial Agents*
https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/0909_TER_The_Bacterial_Challenge_Time_to_React.pdf (2009).

3. 3.

Jevons, M. P. “Celbenin”—resistant Staphylococci. *Br. Med. J.* **1**, 124–125 (1961).

4. 4.

Harkins, C. P. et al. Methicillin-resistant *Staphylococcus aureus* emerged long before the introduction of methicillin into clinical practice. *Genome Biol.* **18**, 130 (2017).

5. 5.

Chambers, H. F. & DeLeo, F. R. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nat. Rev. Microbiol.* **7**, 629–641 (2009).

6. 6.

Price, L. B. et al. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. *mBio* **3**, e00305-11 (2012).

7. 7.

Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics
http://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf?ua=1 (WHO, 2017).

8. 8.

Rasmussen, S. L. et al. European hedgehogs (*Erinaceus europaeus*) as a natural reservoir of methicillin-resistant *Staphylococcus aureus* carrying *mecC* in

Denmark. *PLoS ONE* **14**, e0222031 (2019).

9. 9.

Bengtsson, B. et al. High occurrence of *mecC*-MRSA in wild hedgehogs (*Erinaceus europaeus*) in Sweden. *Vet. Microbiol.* **207**, 103–107 (2017).

10. 10.

García-Álvarez, L. et al. Methicillin-resistant *Staphylococcus aureus* with a novel *meca* homologue in human and bovine populations in the UK and Denmark: a descriptive study. *Lancet Infect. Dis.* **11**, 595–603 (2011).

11. 11.

Paterson, G. K., Harrison, E. M. & Holmes, M. A. The emergence of *mecC* methicillin-resistant *Staphylococcus aureus*. *Trends Microbiol.* **22**, 42–47 (2014).

12. 12.

Marples, M. J. & Smith, J. M. B. The hedgehog as a source of human ringworm. *Nature* **188**, 867–868 (1960).

13. 13.

English, M. P., Evans, C. D., Hewitt, M. & Warin, R. P. “Hedgehog ringworm”. *Br. Med. J.* **1**, 149–151 (1962).

14. 14.

Smith, J. M. B. & Marples, M. J. A natural reservoir of penicillin-resistant strains of *Staphylococcus aureus*. *Nature* **201**, 844 (1964).

15. 15.

Smith, J. M. B. & Marples, M. J. Dermatophyte lesions in the hedgehog as a reservoir of penicillin-resistant staphylococci. *J. Hyg.* **63**, 293–303 (1965).

16. 16.

Smith, J. M. B. *Staphylococcus aureus* strains associated with the hedgehog *Erinaceus europaeus*. *J. Hyg. Camb.* **63**, 293–303 (1965).

17. 17.

Morris, P. & English, M. P. *Trichophyton mentagrophytes* var. *erinacei* in British hedgehogs. *Sabouraudia* **7**, 122–128 (1969).

18. 18.

Le Barzic, C. et al. Detection and control of dermatophytosis in wild European hedgehogs (*Erinaceus europaeus*) admitted to a French wildlife rehabilitation centre. *J. Fungi* **7**, 74 (2021).

19. 19.

Dube, F., Söderlund, R., Salomonsson, M. L., Troell, K. & Börjesson, S. Benzylpenicillin-producing *Trichophyton erinacei* and methicillin resistant *Staphylococcus aureus* carrying the *mecC* gene on European hedgehogs: a pilot-study. *BMC Microbiol.* **21**, 212 (2021).

20. 20.

Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000).

21. 21.

Brockie, R. E. Distribution and abundance of the hedgehog (*Erinaceus europaeus*) L. in New Zealand, 1869–1973. *N. Z. J. Zool.* **2**, 445–462 (1975).

22. 22.

van den Berg, M. A. et al. Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. *Nat. Biotechnol.* **26**, 1161–1168 (2008).

23. 23.

Ullán, R. V., Campoy, S., Casqueiro, J., Fernández, F. J. & Martín, J. F. Deacetylcephalosporin C production in *Penicillium chrysogenum* by expression of the isopenicillin N epimerization, ring expansion, and acetylation genes. *Chem. Biol.* **14**, 329–339 (2007).

24. 24.

Kitano, K. et al. A novel penicillin produced by strains of the genus *Paecilomyces*. *J. Ferment. Technol.* **54**, 705–711 (1976).

25. 25.

Petersen, A. et al. Epidemiology of methicillin-resistant *Staphylococcus aureus* carrying the novel *mecC* gene in Denmark corroborates a zoonotic reservoir with transmission to humans. *Clin. Microbiol. Infect.* **19**, E16–E22 (2013).

26. 26.

Richardson, E. J. et al. Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nat. Ecol. Evol.* **2**, 1468–1478 (2018).

27. 27.

Holden, M. T. G. et al. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.* **23**, 653–664 (2013).

28. 28.

Strauß, L. et al. Origin, evolution, and global transmission of community-acquired *Staphylococcus aureus* ST8. *Proc. Natl Acad. Sci. USA* **114**, E10596–E10604 (2017).

29. 29.

Nübel, U. et al. Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. *Proc. Natl Acad. Sci. USA* **105**, 14130–14135 (2008).

30. 30.

Rasmussen, S. L., Nielsen, J. L., Jones, O. R., Berg, T. B. & Pertoldi, C. Genetic structure of the European hedgehog (*Erinaceus europaeus*) in Denmark. *PLoS ONE* **15**, e0227205 (2020).

31. 31.

Hansen, J. E. et al. LA-MRSA CC398 in dairy cattle and veal calf farms indicates spillover from pig production. *Front. Microbiol.* **10**, 2733 (2019).

32. 32.

Eriksson, J., Espinosa-Gongora, C., Stamphøj, I., Larsen, A. R. & Guardabassi, L. Carriage frequency, diversity and methicillin resistance of in Danish small ruminants. *Vet. Microbiol.* **163**, 110–115 (2013).

33. 33.

Danish Integrated Antimicrobial Resistance Monitoring and Research Programme. *DANMAP 2019: Use of Antimicrobial Agents and Occurrence of Antimicrobial Resistance in Bacteria From Food Animals, Food, and Humans in DENMARK*
https://www.danmap.org/-/media/Sites/danmap/Downloads/Reports/2019/DANMAP_2019.ashx?la=da&hash=AA1939EB449203EF0684440AC1477FFCE2156BA5 (2020).

34. 34.

Veterinary Medicines Directorate. *UK Veterinary Antibiotic Resistance and Sales Surveillance Report*https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950126/UK-VARSS_2019_Report_2020-TPaccessible.pdf (2020).

35. 35.

Harrison, E. M. et al. Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel *mecA* homologue *mecC*. *EMBO Mol. Med.* **5**, 509–515 (2013).

36. 36.

Loncaric, I. et al. Characterization of *mecC* gene-carrying coagulase-negative *Staphylococcus* spp. isolated from various animals. *Vet. Microbiol.* **230**, 138–144 (2019).

37. 37.

Gómez, P. et al. Detection of MRSA ST3061-t843-*mecC* and ST398-t011-*mecA* in white stork nestlings exposed to human residues. *J. Antimicrob. Chemother.* **71**, 53–57 (2016).

38. 38.

Kim, C. et al. Properties of a novel PBP2A protein homolog from *Staphylococcus aureus* strain LGA251 and its contribution to the β-lactam-resistant phenotype. *J. Biol. Chem.* **287**, 36854–36863 (2012).

39. 39.

Tahlan, K. & Jensen, S. E. Origins of the β -lactam rings in natural products. *J. Antibiot.* **66**, 401–419 (2013).

40. 40.

Pantůček, R. et al. *Staphylococcus edaphicus* sp. nov. isolated in Antarctica harbors the *mecC* gene and genomic islands with a suspected role in adaptation to extreme environment. *Appl. Environ. Microbiol.* **84**, e01746-17 (2018).

41. 41.

D’Costa, V. M., et al. Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).

42. 42.

Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A. & Handelsman, J. Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J.* **3**, 243–251 (2009).

43. 43.

Forsberg, K. J. et al. The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111 (2012).

44. 44.

Forsberg, K. J. et al. Bacterial phylogeny structures soil resistomes across habitats. *Nature* **509**, 612–616 (2014).

45. 45.

Coll, F. et al. Definition of a genetic relatedness cutoff to exclude recent transmission of meticillin-resistant *Staphylococcus aureus*: a genomic epidemiology analysis. *Lancet Microbe* **1**, e328–e335 (2020).

46. 46.

Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its application to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

47. 47.

Enright, M. C., Day, N. P., Davies, C. E., Peacock, S. J., Spratt, B. G. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J. Clin. Microbiol.* **38**, 1008–1015 (2000).

48. 48.

Van Wamel, W. J., Rooijakkers, S. H., Ruyken, M. van Kessel, K. P. & Strijp, J. A. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of *Staphylococcus aureus* are located on beta-hemolysin-converting bacteriophages. *J. Bacteriol.* **188**, 1310–1315 (2006).

49. 49.

Viana, D. et al. Adaptation of *Staphylococcus aureus* to ruminant and equine hosts involved SaPI-carried variants of von Willebrand factor-binding protein. *Mol. Microbiol.* **77**, 1583–1594 (2010).

50. 50.

Rooijakkers, S. H. M. et al. Staphylococcal complement inhibitor: structure and active sites. *J. Immunol.* **179**, 2989–2998 (2007).

51. 51.

Arndt, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).

52. 52.

Bortolaia, V. et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* **75**, 3491–3500 (2020).

53. 53.

Clausen, P. T. L. C., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw reads against redundant database with KMA. *BMC Bioinform.* **19**, 397 (2018).

54. 54.

Sahl, J. W. et al. NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. *Microb. Genom.* **2**, e000074 (2016).

55. 55.

Li, H. & Durbin, R. Fast and accurate short read alignment with Burrow-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

56. 56.

McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

57. 57.

DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

58. 58.

Delcher, A. L., Phillippy, A., Carlton, J. & Salzberg, S. L. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* **30**, 2478–2483 (2002).

59. 59.

Kurz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).

60. 60.

Guindon, S. & Gasquel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).

61. 61.

Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

62. 62.

Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genome. *PLoS Comput. Biol.* **11**, e1004041 (2015).

63. 63.

Didelot, X. et al. Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res.* **46**, e134 (2018).

64. 64.

Didelot, X., Siveroni, I. & Volz, E. M. Additive uncorrelated relaxed clock models for the dating of genomic epidemiology phylogenies. *Mol. Biol. Evol.* **38**, 307–317 (2021).

65. 65.

Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11 (2006).

66. 66.

Volz, E. M. & Frost, S. D. Scalable relaxed clock phylogenetic dating. *Virus Evol.* **3**, vex025 (2017).

67. 67.

Wang, M. et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).

68. 68.

Adusumilli, R. & Mallick, P. Data conversion with ProteoWizard msConvert. *Methods Mol. Biol.* **1550**, 339–368 (2017).

Acknowledgements

We thank M. Ganoti, F. Attila-Zoltán, M. Dugar, Z. Pokorná, D. Madsen, S. M. Gamborg, R. Molina-López, C. Rodrigues, T. Vieira, J. Viricel, A. Fingar, K. South, G. Prince, H. Gasser, S. Sebright, N. Ennew, C. Catchpole, E. Acton, N. Acton, K. Horrey and P. Loague for providing hedgehog samples; A. Medina, L. R. H. Kildevang, P. T. Hansen, and S. M. Johansson for technical assistance during the analysis of hedgehog samples; A. E. Henius for help with Figs. 1, 4 and Extended Data Fig. 8; and V. Bortolaia for reading the manuscript. This work is dedicated to V. R. Simpson, who passed away during the study. B.Č.B. was supported by a grant from the Czech University of Life Sciences Prague (no. IGA 20213106). X.D. was funded by a grant

from the National Institute for Health Research (NIHR) Health Protection Research Unit in Genomics and Enabling Data (no. NIHR200892). M.A.H. was supported by grants from the Medical Research Council (nos. G1001787/1, MR/N002660/1 and MR/P007201/1) and the Economic and Social Research Council (no. ES/S000186/1). E.M.H. was supported by a UK Research and Innovation (UKRI) Fellowship (no. MR/S00291X/1). We acknowledge Eurostat for the provision of maps, which are licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence, and administrative boundaries, which are copyright of EuroGeographics.

Author information

Author notes

1. Stien Vandendriessche

Present address: Laboratory for Medical Microbiology, Ghent University Hospital, Ghent, Belgium

2. Birgit Walther

Present address: Advanced Light and Electron Microscopy (ZBS-4), Robert Koch Institute, Berlin, Germany

3. Paolo Moroni

Present address: Quality Milk Production Services, Animal Health Diagnostic Center, Cornell University, Ithaca, NY, USA

4. Fernando García-Garrote

Present address: Servicio de Microbiología, Complejo Asistencial Universitario de Salamanca, Salamanca, Spain

5. Stefan Börjesson

Present address: Department of Microbiology, Public Health Agency of Sweden, Solna, Sweden

6. Andrew S. Waller

Present address: Intervacc AB, Stockholm, Stockholm, Sweden

7. Andrew S. Waller

Present address: Department of Biomedical Science and Veterinary Public Health,
Swedish University of Agricultural Sciences, Uppsala, Sweden

8. Matthew J. Ellington

Present address: Antimicrobial Resistance and Healthcare Associated Infections
Reference Unit, UK Health Security Agency, London, UK

9. Fiona J. E. Morgan

Present address: Department of Physiology, Development & Neuroscience,
University of Cambridge, Cambridge, UK

10. These authors contributed equally: Jesper Larsen, Claire L. Raisen

11. These authors jointly supervised this work: Mark A. Holmes, Ewan M. Harrison,
Anders R. Larsen

Affiliations

1. Department of Bacteria, Parasites & Fungi, Statens Serum Institut, Copenhagen,
Denmark

Jesper Larsen, Maiken C. Arendrup, Rasmus K. Hare, Céline Barnadas, Marc
Stegger, Raphael N. Sieber, Andreas Petersen, Øystein Angen & Anders R.
Larsen

2. Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

Claire L. Raisen, Xiaoliang Ba, David J. Seilly, Fiona J. E. Morgan, Julian
Parkhill, Nazreen F. Hadjirin & Mark A. Holmes

3. Royal Botanic Gardens, Kew, Richmond, UK

Nicholas J. Sadgrove, Guillermo F. Padilla-González & Monique S. J. Simmonds

4. Institute of Microbiology, University of Veterinary Medicine, Vienna, Austria

Igor Loncaric

5. National Reference Center for Antimicrobial Resistance and Nosocomial
Infections, Institute for Hygiene, Microbiology and Tropical Medicine,
Ordensklinikum Linz Elisabethinen, Linz, Austria

Heidrun Kerschner, Petra Apfalter & Rainer Hartl

6. National Reference Centre-Staphylococcus aureus, Department of Microbiology,
Hôpital Erasme, Université libre de Bruxelles, Brussels, Belgium

Ariane Deplano & Stien Vandendriessche
7. Department of Animal Science and Food Processing, Faculty of Tropical
AgriSciences, Czech University of Life Sciences Prague, Prague, Czech Republic

Barbora Černá Bolfíková
8. Department of Zoology, Charles University, Prague, Czech Republic

Pavel Hulva
9. Department of Biology and Ecology, University of Ostrava, Ostrava, Czech
Republic

Pavel Hulva
10. European Programme for Public Health Microbiology Training (EUPHEM),
European Centre for Disease Prevention and Control (ECDC), Stockholm,
Sweden

Céline Barnadas
11. Infectious Disease Preparedness, Statens Serum Institut, Copenhagen, Denmark

Robert L. Skov
12. Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark

Sophie L. Rasmussen
13. Wildlife Conservation Research Unit (WildCRU), Department of Zoology,
University of Oxford, Tubney, UK

Sophie L. Rasmussen
14. Department of Veterinary and Animal Sciences, Faculty of Health and Medical
Sciences, University of Copenhagen, Frederiksberg, Denmark

Carmen Espinosa-Gongora

15. National Food Institute, Technical University of Denmark, Kongens Lyngby, Denmark

Frank M. Aarestrup

16. Expert Microbiology Unit, Department of Health Security, Finnish Institute for Health and Welfare, Helsinki, Finland

Laura J. Lindholm

17. Microbiology Unit, Finnish Food Authority, Helsinki, Finland

Suvi M. Nykäsenoja

18. Bacteriology Department and French National Reference Center for Staphylococci, Hospices Civils de Lyon, University of Lyon, Lyon, France

Frederic Laurent

19. Friedrich Loeffler-Institute of Medical Microbiology, University Medicine Greifswald, Greifswald, Germany

Karsten Becker

20. Institute of Microbiology and Epizootics, Veterinary Faculty, Freie Universität Berlin, Berlin, Germany

Birgit Walther

21. Institute for Veterinary Food Science, Justus-Liebig University Giessen, Giessen, Germany

Corinna Kehrenberg

22. National Reference Centre for Staphylococci and Enterococci, Division Nosocomial Pathogens and Antibiotic Resistances, Department of Infectious Diseases, Robert Koch Institute, Wernigerode, Germany

Christiane Cuny, Franziska Layer, Guido Werner & Wolfgang Witte

23. Vet Med Labor GmbH, Kornwestheim, Germany

Ivonne Stamm

24. Dipartimento di Medicina Veterinaria, Università degli Studi di Milano, Lodi, Italy

Paolo Moroni

25. Norwegian Veterinary Institute, Ås, Norway

Hannah J. Jørgensen

26. Laboratory of Molecular Genetics, ITQB NOVA, Oeiras, Portugal

Hermínia de Lencastre

27. Laboratory of Microbiology and Infectious Diseases, The Rockefeller University, New York, NY, USA

Hermínia de Lencastre

28. Servicio de Microbiología, Hospital Universitario Lucus Augusti, Lugo, Spain

Emilia Cercenado & Fernando García-Garrote

29. Department of Animal Health and Antimicrobial Strategies, National Veterinary Institute (SVA), Uppsala, Sweden

Stefan Börjesson

30. Department of Microbiology, Public Health Agency of Sweden, Solna, Sweden

Sara Hæggman

31. Institute of Veterinary Bacteriology, University of Bern, Bern, Switzerland

Vincent Perreten

32. Department of Bacteriology, Animal and Plant Health Agency, Weybridge, UK

Christopher J. Teale

33. Animal Health Trust, Newmarket, UK

Andrew S. Waller

34. Antimicrobial Resistance and Healthcare Associated Infections Reference Unit, UK Health Security Agency, London, UK

Bruno Pichon

35. Clinical Microbiology and Public Health Laboratory, UK Health Security Agency, Addenbrooke's Hospital, Cambridge, UK

Martin D. Curran & Matthew J. Ellington

36. Department of Genetics, University of Cambridge, Cambridge, UK

John J. Welch

37. Department of Medicine, University of Cambridge, Cambridge, UK

Sharon J. Peacock & Ewan M. Harrison

38. Institute of Infection and Immunity, St George's, University of London, London, UK

Jodi A. Lindsay

39. School of Medicine, University of St Andrews, St Andrews, UK

Matthew T. G. Holden

40. Scottish MRSA Reference Laboratory, NHS Greater Glasgow and Clyde, Stobhill Hospital, Glasgow, UK

Giles F. Edwards

41. SRUC Veterinary Services, Inverness, UK

Geoffrey Foster

42. The Royal (Dick) School of Veterinary Studies and Roslin Institute, University of Edinburgh, Easter Bush, UK

Gavin K. Paterson

43. School of Life Sciences and Department of Statistics, University of Warwick, Warwick, UK

Xavier Didelot

44. Wellcome Sanger Institute, Hinxton, UK

Ewan M. Harrison

45. Department of Public Health and Primary Care, University of Cambridge,
Cambridge, UK

Ewan M. Harrison

Contributions

J.L., M.A.H., E.M.H. and A.R.L. initiated and designed the study. J.L., C.L.R., B.Č.B., P.H. and S.L.R. directed the hedgehog survey. J.L., C.L.R. X.B. and Ø.A. analysed hedgehog samples. C.L.R., I.L., H.K., P.A., R.H., A.D., S.V., R.L.S., A.P., C.E.-G., F.M.A., L.J.L., S.M.N., F. Laurent, K.B., B.W., C.K., C.C., F. Layer, G.W., W.W., I.S., P.M., H.J.J., H.d.L., E.C., F.G.-G., S.B., S.H., V.P., C.J.T., A.S.W., B.P., M.D.C., M.J.E., S.J.P., D.J.S., N.F.H., J.A.L., G.F.E., G.F., G.K.P., M.A.H. and A.R.L. provided bacterial isolates from other sources. X.B., C.B., M.S., F.J.E.M., G.K.P. and M.A.H. conducted whole-genome sequencing of bacterial isolates. J.L., C.L.R., R.N.S., J.J.W., J.P., M.T.G.H., X.D., M.A.H. and E.M.H. performed sequence and phylogenetic analyses of bacterial isolates. J.L., M.C.A. and R.K.H. performed genome sequencing and analysis of *T. erinacei* IMI 101051. C.L.R., X.B. and D.J.S. investigated in vitro antibiotic production by *T. erinacei* IMI 101051. N.J.S. prepared samples, performed LC–MS analysis and guided the identification of the penicillin-like substances. G.F.P.-G. analysed LC–MS data, performed molecular networking analysis and identified the two penicillin-like substances. M.S.J.S. coordinated and supervised the LC–MS analysis. J.L. wrote the manuscript with considerable input from C.L.R., X.B., N.J.S., G.F.P.-G., X.D., M.A.H., E.M.H. and A.R.L. All of the authors reviewed the manuscript.

Corresponding author

Correspondence to [Jesper Larsen](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Ross Fitzgerald 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 Distribution of *mecC*-MRSA clones in European and New Zealand hedgehogs.](#)

The analysis included 828 samples from the nasal area, skin and feet of 276 hedgehogs originating from 16 wildlife rescue centres in ten European countries and two wildlife rescue centres in New Zealand. **a**, Presence of *mecC*-MRSA in hedgehogs ($n = 276$). Presence and absence are shown as black and white boxes, respectively. **b**, Distribution of *mecC*-MRSA clones in hedgehog samples ($n = 828$). **c**, Distribution of MSSA clones in MRSA-negative hedgehog samples ($n = 606$).

[Source data](#)

[Extended Data Fig. 2 Detection of penicillin G in *T. erinacei* IMI 101051 culture broth.](#)

a, Left and right panels show extracted ion chromatogram of a pure standard of penicillin G and in *T. erinacei* culture broth, respectively. **b**, Upper and lower panels show MS² spectra of penicillin G in a *T. erinacei* culture broth and a pure standard of penicillin G, respectively.

Extended Data Fig. 3 Detection of KPN in *T. erinacei* IMI 101051 culture broth.

- a**, Left and right panels show extracted ion chromatogram of a pure standard of penicillin G and KPN in *T. erinacei* culture broth, respectively.
b, Upper and lower panels show MS² spectra of KPN in a *T. erinacei* culture broth and a pure standard of penicillin G, respectively.

Extended Data Fig. 4 Antibiotic activity of *T. erinacei* IMI 101051.

T. erinacei inhibition zones against a collection of *S. aureus* control strains and two *mecC*-MRSA wild-type strains belonging to CC130 (02.5099.D) and CC425 (LGA251) and their isogenic mutants. The numbers on the plates refer to each of four biologically independent fungal culture extracts.

[Source data](#)

Extended Data Fig. 5 Root-to-tip linear regression analysis of the type XI SCC*mec* dataset.

The correlation between root-to-tip distances and isolation dates is very weak with a coefficient of determination $R^2 = -0.05$.

Extended Data Fig. 6 Number of *mecC*-MRSA CC1943 sublineages.

The bars show the number of sublineages of *mecC*-MRSA CC1943:C1, *mecC*-MRSA CC1943:C2 and *mecC*-MRSA CC1943:C3 at different time points.

[Source data](#)

Extended Data Fig. 7 Geographical distribution and host range of major *mecC*-MRSA CC130, CC425 and CC1943 lineages.

The study collection included 991 *mecC*-MRSA CC130, CC425 and CC1943 isolates originating from 16 European countries. The countries of sampling and the eight largest (≥ 25 isolates) *mecC*-MRSA lineages are shown. Isolates belonging to the eight minor *mecC*-MRSA lineages are grouped together (others). The pie charts depict the proportion of *mecC*-MRSA isolates from hedgehogs (black), humans (grey) and other sources (white). The area of the pie chart is proportional to the number of *mecC*-MRSA isolates from that country.

[Source data](#)

Extended Data Fig. 8 Sampling locations of Danish *mecC*-MRSA isolates from hedgehogs and humans.

The maps of Denmark relate to the map shown in Fig. 4. **a**, *mecC*-MRSA recovered from hedgehogs ($n = 141$). **b**, *mecC*-MRSA recovered from humans ($n = 327$). The location of each sample is given at the zip code area level. Maps were provided by Eurostat under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence; the administrative boundaries are copyright of EuroGeographics.

Extended Data Fig. 9 Frequency of potential transmission events of *mecC*-MRSA CC130 isolates within wildlife rescue centres.

The bars show the proportion of isolates that belong to clusters defined at different pairwise SNP distance thresholds.

[Source data](#)

Extended Data Table 1 Penicillin biosynthetic genes in *T. erinacei* IMI 101051

Supplementary information

Supplementary Information

Supplementary Figs. 1–11 and their accompanying legends.

Reporting Summary

Supplementary Table 1

Information about the 1,157 *S. aureus* isolates used in this study.

Supplementary Data 1–10

Supplementary Data 1–10 and their accompanying legends.

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. 4

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 7

Source Data Extended Data Fig. 9

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

Larsen, J., Raisen, C.L., Ba, X. *et al.* Emergence of methicillin resistance predates the clinical use of antibiotics. *Nature* **602**, 135–141 (2022).
<https://doi.org/10.1038/s41586-021-04265-w>

- Received: 12 May 2021
- Accepted: 18 November 2021
- Published: 05 January 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04265-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

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04265-w>

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

- Article
- [Published: 26 January 2022](#)

Petabase-scale sequence alignment catalyses viral discovery

- [Robert C. Edgar](#) [ORCID: orcid.org/0000-0001-7355-2541](#)^{1 na1},
- [Jeff Taylor](#)^{2 na1},
- [Victor Lin](#) [ORCID: orcid.org/0000-0002-8670-3377](#)^{3 na1},
- [Tomer Altman](#) [ORCID: orcid.org/0000-0002-6521-5454](#)^{4 na1},
- [Pierre Barbera](#) [ORCID: orcid.org/0000-0002-3437-150X](#)^{5 na1},
- [Dmitry Meleshko](#)^{6,7 na1},
- [Dan Lohr](#)^{8 na1},
- [Gherman Novakovsky](#)^{9 na1},
- [Benjamin Buchfink](#)^{10 na1},
- [Basem Al-Shayeb](#)^{11 na1},
- [Jillian F. Banfield](#) [ORCID: orcid.org/0000-0001-8203-8771](#)^{12 na1},
- [Marcos de la Peña](#) [ORCID: orcid.org/0000-0002-7949-8459](#)^{13 na1},
- [Anton Korobeynikov](#)^{6,14 na1},
- [Rayan Chikhi](#)^{15 na1} &
- [Artem Babaian](#) [ORCID: orcid.org/0000-0002-4315-6262](#)^{2 na1}

Nature volume 602, pages 142–147 (2022)

- 28k Accesses
- 980 Altmetric
- [Metrics details](#)

Subjects

- [Computational platforms and environments](#)
- [Data mining](#)
- [Hepatitis B virus](#)
- [Metagenomics](#)
- [SARS-CoV-2](#)

Abstract

Public databases contain a planetary collection of nucleic acid sequences, but their systematic exploration has been inhibited by a lack of efficient methods for searching this corpus, which (at the time of writing) exceeds 20 petabases and is growing exponentially¹. Here we developed a cloud computing infrastructure, Serratus, to enable ultra-high-throughput sequence alignment at the petabase scale. We searched 5.7 million biologically diverse samples (10.2 petabases) for the hallmark gene RNA-dependent RNA polymerase and identified well over 10^5 novel RNA viruses, thereby expanding the number of known species by roughly an order of magnitude. We characterized novel viruses related to coronaviruses, hepatitis delta virus and huge phages, respectively, and analysed their environmental reservoirs. To catalyse the ongoing revolution of viral discovery, we established a free and comprehensive database of these data and tools. Expanding the known sequence diversity of viruses can reveal the evolutionary origins of emerging pathogens and improve pathogen surveillance for the anticipation and mitigation of future pandemics.

[Download PDF](#)

Main

Viral zoonotic disease has had a major impact on human health over the past century, with notable examples including the 1918 Spanish influenza, AIDS, SARS, Ebola and COVID-19. There are an estimated 3×10^5 mammalian virus species from which infectious diseases in humans may arise², of which only a fraction are known at present. Global surveillance of virus diversity is required for improved prediction and prevention of future epidemics, and is the focus of international consortia and hundreds of research laboratories^{3,4}.

Pioneering works expanding the virome of the Earth have each uncovered thousands of novel viruses, with the rate of virus discovery increasing exponentially and driven largely by the increased availability of high-throughput sequencing^{5,6,7,8,9,10,11}. Sequence analysis remains computationally expensive, in particular the assembly of short reads into contigs, which limits the breadth of samples analysed. Here we propose an alternative alignment-based strategy that is considerably cheaper than assembly and enables processing of massive datasets.

Petabases (1×10^{15} bases) of sequencing data are freely available in public databases such as the Sequence Read Archive (SRA)¹, in which viral nucleic acids are often captured incidental to the goals of the original studies¹². To catalyse global virus discovery, we developed the Serratus cloud computing infrastructure for ultra-high-

throughput sequence alignment, screening 5.7 million ecologically diverse sequencing libraries or 10.2 petabases of data.

Identification of Earth's virome is a fundamental step in preparing for the next pandemic. We lay the foundations for future research by enabling direct access to 883,502 RNA-dependent RNA polymerase (RdRP)-containing sequences, which include the RdRP from 131,957 novel RNA viruses (sequences with greater than 10% divergence from a known RdRP), including 9 novel coronaviruses. Altogether this captures the collective efforts of over a decade of sequencing studies in a free repository, available at <https://serratus.io>.

Accessing the planetary virome

Serratus is a free, open-source cloud-computing infrastructure optimized for petabase-scale sequence alignment against a set of query sequences. Using Serratus, we aligned more than one million short-read sequencing datasets per day for less than 1 US cent per dataset (Extended Data Fig. 1). We used a widely available commercial computing service to deploy up to 22,250 virtual CPUs simultaneously (see Methods), leveraging SRA data mirrored onto cloud platforms as part of the NIH STRIDES initiative¹³.

Our search space spans data deposited over 13 years from every continent and ocean, and all kingdoms of life (Fig. 1). We applied Serratus in two of many possible configurations. First, to identify libraries that contain known or closely related viruses, we searched 3,837,755 (around May 2020) public RNA sequencing (RNA-seq), meta-genome, meta-transcriptome and meta-virome datasets (termed sequencing runs¹) against a nucleotide pangenome of all coronavirus sequences and RefSeq vertebrate viruses. We then aligned 5,686,715 runs (January 2021) against all known viral RdRP amino acid sequences using a specially optimized version of DIAMOND v2 (ref. 14, Methods); this search was completed within 11 days, at a cost of US\$23,980 (Fig. 1a, Methods).

Fig. 1: Searching the planetary virome.

 **figure 1**

a, Total bases searched from the 5,686,715 SRA sequencing runs analysed in the viral RdRP search grouped by sample taxonomy, where available (see Extended Data Figs. 1, 3, [Supplementary Table 1](#)). A total of 8,871 out of 15,016 (59%) of known RdRP sOTUs were observed in the SRA, and 131,957 unique and novel RdRP sOTUs were identified (see Extended Data Fig. 2). sOTUs identified in multiple taxonomic groups are counted in each group separately; numbers shown indicate the number of novel sOTUs in each group. WGS, whole-genome sequencing. **b**, Release dates of the runs included in the analysis reflecting the growth rate of available data. **c**, Sample locations for 635,656 RdRP-containing contigs (27.8% of samples lacked geographical metadata). The high density of RdRP seen in North America, western Europe and eastern Asia reflects the substantial acquisition bias for samples originating from these regions. Interactive RdRP map is available at <https://serratus.io/geo>.

Previous approaches for identifying sequences across the entire SRA rely on pre-computed indexes^{15,16} that require exact substring or hash-based matches, which limits their sensitivity to diverged sequences (Extended Data Fig. 1f). Pre-assembled reads (for example, the NCBI Transcriptome Shotgun Assembly database) enable efficient alignment-based searches⁵, but are at present available for only a small fraction of the SRA. Serratus aligns a query of up to hundreds of megabytes against unassembled libraries, achieving greater sensitivity to diverged viruses compared to substring (k -mer) indexes while using far fewer computational resources than de novo assembly (Fig. 1g, Methods).

A sketch of RdRP

Viral RdRP is a hallmark gene of RNA viruses that lack a DNA stage of replication¹⁷. We identified RdRP by a well-conserved amino acid sub-sequence that we call the ‘palmprint’. Palmprints are delineated by three essential motifs that together form the catalytic core in the RdRP structure¹⁸ (Fig. 2). We constructed species-like operational taxonomic units (sOTUs) by clustering palmprints at a threshold of 90% amino acid identity, chosen to approximate taxonomic species¹⁸.

Fig. 2: RNA-dependent RNA polymerase in the SRA.

 figure 2

a, The RdRP palmprint is the protein sequence spanning three well-conserved sequence motifs (A, B and C), including intervening variable regions, exemplified within the full-length poliovirus RdRP structure with essential aspartic acid residues (asterisks) (Protein Data Bank code: 1RA6⁴⁹). Conservation was calculated from RdRP alignment in a previous study¹⁹, trimmed to the poliovirus sequence; motif sequence logos are shown below. aa, amino acids. **b**, Per-phylum histogram of amino acid identity of novel sOTUs aligned to the NCBI non-redundant protein database. Extended Data Figure 3c shows the per-order distribution. Inset, Preston plot and linear regression of palmprint abundances indicates that singleton palmprints (that is, observed in exactly one run) occur within 95% confidence intervals of the value predicted by extrapolation from high-abundance palmprints (linear regression applied to log-transformed data), and this distribution is consistent through time (Extended Data Fig. 2). NA, not applicable; uncl, unclassified.

A total of 3,376,880 (59.38%) sequencing runs contained one or more reads that mapped to the RdRP query ($E\text{-value} \leq 1 \times 10^{-4}$). We assembled RdRP aligned reads from each library (and their mate-pairs when available), which yielded 4,261,616 ‘microassembly’ contigs. Of these, 881,167 (20.7%) contained a high-confidence

palmprint identified by Palmscan (false discovery rate = 0.001)¹⁸, representing 260,808 unique palmprints. Applying Palmscan to reference databases^{1,7,19}, we obtained 45,824 unique palmprints, which clustered into 15,016 known sOTUs. If a newly acquired palmprint aligned to a known palmprint at an identity of 90% or greater, it was assigned membership to that reference sOTU; otherwise, it was designated as novel. We clustered novel palmprints at 90% identity and obtained 131,957 novel sOTUs, representing an increase in the number of known RNA viruses by a factor of 9.8. Clustering novel palmprints at genus-like 75% and family-like 40% thresholds yielded 78,485 and 3,599 novel OTUs, which represent increases of 8.0× and 1.9×, respectively (Fig. 2b).

We extracted host, geospatial and temporal metadata for each biological sample when available (Fig. 1c), noting that the majority (88%) of novel RdRP sOTUs were observed from metagenomic or environmental runs in which accurate host inference is challenging. Mapping observations of virus marker genes across time and space suggests ecological niches for these viruses, and improved characterization of sequence diversity can improve PCR primer design for *in situ* virus identification.

We estimate that around 1% of sOTUs are endogenous virus elements (EVEs); that is, viral RdRPs that have reverse-transcribed into a host germline. We did not attempt to systematically distinguish EVEs from viral RdRPs, noting that EVEs with intact catalytic motifs are likely to be recent insertions that can serve as a representative sequence for related exogenous viruses. Most (60.5%) recovered palmprints were found in exactly one run (singletons), and are observed within the expected frequency range predicted by extrapolating from more abundant sequences (Fig. 2b).

The abundance distribution of distinct palmprints is consistent with log-log-linear for each year from 2015 to 2020 (Extended Data Fig. 2e), and over time, singletons are confirmed by subsequent runs at an approximately constant rate (Extended Data Fig. 2g). The majority of novel viruses will be singletons until the diversity represented by the search query and the fraction of the planetary virome sampled in the SRA both approach saturation. Extrapolating one year forward, by when the SRA is expected to have doubled in size, we predict that 430,000 (95% confidence interval [330,000, 561,000]) additional unique palmprints could be identified by running Serratus with its current query (Fig. 2b).

RNA viruses have highly divergent sequences, even within the conserved RdRP¹⁷. Amino acid sequence alignment can recover the majority of RdRP short reads above 60% identity, but sensitivity falls as sequences diverge further (Extended Data Fig. 2f). Subsequent microassembly fragmentation can in part account for the decreased abundance of novel sOTUs below 60% identity (Fig. 2b); thus, the sensitivity to highly diverged (less than 50% identity) RdRP sequences is limited in the present study. Saturation of virus discovery within the SRA is far from complete, even if data-

growth rates are ignored. Intensive searches for so-called highly diverged or ‘dark’ viruses²⁰, in combination with iterative reanalysis (conceptually similar to PSI-BLAST²¹), are likely to yield further expansion of the known virome.

The total number of virus species is estimated to be 10^8 to 10^{12} (ref.²²), so our data captured at most 0.1% of the global virome. However, if exponential data growth combined with increased search sensitivity continues, we are at the cusp of identifying a notable fraction of Earth’s total genetic diversity with tools such as Serratus.

Expanding known *Coronaviridae*

The SARS-CoV-2 pandemic has severely affected human society. We further exemplify the potential of Serratus for virus discovery with the *Coronaviridae* (CoV) family, including a recently proposed subfamily²³ that contains a CoV-like virus, Microhyla alphaletovirus 1 (MLeV), in the frog *Microhyla fissipes*, and Pacific salmon nidovirus (PsNV) described in the endangered *Oncorhynchus tshawytscha*²⁴.

First, we identified 52,772 runs that contain 10 or more CoV-aligned reads or 2 or more CoV *k*-mers (32-mer,¹⁶). These runs were de-novo-assembled with a new version of synteny-informed SPAdes called coronaSPAdes²⁵. This yielded 11,120 identifiable CoV contigs that we annotated for a comprehensive assemblage of *Coronaviridae* in the SRA (see Methods for discussion). With these training data we defined a scoring function to predict the subsequent success of assembly (Extended Data Fig. 3b).

CoV and neighbouring palmprints comprise 70 sOTUs, 44 of which are described in public databases. Seventeen CoV sOTUs contained partial RdRP (inclusive of full palmprint) from an amplicon-based virus discovery study for which the data had not been publicly deposited at the time of writing²⁶. The remaining nine sOTUs are novel viruses, with protein domains consistent with a CoV or CoV-like genome organization (Extended Data Fig. 4).

We operationally designate MLeV, PsNV and the nine novel viruses broadly as group E, noting that all were found in samples from non-mammalian aquatic vertebrates (Fig. 3). Notably, *Ambystoma mexicanum* (axolotl) nidovirus (AmexNV) was assembled in 18 runs, 11 of which yielded common contigs of approximately 19 kb. Easing the criteria of requiring an RdRP match in a contig, 28 out of 44 (63.6%) of the runs from the associated studies were AmexNV-positive^{27,28}. Consistent assembly break points in AmexNV, PsNV and similar viruses suggest that the viral genomes of this clade of CoV-like viruses are organized in at least two segments, one containing ORF1ab with RdRP, and a shorter segment containing a lamin-associated domain protein, spike and N' accessory genes (Fig. 3). An assembly gap with common break

points is present in the published PsNV genome²⁴. Together, these seven monophyletic species possibly represent a distinct clade of segmented CoV-like nidoviruses, although molecular validation of this hypothesis is required.

Fig. 3: Expanding Coronaviridae.

 figure 3



a, Phylogram for group-E sequences. Six viruses were similar to PsNV in *Ambystoma mexicanum* (axolotl; AmexNV), *Puntigrus tetrazona* (tiger barb; PtetNV), *Hippocampus kuda* (seahorse; HkudNV), *Syngnathus typhle* (broad-nosed pipefish; StypNV), *Takifugu pardalis* (fugu fish; TparNV) and the *Acanthemblemaria* sp. (blenny; AcaNV). More-distant members identified were in *Hopomesus transpacificus* (the endangered delta smelt; HtraNV), *Silurus* sp. (catfish; SiNV) and *Monopterus albus* (asian swamp eel; MalbNV). **b**, Unrooted phylogram for Coronaviridae

annotated with genera (Greek letters) and group-E CoV-like nidoviruses (see also Extended Data Fig. 4). Maximum likelihood tree generated by clustering the RdRP amino acid sequences at 97% identity to show sub-species variability. c, Genome structure of AmexNV and the contigs recovered from group-E CoV-like viruses annotated with HMM matches. AmexNV contigs contain an identical 129-nt trailing sequence (Tr). All the putatively segmented CoV-like are monophyletic with PsNV. A gap in the PsNV reference sequence²⁴ is shown with circles, overlapping the common contig ends seen in these viruses.

While our manuscript was under review, public transcriptome screening by Miller et al.²⁹ identified three group-E CoV sequences that are not included in our sOTU analysis. One CoV⁺ library had failed at the alignment step, and microassembly from two others yielded incomplete palmprint sub-sequences and therefore lacked the required specificity for the systematic palmprint classification. A high-sensitivity reanalysis of microassemblies for any group-E RdRP sequence fragment captured the two CoV sequences that we missed from the Miller et al. study²⁹, and found another approximately 25 putative-novel CoV species from 53 fragmented contigs (Supplementary Table 1e).

In addition to identifying genetic diversity within CoV, we cross-referenced CoV⁺ library metadata to identify possible zoonoses and vectors of transmission. Discordant libraries—ones in which a CoV is identified and the viral expected host³⁰ does not match the sequencing library source taxa—were rare, accounting for only 0.92% of cases (Supplementary Table 1f).

An important limitation for these analyses is that the nucleic acid reads do not prove that viral infection has occurred in the nominal host species. For example, we identified five libraries in which a porcine, avian, or bat coronavirus was found in plant samples. The parsimonious explanation is that CoV was present in faeces or fertilizer originating from a mammalian or avian host applied to these plants. However, this exemplifies a merit of exhaustive search in identifying transmission vectors and for monitoring the geotemporal distribution of viruses.

Rapid expansion into the viral unknowns

The global mortality from viral hepatitis exceeds that of HIV/AIDS, tuberculosis or malaria³¹. Hepatitis delta virus (HDV) has a small circular RNA genome (around 1,700 nucleotides (nt)) that folds into a rod-like shape and encodes three genes: a delta antigen protein, and two self-cleaving delta ribozymes (drbz)³².

Before 2018, HDV was the sole known member of its genus; 13 drbz-containing members have since been characterized^{33,34,35,36,37,38}, and recently a second class of

ribozyme (known as hammerhead or hhrbz) characteristic of plant viroids was identified in delta-like viruses that we refer to as epsilon viruses³⁹. By sequence search for the delta antigen protein and ribozymes, we identified 14 delta viruses, 39 epsilon viruses and 311 enigmatic sequences with delta-virus-like synteny that we term zeta viruses (Fig. 4, Extended Data Fig. 5). The evolutionary histories of these mammalian delta viruses are explored further elsewhere³⁷.

Fig. 4: Expanding delta viruses and huge phages.

 figure 4

a, Genome structure for the *Marmota monax* delta virus (MmonDV) and a delta-virus-like genome detected in an environmental dataset, each containing a negative-sense delta-antigen (δ Ag) ORF; two delta ribozymes (drbz); and characteristic rod-like folding, where each line shows the predicted base-pairing within the RNA genome, coloured by base-pairing confidence score (p-num)⁵⁰. **b**, Similar genome structure for the Sulabanus spp. epsilon virus-like (SulaEV) and an epsilon-virus-like genome from an environmental dataset, each containing a negative-sense epsilon-antigen (ϵ Ag) ORF; two hammerhead ribozymes (hhrbz); and rod-like folding. **c**, Example of the compact genome structure of a Zeta virus-like from an environmental dataset containing two predicted zeta-antigen (ζ Ag $^{+/-}$; protein alignment is shown in the outer circles) ORFs without stop codons; two hhrbz overlapping with the ORFs; and rod-like folding. Further novel genomes are shown in Extended Data Figs. 5, 6. **d**, Maximum-likelihood phylogenetic tree of delta viruses derived from a delta-antigen protein alignment with bootstrap values. Two divergent environmental delta viruses could not yet be placed. **e**, Tree showing huge phage clade expansion. Black dots indicate branches with bootstrap values greater than 90. Outer ring indicates genome or genome fragment length: grey are sequences from Al-Shayeb et al.⁴² and reference sequences, shadings indicate previously defined clades of phages with very large

genomes (200–735 kb). The Kabirphages (light purple) are shown in expanded view in Extended Data Fig. 7.

The zeta virus circular genomes are highly compressed, ranging from 324 to 789 nt and predicted to fold into rod-like structures. They contain a hhrbz in each orientation and encode two open reading frames (ORFs), one sense and one anti-sense. Both ORFs generally lack stop codons and encompass the entire genome, potentially producing an endless tandem repeat of antigen. The atypical coiled-coil domain of the HDV antigen⁴⁰ is conserved in the antigens of new delta and epsilon viruses, whereas epsilon and zeta genomes show analogous hhrbz (Extended Data Fig. 6), suggesting that these sequences share common ancestry. These abundant elements may help to solve a long-standing question about the origins of circular RNA subviral agents in higher eukaryotes (Extended Data Fig. 6), historically regarded as molecular fossils of a prebiotic RNA world⁴¹.

To evaluate the feasibility of applying Serratus in the context of microbiome research, we sought to locate bacteriophages that are related to recently reported huge phages⁴², searching for terminase amino acid sequences. Targeted assembly of 287 high-scoring runs returned 252 terminase-containing contigs of greater than 140 kb. Phylogenetics of these sequences resolved new groups of phages with large genomes (Fig. 4e). Although most phages were from a single animal genus, we identified closely related phages that crossed animal orders, including related phages in a human from Bangladesh (ERR866585) and in groups of cats (PRJEB9357) and dogs (PRJEB34360) from England, sampled five years apart. Similarly, we recovered two approximately 554-kb Lak megaphage genomes (among the largest animal microbiome phages reported so far) that are extremely closely related to sequences previously reported from pigs, baboons and humans⁴³ (Extended Data Fig. 7). These two genomes were circularized and manually curated to completion. The large carrying capacity of such phages and broad distribution underlines their potential for extensive lateral gene transfer amongst animal microbiomes and modification of host bacterial function. These sequences substantially expand the inventory of phages with genomes whose length range overlaps with those of bacteria.

Discussion

Since the completion of the human genome, the growth of DNA sequencing databases has outpaced Moore's Law. Serratus provides rapid and focused access to genomic sequences captured over more than a decade by the global research community, which would otherwise be inaccessible in practice. This work and further extensions of petabase-scale genomics^{15,16,44} are shaping a new era in computational biology, enabling expansive gene discovery, pathogen surveillance and pangenomic evolutionary analyses.

Optimal translation of such massive datasets into meaningful biomedical advances requires free and open collaboration among scientists⁴⁵. The current pandemic underscores the need for prompt, unrestricted and transparent data sharing. With these goals in mind, we deposited 7.3 terabytes of virus alignments and assemblies into an open-access database that can be explored via a graphical web interface at <https://serratus.io> or programmatically through the Tantalus R package and its PostgreSQL interface.

The ‘metagenomics revolution’ of virus discovery is accelerating^{7,11}. Innovative fields such as high-throughput viromics⁴⁶ can leverage vast collections of virus sequences to inform policies that predict and mitigate emerging pandemics⁴⁷. Combining ecoinformatics with virus, host and geotemporal metadata offers a proof-of-concept for a global pathogen surveillance network, arising as a by-product of centralized and open data sharing.

Human population growth and encroachment on animal habitats is bringing more species into proximity, leading to an increased rate of zoonosis² and accelerating the Anthropocene mass extinction⁴⁸. While Serratus enhances our capability to chronicle the full genetic diversity of our planet, the genetic diversity of the biosphere is diminishing. Thus, investment in the collection and curation of biologically diverse samples, with an emphasis on geographically underrepresented regions, has never been more pressing—if not for the conservation of endangered species, then to better conserve our own.

Methods

Serratus alignment architecture

Serratus (v0.3.0) (<https://github.com/ababaian/serratus>) is an open-source cloud-infrastructure designed for ultra-high-throughput sequence alignment against a query sequence or pangenome (Extended Data Fig. 1). Serratus compute costs are dependent on search parameters (expanded discussion available:

https://github.com/ababaian/serratus/wiki/pangenome_design). The nucleotide vertebrate viral pangenome search (bowtie2, database size: 79.8 MB) reached processing rates of 1.29 million SRA runs in 24 h at a cost of US\$0.0062 per dataset (Extended Data Fig. 1). The translated-nucleotide RdRP search (DIAMOND¹⁴; database size: 7.1 MB) reached processing rates exceeding 0.5 million SRA runs in 12 h at a cost of US\$0.0042 per dataset. All 5,686,715 runs analysed in the RdRP search were completed within 11 days for a total cost of US\$23,980 or around US\$2,350 per petabase. For a detailed breakdown of Serratus project costs and recommendations for managing cloud-computing costs, see Serratus wiki:

<https://github.com/ababaian/serratus/wiki/budget>. Tutorials on how to find particular novel viruses using Serratus data are available at https://github.com/ababaian/serratus/wiki/Find_novel_viruses.

Computing cluster architecture

The processing of each sequencing library is split into three modules: ‘dl’ (download), ‘align’ and ‘merge’. The dl module acquires compressed data (.sra format) via prefetch (v2.10.4), from the Amazon Web Services (AWS) Simple Storage Service (S3) mirror of the SRA, decompresses to FASTQ with fastq-dump (v2.10.4) and splits the data into chunks of 1 million reads or read-pairs (‘fq-blocks’) into a temporary S3 cache bucket. To mitigate excessive disk usage caused by a few large datasets, a total limit of 100 million reads per dataset was imposed. The align module reads individual fq-blocks and aligns to an indexed database of user-provided query sequences using either bowtie2 (v2.4.1, *--very-sensitive-local*)⁵¹ for nucleotide search, or DIAMOND (v2.0.6 development version, *--mmap-target-index --target-indexed --masking 0 --mid-sensitive -s 1 -c1 -p1 -k1 -b 0.75*)¹⁴ for translated-protein search. Finally, the merge module concatenates the aligned blocks into a single output file (.bam for nucleotide, or .pro for protein) and generates alignment statistics with a Python script (see details about Summarizer in ‘Generating viral summary reports’ below).

Computing resource allocation

Each component is launched from a separate AWS autoscaling group with its own launch template, allowing the user to tailor instance requirements per task. This enabled us to minimize the use of costly block storage during compute-bound tasks such as alignment. We used the following Spot instance types; dl: 250 GB SSD block storage, 8 virtual CPUs (vCPUs), 32 GB RAM (r5.xlarge) around 1,300 instances; align: 10 GB SSD block storage, 8 vCPUs, 8 GB RAM (c5.xlarge) around 4,300 instances; merge: 150 GB SSD block storage, 4 vCPUs, 4 GB RAM (c5.large) around 60 instances. Users should note that it may be necessary to submit a service ticket to access more than the default EC2 instance limit.

AWS Elastic Compute Cloud (EC2) instances have higher network bandwidth (up to 1.25 GB s^{-1}) than block storage bandwidth (250 MB s^{-1}). To exploit this, we used S3 buckets as a data buffering and streaming system to transfer data between instances following methods developed in a previous cloud architecture (<https://github.com/FredHutch/sra-pipeline>). This, combined with splitting of FASTQ files into individual blocks, effectively eliminated file input/output (i/o) as a bottleneck, as the available i/o is multiplied per running instance (conceptually analogous to a RAID0 configuration or a Hadoop distributed file system⁵²).

Using S3 as a buffer also allowed us to decouple the input and output of each module. S3 storage is cheap enough that in the event of unexpected issues (for example, exceeding EC2 quotas) we could resolve system problems in real time and resume data processing. For example, shutting down the align modules to hotfix a genome indexing problem without having to re-run the dl modules, or if an alignment instance is killed by a Spot termination, only that block needs to be reprocessed instead of the entire sequencing run.

Work queue and scheduling

The Serratus scheduler node controls the number of desired instances to be created for each component of the workflow, based on the available work queue. We implemented a pull-based work queue. After boot-up, each instance launches a number of ‘worker’ threads equal to the number of CPUs available. Each worker independently manages itself via a boot script, and queries the ‘scheduler’ for available tasks. Upon completion of the task, the worker updates the scheduler of the result: success, or fail, and queries for a new task. Under ideal conditions, this allows for a worst-case response rate in the hundreds of milliseconds, keeping cluster throughput high. Each task typically lasts several minutes depending on the pangenome.

The scheduler itself was implemented using Postgres (for persistence and concurrency) and Flask (to pool connections and translate REST queries into SQL). The Flask layer allowed us to scale the cluster past the number of simultaneous sessions manageable by a single Postgres instance. The work queue can also be managed manually by the user, to perform operations such as re-attempting the downloading of an SRA accession after a failure or to pause an operation while debugging. Up to 300,000 SRA jobs can be processed in the work queue per batch process.

The system is designed to be fully self-scaling. An ‘autoscaling controller’ was implemented, which scales-in or scales-out the desired number of instances per task every five minutes on the basis of the work queue. As a backstop, when all workers on an instance fail to receive work instructions from the scheduler, the instance self shuts-down. Finally a ‘job cleaner’ component checks the active jobs against currently running instances. If an instance has disappeared owing to SPOT termination or manual shutdown, it resets the job allowing it to be processed up by the next available instance.

To monitor cluster performance in real-time, we used Prometheus (v2.5.0) and node exporter to retrieve CPU, disk, memory and networking statistics from each instance, to expose performance information about the work queue, and Python exporter to export information from the Flask server. This allowed us to identify and diagnose performance problems within minutes to avoid costly overruns.

Generating viral summary reports

We define a viral pangenome as the entire collection of reference sequences belonging to a taxonomic viral family, which may contain both full-length genomes and sequence fragments such as those obtained by RdRP amplicon sequencing.

We developed a Summarizer module written in Python to provide a compact, human- and machine-readable synopsis of the alignments generated for each SRA dataset. The method was implemented in `Serratus_summarizer.py` for nucleotide alignment and `Serratus_psummarizer.py` for amino acid alignments. Reports generated by the Summarizer are text files with three sections described in detail online (<https://github.com/ababaian/serratus/wiki/.summary-Reports>). In brief, each contains a header section with alignment metadata and one-line summaries for each virus family pangenome, reference sequence and gene, respectively, with gene summaries provided for protein alignments only.

For each summary line we include descriptive statistics gathered from the alignment data such as the number of aligned reads, estimated read depth, mean alignment identity and coverage; that is, the distribution of reads across each reference sequence or pangenome. Coverage is measured by dividing a reference sequence into 25 equal bins and depicted as an ASCII text string of 25 symbols, one per bin; for example *oaooomoUU:oWWUUWOWamWAauW*. Each symbol represents $\log_2(n + 1)$, where n is the number of reads aligned to a bin in this order $_.:uwaomUWAOM^$. Thus, ‘’ indicates no reads, ‘.’ exactly one read, ‘:’ two reads, ‘u’ 3–4 reads, ‘w’ 5–7 reads and so on; ‘^’ represents $>2^{13} = 8,192$ reads in the bin. For a pangenome, alignments to its reference sequences are projected onto a corresponding set of 25 bins. For a complete genome, the projected pangenome bin number 1, 2, ..., 25 is the same as the reference sequence bin number. For a fragment, a bin is projected onto the pangenome bin implied by the alignment of the fragment to a complete genome. For example, if the start of a fragment aligns halfway into a complete genome, bin 1 of the fragment is projected to bin $\text{floor}(25/2) = 12$ of the pangenome. The introduction of pangenome bins was motivated by the observation that bowtie2 selects an alignment at random when there are two or more top-scoring alignments, which tends to distribute coverage over several reference sequences when a single viral genome is present in the reads. Coverage of a single reference genome may therefore be fragmented, and binning to a pangenome better assesses coverage over a putative viral genome in the reads while retaining pangenome sequence diversity for detection.

Identification of viral families within a sequencing dataset

The Summarizer implements a binary classifier predicting the presence or absence of each virus family in the query on the basis of pangenome-aligned short reads. For a

given family F , the classifier reports a score in the range [0,100] with the goal of assigning a high score to a dataset if it contains F and a low score if it does not. Setting a threshold on the score divides datasets into disjoint subsets representing predicted positive and negative detections of family F . The choice of threshold implies a trade-off between false positives and false negatives. Sorting by decreasing score ranks datasets in decreasing order of confidence that F is present in the reads.

Naively, a natural measure of the presence of a virus family is the number of alignments to its reference sequences. However, alignments may be induced by non-homologous sequence similarity, for example low-complexity sequence.

The score for a family was therefore designed to reflect the overall coverage of a pangenome because coverage across all or most of a pangenome is more likely to reflect true homology; that is, the presence of a related virus. Ideally, coverage would be measured individually for each base in the reference sequence, but this could add undesirable overhead in compute time and memory for a process that is executed in the Linux alignment pipe (FASTQ decompression → aligner → Summarizer → alignment file compression). Coverage was therefore measured by binning as described above, which can be implemented with minimal overhead.

A virus that is present in the reads with coverage too low to enable an assembly may have less practical value than an assembled genome. Also, genomes with lower identity to previously known sequences will tend to contain more novel biological information than genomes with high identity but highly diverged genomes will tend to have fewer aligned reads. With these considerations in mind, the classifier was designed to give higher scores when coverage is high, read depth is high and/or identity is low. This was accomplished as follows. Let H be the number of bins with at least 8 alignments to F , and L be the number of bins with from 1 to 7 alignments. Let S be the mean alignment percentage identity, and define the identity weight $w = (S/100)^{-3}$, which is designed to give higher weight to lower identities, noting that w is close to 1 when identity is close to 100% and increases rapidly at lower identities. The classification score for family F is calculated as $Z_F = \max(w(4H + L)), 100$. By construction, Z_F has a maximum of 100 when coverage is consistently high across a pangenome, and is also high when identity is low and coverage is moderate, which may reflect high read depth but many false negative alignments due to low identity. Thus, Z_F is greater than zero when there is at least one alignment to F and assigns higher scores to SRA datasets that are more likely to support successful assembly of a virus belonging to F .

Sensitivity to novel viruses as a function of identity

We aimed to assess the sensitivity of our pipeline as a function of sequence identity by asking what fraction of novel viruses is detected at increasingly low identities compared to the reference sequences used for the search. Several variables other than identity affect sensitivity, including read length, whether reads are mate-paired, sequencing error rate, coverage bias and the presence of other similar viruses that may cause some variants to be unreported in the contigs. Coverage bias can render a virus with high average read depth undetectable, in particular if the query is RdRP-only and the RdRP gene has low coverage or is absent from the reads. Successful detection might be defined in different ways, depending on the goals of the search; for example, a single local alignment of a reference to a read (maximizing sensitivity, but not always useful in practice); a microassembled palmprint; a full assembly contig that contains a complete palmprint or otherwise classifiable fragment of a marker gene; or an assembly of a complete genome. We assessed alignment sensitivity of *bowtie2* -- *very-sensitive-local* and Serratus-optimized DIAMOND¹⁴ as a function of identity by simulating typical examples in a representative scenario: unpaired reads of length 100 with a base call error rate of 1%. We manually selected test-reference pairs of RefSeq complete *Ribovirus* genomes at RdRP amino acid identities 100%, 95% ... 20%, generating simulated length-100 reads at uniformly distributed random locations in the test genome with a mean coverage of 1,000×. For *bowtie2*, the complete reference genome was used as a reference; for DIAMOND the reference was the translated amino acid sequence of the RdRP gene (400 amino acids), which was identified by aligning to the ‘wolf18’ dataset. These choices model the coronavirus pangenome used as a *bowtie2* query and the *rdrp1* protein reference used as a DIAMOND query, respectively. Sensitivity was assessed as the fraction of reads aligned to the reference. With *bowtie2*, the number of unmapped reads reflects a combination of lack of alignment sensitivity and divergence in gene content as some regions of the genome may lack homology to the reference. With DIAMOND, the number of unmapped reads reflects a combination of lack of alignment sensitivity and the fraction of the genome that is not RdRP, which varies by genome length 1g. They show that the fraction of aligned reads by *bowtie2* drops to around 2% to 4% at 90% RdRP amino acid identity, and maps no reads for most of the lower identity test–reference pairs. DIAMOND maps around 5% to 10% of reads down to 50% RdRP amino acid identity, then less than 1% at lower identities; around 30% to 35% is the lower limit of practical detection.

Defining viral pangomes and the SRA search space

Nucleotide search pangomes

To create a collection of viral pangomes, a comprehensive set of complete and partial genomes representing the genetic diversity of each viral family, we used two approaches.

For *Coronaviridae*, we combined all RefSeq ($n = 64$) and GenBank ($n = 37,451$) records matching the NCBI Nucleotide⁵³ server query “txid11118[Organism:exp]” (date accessed: 1 June 2020). Sequences of fewer than 200 nt were excluded as well as sequences identified to contain non-CoV contaminants during preliminary testing (such as plasmid DNA or ribosomal RNA fragments). Remaining sequences were clustered at 99% identity with UCLUST (USEARCH: v11.0.667)⁵⁴ and masked by Dustmasker (ncbi-blast:2.10.0) (--window 30 and --window 64)⁵⁵. The final query contained 10,101 CoV sequences (accessions in [Supplementary Table 1a](#); masked coordinates in [Supplementary Table 1b](#)). SeqKit (v0.15) was used for working with fasta files⁵⁶.

For all other vertebrate viral family pangenomes, RefSeq sequences ($n = 2,849$) were downloaded from the NCBI Nucleotide server with the query "Viruses[Organism] AND srcref[PROP] NOT wgs[PROP] NOT cellular organisms[ORGN] NOT AC 000001:AC 999999[PACC] AND ("vhost human"[Filter] AND "vhost vertebrates"[Filter])" (date accessed: 17 May 2020). Retroviruses ($n = 80$) were excluded as preliminary testing yielded excessive numbers of alignments to transcribed endogenous retroviruses. Each sequence was annotated with its taxonomic family according to its RefSeq record; those for which no family was assigned by RefSeq ($n = 81$) were designated as ‘unknown’.

The collection of these pangenomes was termed ‘cov3m’, and was the nucleotide sequence reference used for this study.

Amino acid viral RdRP search panproteome

For the translated-nucleotide search of viral RNA-dependent RNA polymerase (RdRP; hereinafter viral RdRP is implied) we combined sequences from several sources. (1) The ‘wolf18’ collection is a curated snapshot (around 2018) of RdRP from GenBank (ref. [19](#) accessed: ftp://ftp.ncbi.nlm.nih.gov/pub/wolf/_suppl/rnavir18/RNAVirome.S2.afa). (2) The ‘wolf20’ collection is RdRPs from assembled from marine metagenomes (ref. [7](#) accessed: ftp://ftp.ncbi.nlm.nih.gov/pub/wolf/_suppl/yangshan/gb_rdp.afa). (3) All viral GenBank protein sequences were aligned with DIAMOND --ultra-sensitive¹⁴ against the combined wolf18 and wolf20 sequences (E -value $< 1 \times 10^{-6}$). These produced local alignments that contained truncated RdRP, so each RdRP-containing GenBank sequence was then re-aligned to the wolf18 and wolf20 collection to ‘trim’ them to ‘wolf’ RdRP boundaries. (4) The above algorithm was also applied to all viral GenBank nucleotide records to capture additional RdRP not annotated as such by GenBank. A region of HCV capsid protein shares similarity to HCV RdRP; sequences annotated as HCV capsid were therefore removed. Eight novel coronavirus RdRP

sequences identified in a pilot experiment were added manually. The combined RdRP sequences from the above collections were clustered (UCLUST) at 90% amino acid identity and the resulting representative sequences (centroids, $n = 14,653$) used as the rdp1 search query.

In addition, we added delta virus antigen proteins from NC 001653, M21012, X60193, L22063, AF018077, AJ584848, AJ584847, AJ584844, AJ584849, MT649207, MT649208, MT649206, NC 040845, NC 040729, MN031240, MN031239, MK962760, MK962759 and eight additional homologues we identified in a pilot experiment.

SRA search space and queries

To run Serratus, a target list of SRA run accessions is required. We defined 11 (not-mutually exclusive) queries as our search space, which were named human, mouse, mammal, vertebrate, invertebrate, eukaryotes, prokaryotes/others, bat (including genomic sequences), virome, metagenome and mammalian genome ([Supplementary Table 1c](#)). Our search was restricted to Illumina sequencing technologies and to RNA-seq, meta-genomic and meta-transcriptome library types for these organisms (except for the mammalian genome query, which was genome or exome). Before each Serratus deployment, target lists were depleted of accessions already analysed. Reprocessing of a failed accession was attempted at least twice. In total, we aligned 3,837,755/4,059,695 (94.5%) of the runs in our nucleotide-pangenome search (around May 2020) and 5,686,715/5,780,800 (98.37%) of the runs in our translated-nucleotide RdRP search (around January 2021).

User interfaces for the Serratus databases

We implemented an on-going, multi-tiered release policy for code and data generated by this study, as follows. All code, electronic notebooks and raw data are immediately available at <https://github.com/ababaian/serratus> and on the s3://serratus-public bucket, respectively. Upon completion of a project milestone, a structured data release is issued containing raw data into our viral data warehouse s3://lovelywater/. For example, the .bam nucleotide alignment files from 3.84 million SRA runs are stored in s3://lovelywater/bam/X.bam; and the protein .summary files are in s3://lovelywater/psummary/X.psummary, where X is a SRA run accession. These structured releases enable downstream and third-party programmatic access to the data.

Summary files for every searched SRA dataset are parsed into a publicly accessible AWS Relational Database (RDS) instance that can be queried remotely via any PostgreSQL client. This enables users and programs to perform complex operations

such as retrieving summaries and metadata for all SRA runs matching a given reference sequence with above a given classifier score threshold. For example, one can query for all records containing at least 20 aligned reads to hepatitis delta virus (NC 001653.2) and the associated host taxonomy for the corresponding SRA datasets:

```
SELECT sequence_accession, run_id, tax_id, n_reads FROM nsequence JOIN srarun  
ON (nsequence.run_id = srarun.run) WHERE n_reads >= 20
```

For users unfamiliar with SQL, we developed Tantalus (<https://github.com/serratus-bio/tantalus>), an R programming-language package that directly interfaces the Serratus PostgreSQL database to retrieve summary information as data-frames. Tantalus also offers functions to explore and visualize the data.

Finally, the Serratus data can be explored via a graphical web interface by accession, virus or viral family at <https://serratus.io/explorer>. Under the hood, we developed a REST API to query the database from the website. The website uses React+D3.js to serve graphical reports with an overview of viral families found in each SRA accession matching a user query.

All four data access interfaces are under ongoing development, receiving community feedback via their respective GitHub issue trackers to facilitate the translation of this data collection into an effective viral discovery resource. Documentation for data access methods is available at <https://serratus.io/access>.

Geocoding BioSamples

To generate the map in Fig. 1c, we parsed and extracted geographical information from all 16 million BioSample XML submissions⁵⁷. Geographic information is either in the form of coordinates (latitude and longitude) or freeform text (for example, ‘France’, ‘Great Lakes’). For each BioSample, coordinate extraction was attempted using regular expressions. If that failed, text extraction was attempted using a manually curated list of keywords that capture BioSample attribute names that are likely to contain geographical information. If that failed, then we were unable to extract geographical information for that BioSample. Geocoding the text to coordinates was done using Amazon Location Service on a reduced set of distinct filtered text values (52,028 distinct values from 2,760,241 BioSamples with potential geographical text). BioSamples with geocoded coordinates were combined with BioSamples with submitted coordinate information to form a set of 5,325,523 geospatial BioSamples. This is then cross-referenced with our subset of SRA accessions with an RdRP match to generate the figure.

All intermediate and resulting data from this step are stored on the SQL database described above. Development work is public at <https://github.com/serratus-bio/biosample-sql>.

Viral alignment, assembly and annotation

Upon identification of CoV reads in a run from alignment, we assembled 52,772 runs containing at least 10 reads that aligned to our CoV pangenome or at least 2 reads with CoV-positive k -mers¹⁶. A total of 11,120 of the resulting assemblies contained identifiable CoV contigs, of which only 4,179 (37.58%) contained full-length CoV RdRP ([Supplementary Table 1d](#)). The discrepancy between alignment-positive, assembly-positive and RdRP-positive libraries arises owing to random sampling of viral reads and assembly fragmentation. In this respect, alignment or k -mer based methods are more sensitive than assembly in detecting for the presence of low-abundance viruses (genome coverage < 1) with high identity to a reference sequence. Scoring libraries for genome coverage and depth is a good predictor of ultimate assembly success (Extended Data Fig. 3); thus, it can be used to efficiently prioritize computationally expensive assembly in the future, as has been previously demonstrated for large-scale SRA alignment analyses⁵⁸.

DIAMOND optimization and output

To optimize DIAMOND¹⁴ for small (<10 MB) databases such as the RdRP search database, we built a probabilistic hash set that stores 8-bit hash values for the database seeds, using SIMD instructions for fast probing. This index is loaded as a memory mapped file to be shared among processes and allows us to filter the query reads for seeds contained in the database, thus omitting the full construction of the query seed table. We also eliminated the overhead of building seed distribution histograms that is normally required to allocate memory and construct the query table in a single pass over the data using a deque-like data structure. In addition, query reads were not masked for simple repeats, as the search database is already masked. These features are available starting from DIAMOND v2.0.8 with the command line flags --target-indexed --masking 0. In a benchmark of 4 sets of 1 million reads from a bat metagenome (ERR2756788), the implemented optimization produced a speed-up of ×1.47 and reduced memory use by 64%, compared to the public unmodified DIAMOND v2.0.6, using our optimized set of parameters in both cases (see 1.1.1). Together, the optimized parameters and implementation reduced DIAMOND runtime against RdRP search from 197.96 s (s.d. = 0.18 s), to 21.29 s (s.d. = 0.23 s) per million reads, a speed-up of a factor of 9.3. This effectively reduced the computational cost of translated-nucleotide search for Serratus from US\$0.03 to US\$0.0042 per library.

DIAMOND output files (we label .pro) were specified with the command -f 6 qseqid qstart qend qlen qstrand sseqid sstart send slen pident evalue cigar qseq_translated full_qseq full_qseq_mate.

coronaSPAdes

RNA viral genome assembly faces several distinct challenges stemming from technical and biological bias in sequencing data. During library preparation, reverse transcription introduces 5' end coverage bias, and GC-content skew and secondary structures lead to unequal PCR amplification⁵⁹. Technical bias is confounded by biological complexity such as intra-sample sequence variation due to transcript isoforms and/or to the presence of multiple strains.

To address the assembly challenges specific to RNA viruses, we developed coronaSPAdes (v3.15.3), which is described in detail in a companion manuscript²⁵. In brief, rnaviralSPAdes and the more specialized variant, coronaSPAdes, combines algorithms and methods from several previous approaches based on metaSPAdes⁶⁰, rnaSPAdes⁶¹ and metaviralSPAdes⁶² with a HMMPPathExtension step. coronaSPAdes constructs an assembly graph from an RNA-seq dataset (transcriptome, meta-transcriptome, and meta-virome are supported), removing expected sequencing artifacts such as low complexity (poly-A/poly-T) tips, edges, single-strand chimeric loops or double-strand hairpins⁶¹ and subspecies-bases variation⁶².

To deal with possible misassemblies and high-covered sequencing artefacts, a secondary HMMPPathExtension step is performed to leverage orthogonal information about the expected viral genome. Protein domains are identified on all assembly graphs using a set of viral hidden Markov models (HMMs), and similar to biosyntheticSPAdes⁶³, HMMPPathExtension attempts to find paths on the assembly graph that pass through significant HMM matches in order.

coronaSPAdes is bundled with the Pfam SARS-CoV-2 set of HMMs⁶⁴, although these may be substituted by the user. This latter feature of coronaSPAdes was used for HDV assembly, in which the HMM model of HDAg, the hepatitis delta antigen, was used instead of the Pfam SARS-CoV-2 set. Note that despite the name, the HMMs from this set are quite general, modelling domains found in all coronavirus genera in addition to RdRP, which is found in many RNA virus families. Hits from these HMMs cover most bases in most known coronavirus genomes, enabling the recovery of strain mixtures and splice variants.

Microassembly of RdRP-aligned reads

Reads aligned by DIAMOND¹⁴ in the translated-nucleotide RdRP search are stored in the .pro alignment file. All sets of mapped reads (3,379,127 runs) were extracted, and each non-empty set was assembled with rnaviralSPAdes (v3.15.3)²⁵ using default parameters. This process is referred to as ‘microassembly’, as a collection of DIAMOND hits is orders of magnitude smaller than the original SRA accession (40 ± 534 KB compressed size, ranging from a single read up to 53 MB). Then bowtie2⁵¹ (default parameters) was used to align the DIAMOND read hits of an accession back to the microassembled contigs of that accession. Palmscan (v1.0.0, -rdrp -hicon)¹⁸ was run on microassembled contigs, resulting in high-confidence palmprints for 337,344 contigs. Finally mosdepth (v0.3.1)⁶⁵ was used to calculate a coverage pileup for each palmprint hit region within microassembled contigs.

Classification of assembled RdRP sequences

Our methods for RdRP classification are described and validated in a companion paper¹⁸. In brief, we defined a barcode sequence, the polymerase palmprint (PP), as an approximately 100-amino-acid segment of the RdRP palm subdomain delineated by well-conserved catalytic motifs. We implemented an algorithm, Palmscan, to identify palmprint sequences and discriminate RdRPs from reverse transcriptases. The combined set of RdRP palmprints from public databases and our assemblies was classified by clustering into operational taxonomic units (OTUs) at 90%, 75% and 40% identity, giving species-like, genus-like and family-like clusters (sOTUs, gOTUs and fOTUs), respectively. Tentative taxonomy of novel OTUs was assigned by aligning to palmprints of named viruses and taking a consensus of the top hits above the identity threshold for each rank.

Quality control of assembled RdRP sequences

Our goal was to identify novel viral RdRP sequences and novel sOTUs in SRA libraries. From this perspective, we considered the following to be erroneous to varying degrees: sequences that are (a) not polymerases; (b) not viral; (c) with differences due to experimental artefacts; or (d) with sufficient differences to cause a spurious inference of a novel sOTU. We categorized potential sources of such errors and implemented quality control procedures to identify and mitigate them, as follows.

Point errors are single-letter substitution and indel errors that may be caused by PCR or sequencing per se. Random point errors are not reproduced in multiple non-PCR duplicate reads and are unlikely to assemble because such errors almost always induce identifiable structures in the assembly graph (tips and bubbles) that are pruned during graph simplification. In rare cases, a contig may contain a read with random point errors. Such contigs will have low coverage of around 1, and we therefore recorded

coverage as a quality control metric and assessed whether low-coverage assemblies were anomalous compared to high-coverage assemblies by measures such as the frequencies with which they are reproduced in multiple libraries compared to exactly one library, finding no noticeable difference when coverage is low.

Chimeras of polymerases from different species could arise from PCR amplification or assembly. We used the UCHIME2 (usearch v8.0.1623) algorithm⁶⁶ to screen assembled palmprint sequences, finding no high-scoring putative chimeras. Mosaic sequences formed by joining a polymerase to unrelated sequence would either have an intact palmprint, in which case the mosaic would be irrelevant to our analysis, or would be rejected by Palmscan owing to the lack of delimiting motifs.

Reverse transcriptases are homologous to RdRP. Retroviral insertions into host genomes induce ubiquitous sequence similarity between host genomes and viral RdRP. Palmscan was designed to discriminate RdRP from sequences of reverse transcriptase origin. Testing on a large decoy set of non-RdRP sequences with recognizable sequence similarity showed that the Palmscan false discovery rate for RdRP identification is 0.001. We estimated the probability of false positive matches in unrelated sequence by generating sufficient random nucleotide and amino acid sequences to show that the expected number of false positive palmprint identifications is zero in a dataset of comparable size to our assemblies. We also regard the low observed frequency of palmprints in DNA whole-genome sequencing data (in 2.6 Pbp or 25.8% of reads, accounted for 100 known palmprints and 95 novel palmprints or 0.13% of the total identified) as a de facto confirmation of the low probability false positives in unrelated sequence.

Endogenous viral elements (EVEs; that is, insertions of viral sequence into host genomes that are potentially degraded and non-functional) cannot be distinguished from viral genomes on the basis of the palmprint sequence alone. To assess the frequency of EVEs in our data, we re-assembled 890 randomly chosen libraries yielding one or more palmprints using all reads, extracted the 23,530 resulting contigs with a positive palmprint hit by Palmscan, and classified them using Virosorter2 (v2.1)⁶⁷. Of these contigs, 11,914 were classified as viral, confirming the Palmscan identification; 49 as Viridiplantae (green plants); 46 as Metazoa; 25 as Fungi and the remainder were unclassified. Thus, $120/12,034 = 1\%$ of the classified contigs were predicted as non-viral, suggesting that the frequency of EVEs in the reported palmprints is around 1%.

Annotation of CoV assemblies

Accurate annotation of CoV genomes is challenging owing to ribosomal frameshifts and polyproteins that are cleaved into maturation proteins⁶⁸, and thus previously

annotated viral genomes offer a guide to accurate gene-calls and protein functional predictions. However, although many of the viral genomes we were likely to recover would be similar to previously annotated genomes in Refseq or GenBank, we anticipated that many of the genomes would be taxonomically distant from any available reference. To address these constraints, we developed an annotation pipeline called DARTH (version maul)⁶⁹ which leverages both reference-based and ab initio annotation approaches.

In brief, DARTH consists of the following phases: standardize the ordering and orientation of assembly contigs using conserved domain alignments, perform reference-based annotation of the contigs, annotate RNA secondary structure, ab initio gene-calling, generate files for aiding assembly and annotation diagnostics, and generate a master annotation file. It is important to put the contigs in the ‘expected’ orientation and ordering to facilitate comparative analysis of synteny and as a requirement for genome deposition. To perform this standardization, DARTH generates the six-frame translation of the contigs using the transeq (EMBOSS:6.6.0.0)⁷⁰ and uses HMMER3 (v3.3.2)⁷¹ to search the translations for Pfam domain models specific to CoV⁶⁴. DARTH compares the Pfam accessions from the HMMER alignment to the NCBI SARS-CoV-2 reference genome (NCBI Nucleotide accession NC_045512.2) to determine the correct ordering and orientation, and produces an updated assembly FASTA file. DARTH performs reference-based annotation using VADR (v1.1)⁷², which provides a set of genome models for all CoV RefSeq genomes⁷³. VADR provides annotations of gene coordinates, polyprotein cleavage sites, and functional annotation of all proteins. DARTH supplements the VADR annotation by using Infernal⁷⁴ to scan the contigs against the SARS-CoV-2 Rfam release⁷⁵ which provides updated models of CoV 50 and 30 untranslated regions (UTRs) along with stem-loop structures associated with programmed ribosomal frame-shifts. Although VADR provides reference-based gene-calling, DARTH also provides ab initio gene-calling by using FragGeneScan (v1.31)⁷⁶, a frameshift-aware gene caller. DARTH also generates auxiliary files that are useful for assembly quality and annotation diagnostics, such as indexed BAM files created with SAMtools (v1.7)⁷⁷ representing self-alignment of the trimmed reads to the canonicalized assembly using bowtie2⁵¹, and variant-calls using bcftools from SAMtools. DARTH generates these files so that they can be easily loaded into a genome browser such as JBrowse⁷⁸ or IGV⁷⁹. As the final step DARTH generates a single Generic Feature Format (GFF) 3.0 file⁸⁰ containing combined set of annotation information described above, ready for use in a genome browser, or for submitting the annotation and sequence to a genome repository.

Phage assembly

Each metagenomic dataset was individually de-novo-assembled using MEGAHIT (v1.2.9)⁸¹, and filtered to remove contigs smaller than 1 kb in size. ORFs were then predicted on all contigs using Prodigal (v2.6.3)⁸² with the following parameters: -m -p meta. Predicted ORFs were initially annotated using USEARCH⁵⁴ to search all predicted ORFs against UniProt⁸³, UniRef90 and KEGG⁸⁴. Sequencing coverage of each contig was calculated by mapping raw reads back to assemblies using bowtie2⁵¹. Terminase sequences from Al-Shayeb et al.⁴² were clustered at 90% amino acid identity to reduce redundancy using CD-HIT (v4.8.1)⁸⁵, and HMM models were built with hmmbuild (from the HMMER3 suite⁷¹) from the resulting set. Terminases in the assemblies from Serratus were identified using hmmsearch, retaining representatives from contigs greater than 140 kb in size. Some examples of prophage and large phages that did not co-cluster with the sequences from Al-Shayeb et al. were also recovered because they were also present in a sample that contained the expected large phages. The terminases were aligned using MAFFT (v7.407)⁸⁶ and filtered by TrimAL (v1.14)⁸⁷ to remove columns comprising more than 50% gaps, or 90% gaps, or using the automatic gappyout setting to retain the most conserved residues. Maximum likelihood trees were built from the resulting alignments using IQTREE (v1.6.6)⁸⁸.

Deploying the assembly and annotation workflow

The Serratus search for known or closely related viruses identified 37,131 libraries (14,304 by nucleotide and 23,898 by amino acid) as potentially positive for CoV (score ≥ 20 and ≥ 10 reads). To supplement this search we also used a recently developed index of the SRA called STAT¹⁶, which identified an additional 18,584 SRA datasets not in the defined SRA search space. The STAT BigQuery (accessed 24 June 2020) was: *WHERE tax id=11118 AND total count >1*.

We used AWS Batch to launch thousands of assemblies of NCBI accessions simultaneously. The workflow consists of four standard parts: a job queue, a job definition, a compute environment, and finally, the jobs themselves. A CloudFormation template (https://gitlab.pasteur.fr/rchikhi_pasteur/serratus-batch-assembly/-/blob/10934001/template/template.yaml) was created for building all parts of the cloud infrastructure from the command line. The job definition specifies a Docker image, and asks for 8 virtual CPUs (vCPUs, corresponding to threads) and 60 GB of memory per job, corresponding to a reasonable allocation for coronaSPAdes. The compute environment is the most involved component. We set it to run jobs on cost-effective Spot instances (optimal setting) with an additional cost-optimization strategy (SPOT_CAPACITY_OPTIMIZED setting), and allowing up to 40,000 vCPUs total. In addition, the compute environment specifies a launch template which, on each instance, (i) automatically mounts an exclusive 1 TB EBS volume, allowing sufficient

disk space for several concurrent assemblies, and (ii) downloads the 5.4 GB CheckV (v0.6.0)⁸⁹ database, to avoid bloating the Docker image.

The peak AWS usage of our Batch infrastructure was around 28,000 vCPUs, performing around 3,500 assemblies simultaneously. A total of 46,861 accessions out of 55,715 were assembled in a single day. They were then analysed by two methods to detect putative CoV contigs. The first method is CheckV⁸⁹, followed selecting contigs associated to known CoV genomes. The second method is a custom script (https://gitlab.pasteur.fr/rchikhi_pasteur/serratus-batch-assembly/-/blob/10934001/stats/bgc_parse_and_extract.py) that parses coronaSPAdes BGC candidates and keeps contigs containing CoV domain(s). For each accession, we kept the set of contigs obtained by the first method (CheckV) if it is non-empty, and otherwise we kept the set of contigs from the second method (BGC).

A majority (76%) of the assemblies were discarded for one of the following reasons: (i) no CoV contigs were found by either filtering method; (ii) reads were too short to be assembled; (iii) Batch job or SRA download failed; or (iv) coronaSPAdes ran out of memory. A total of 11,120 assemblies were considered for further analysis.

The average cost of assembly was between US\$0.30 and US\$0.40 per library, varying depending on library type (RNA-seq versus metagenomic). This places an estimate of 46–95-fold higher cost for assembly alone compared to a cost of US\$0.0042 or US\$0.0065 for an alignment-based search.

Taxonomic and phylogenetic analyses

Taxonomy prediction for coronavirus genomes

We developed a module, SerraTax, to predict taxonomy for CoV genomes and

assemblies (<https://github.com/ababaian/serratus/tree/master/containers/serratax>). SerraTax was designed with the following requirements in mind: provide taxonomy predictions for fragmented and partial assemblies in addition to complete genomes; report best-estimate predictions balancing over-classification and under-classification (too many and too few ranks, respectively); and assign an NCBI Taxonomy Database⁹⁰ identifier (TaxID).

Assigning a best-fit TaxID was not supported by any previously published taxonomy prediction software to the best of our knowledge; this requires assignment to intermediate ranks such as sub-genus and ranks below species (commonly called strains, but these ranks are not named in the Taxonomy database), and to unclassified taxa, for example, TaxID 2724161, unclassified Buldecovirus, in cases in which the genome is predicted to fall inside a named clade but outside all named taxa within that clade.

SerraTax uses a reference database containing domain sequences with TaxIDs. This database was constructed as follows. Records annotated as CoV were downloaded from UniProt⁸³, and chain sequences were extracted. Each chain name, for example Helicase, was considered to be a separate domain. Chains were aligned to all complete coronavirus genomes in GenBank using UBLAST (usearch: v11.0.667)⁵⁴ to expand the repertoire of domain sequences. The reference sequences were clustered using UCLUST⁵⁴ at 97% sequence identity to reduce redundancy.

For a given query genome, ORFs are extracted using the getorf (EMBOSS:6.6.0) software⁷⁰. ORFs are aligned to the domain references and the top 16 reference sequences for each domain are combined with the best-matching query ORF. For each domain, a multiple alignment of the top 16 matches plus query ORF is constructed on the fly by MUSCLE

(v3.8.31⁹¹) and a neighbour-joining tree is inferred from the alignment, also using MUSCLE. Finally, a consensus prediction is derived from the placement of the ORF in the domain trees. Thus, the presence of a single domain in the assembly suffices to enable a prediction; if more domains are present they are combined into a consensus.

Taxonomic assignment by phylogenetic placement

To generate an alternate taxonomic annotation of an assembled genome, we created a pipeline based on phylogenetic placement, SerraPlace.

To perform phylogenetic placement, a reference phylogenetic tree is required. To this end, we collected 823 reference amino acid RdRP sequences, spanning all *Coronaviridae*. To this set we added an outgroup RdRP sequence from the Torovirus family (NC 007447). We clustered the sequences to 99% identity using USEARCH (ref. ⁵⁴, UCLUST algorithm, v11.0.667), resulting in 546 centroid sequences. Subsequently, we performed multiple sequence alignment on the clustered sequences using MUSCLE. We then performed maximum likelihood tree inference using RAxML-NG (ref. ⁹², ‘PROTGTR+FO+G4’, v0.9.0), resulting in our reference tree.

To apply SerraPlace to a given genome, we first use HMMER (ref. ⁷¹, v3.3) to generate a reference HMM, based on the reference alignment. We then split each contig into ORFs using esl-translate, and use hmmsearch (P value cut-off 0.01) and seqtk (commit 7c04ce7) to identify those query ORFs that align with sufficient quality to the previously generated reference HMM. All ORFs that pass this test are considered valid input sequences for phylogenetic placement. This produces a set of likely placement locations on the tree, with an associated likelihood weight. We then use Gappa (v0.6.1, ⁹³) to assign taxonomic information to each query, using the taxonomic information for the reference sequences. Gappa assigns taxonomy by first labelling the interior nodes of the reference tree by a consensus of the taxonomic labels of all descendant leaves of that node. If 66% of leaves share the same taxonomic label up to some level, then the internal node is assigned that label. Then, the likelihood weight associated

with each sequence is assigned to the labels of internal nodes of the reference tree, according to where the query was placed.

From this result, we select that taxonomic label that accumulated the highest total likelihood weight as the taxonomic label of a sequence. Note that multiple ORFs of the same genome may result in a taxonomic label, in which case, we select the longest sequence as the source of the taxonomic assignment of the genome.

Phylogenetic inference

We performed phylogenetic inferences using a custom snakemake (v6.6.0) pipeline (available at <https://github.com/lczech/nidhoggr>), using ParGenes (v1.1.2)⁹⁴. ParGenes is a tree search orchestrator, combining ModelTestNG (v0.1.3)⁹⁵ and RAxML-NG, and enabling higher levels of parallelization for a given tree search.

To infer the maximum likelihood phylogenetic trees, we performed a tree search comprising 100 distinct starting trees (50 random, 50 parsimony), as well as 1,000 bootstrap searches. We used ModelTest-NG to automatically select the best evolutionary model for the given data. The pipeline also automatically produces versions of the best maximum likelihood tree annotated with Felsenstein's Bootstrap⁹⁶ support values, and Transfer Bootstrap Expectation values⁹⁷.

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

All Serratus data, raw and processed, are released into the public domain immediately in accordance with the Bermuda Principles and freely available at <https://serratus.io/access>. Assembled genomes for this study are available on GenBank under project [PRJEB44047](#).

Code availability

Serratus (v0.3.0) is available at <https://github.com/ababaian/serratus>. Archival copies of all code and software generated for this study are freely available at <https://github.com/serratus-bio>. Electronic notebooks for experiments are available at <https://github.com/ababaian/serratus>.

References

1. 1.
Leinonen, R., Sugawara, H. & Shumway, M. The Sequence Read Archive. *Nucleic Acids Res.* **39**, D19–D21 (2011).
2. 2.
Anthony, S. J. et al. A strategy to estimate unknown viral diversity in mammals. *mBio* **4**, e00598-13 (2013).
3. 3.
Johnson, C. K. et al. Global shifts in mammalian population trends reveal key predictors of virus spillover risk. *Proc. R. Soc. B* **287**, 20192736 (2020).
4. 4.
Carroll, D. et al. The Global Virome Project. *Science* **359**, 872–874 (2018).
5. 5.
Shi, M. et al. The evolutionary history of vertebrate RNA viruses. *Nature* **556**, 197–202 (2018).
6. 6.

Wahba, L. et al. An extensive meta-metagenomic search identifies SARS-CoV-2-homologous sequences in pangolin lung viromes. *mSphere* **5**, 00160-20 (2020).

7. 7.

Wolf, Y. I. et al. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat. Microbiol.* **5**, 1262–1270 (2020).

8. 8.

Mitchell, A. L. et al. MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Res.* **48**, D570–D578 (2020).

9. 9.

Chen, I.-M. A. et al. The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* **49**, D751–D763 (2021).

10. 10.

Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D. Massive expansion of human gut bacteriophage diversity. *Cell* **184**, 1098–1109 (2021).

11. 11.

Nayfach, S. et al. A genomic catalog of Earth’s microbiomes. *Nat. Biotechnol.* **39**, 499–509 (2021).

12. 12.

Moore, R. A. et al. The sensitivity of massively parallel sequencing for detecting candidate infectious agents associated with human tissue. *PLoS One* **6**, e19838 (2011).

13. 13.

NIH. *STRIDES Initiative—Data Science at NIH*
<https://datascience.nih.gov/strides> (2021).

14. 14.

Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).

15. 15.

Karasikov, M. et al. MetaGraph: indexing and analysing nucleotide archives at petabase-scale. Preprint at
<https://www.biorxiv.org/content/10.1101/2020.10.01.322164v2> (2020).

16. 16.

Katz, K. S. et al. STAT: a fast, scalable, MinHash-based k-mer tool to assess Sequence Read Archive next-generation sequence submissions. *Genome Biol.* **22**, 270 (2021).

17. 17.

Koonin, E. V. & Dolja, V. V. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiol. Mol. Biol. Rev.* **78**, 278–303 (2014).

18. 18.

Babaian, A. & Edgar, R. C. Ribovirus classification by a polymerase barcode sequence. Preprint at
<https://www.biorxiv.org/content/10.1101/2021.03.02.433648v1> (2021).

19. 19.

Wolf, Y. I. et al. Origins and evolution of the global RNA virome. *mBio* **9**, e0239-18 (2018).

20. 20.

Obbard, D. J., Shi, M., Roberts, K. E., Longdon, B. & Dennis, A. B. A new lineage of segmented RNA viruses infecting animals. *Virus Evol.* **6**, vez061 (2020).

21. 21.

Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

22. 22.

Koonin, E. V. et al. Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* **84**, e00061-19 (2020).

23. 23.

Bukhari, K. et al. Description and initial characterization of metatranscriptomic nidovirus-like genomes from the proposed new family Abyssoviridae, and from a sister group to the *Coronavirinae*, the proposed genus Alphaletoivirus. *Virology* **524**, 160–171. (2018).

24. 24.

Mordecai, G. J. et al. Endangered wild salmon infected by newly discovered viruses. *eLife* **8**, e47615 (2019).

25. 25.

Meleshko, D., Hajirasouliha, I. & Korobeynikov, A. coronaSPAdes: from biosynthetic gene clusters to RNA viral assemblies. *Bioinformatics* **38**, 1–8 (2022).

26. 26.

Tao, Y. et al. Broad-range virus detection and discovery using microfluidic PCR coupled with high-throughput sequencing. Preprint at <https://www.biorxiv.org/content/10.1101/2020.06.10.145052v1> (2020).

27. 27.

Tsai, S. L., Baselga-Garriga, C. & Melton, D. A. Midkine is a dual regulator of wound epidermis development and inflammation during the initiation of limb regeneration. *eLife* **9**, e50765 (2020).

28. 28.

Sabin, K. Z., Jiang, P., Gearhart, M. D., Stewart, R. & Echeverri, K. AP-1 cFos/JunB /miR-200a regulate the pro-regenerative glial cell response during axolotl spinal cord regeneration. *Commun. Biol.* **2**, 91 (2019).

29. 29.

Miller, A. K. et al. Slippery when wet: cross-species transmission of divergent coronaviruses in bony and jawless fish and the evolutionary history of the Coronaviridae. *Virus Evol.* **7**, veab050 (2021).

30. 30.

Mukherjee, S. et al. Genomes OnLine Database (GOLD) v.8: overview and updates. *Nucleic Acids Res.* **49**, D723–D733 (2021).

31. 31.

Stanaway, J. D. et al. The global burden of viral hepatitis from 1990 to 2013: findings from the Global Burden of Disease Study 2013. *Lancet* **388**, 1081–1088 (2016).

32. 32.

Taylor, J. M. Infection by hepatitis delta virus. *Viruses* **12**, 648 (2020).

33. 33.

Szirovicza, L. et al. Snake deltavirus utilizes envelope proteins of different viruses to generate infectious particles. *mBio* **11**, e03250-19 (2020).

34. 34.

Wille, M. et al. A divergent hepatitis D-like agent in birds. *Viruses* **12**, 720 (2018).

35. 35.

Chang, W.-S. et al. Novel hepatitis D-like agents in vertebrates and invertebrates. *Virus Evol.* **5**, vez021 (2019).

36. 36.

Paraskevopoulou, S. et al. Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent *Proechimys semispinosus*. *Proc. Natl Acad. Sci. USA* **117**, 17977–17983 (2020).

37. 37.

Bergner, L. M. et al. Diversification of mammalian deltaviruses by host shifting. *Proc. Natl Acad. Sci. USA* **118**, e2019907118 (2021).

38. 38.

Iwamoto, M. et al. Identification of novel avian and mammalian deltaviruses provides new insights into deltavirus evolution. *Virus Evol.* **7**, veab003 (2021).

39. 39.

De la Peña, M., Ceprián, R., Casey, J. L. & Cervera, A. Hepatitis delta virus-like circular RNAs from diverse metazoans encode conserved hammerhead ribozymes. *Virus Evol.* **7**, veab016 (2021).

40. 40.

Zuccola, H. J., Rozzelle, J. E., Lemon, S. M., Erickson, B. W. & Hogle, J. M. Structural basis of the oligomerization of hepatitis delta antigen. *Structure* **6**, 821–830 (1998).

41. 41.

Flores, R., Gago-Zachert, S., Serra, P., Sanjuán, R. & Elena, S. F. Viroids: survivors from the RNA world? *Annu. Rev. Microbiol.* **68**, 395–414 (2014).

42. 42.

Al-Shayeb, B. et al. Clades of huge phages from across Earth's ecosystems. *Nature* **578**, 425–431 (2020).

43. 43.

Devoto, A. E. et al. Megaphages infect *Prevotella* and variants are widespread in gut microbiomes. *Nat. Microbiol.* **4**, 693–700 (2019).

44. 44.

Bradley, P., Den Bakker, H. C., Rocha, E. P. C., McVean, G. & Iqbal, Z. Ultra-fast search of all deposited bacterial and viral genomic data. *Nat. Biotechnol.* **37**, 152–159 (2019).

45. 45.

Baker, D. et al. No more business as usual: agile and effective responses to emerging pathogen threats require open data and open analytics. *PLoS Pathog.* **16**, e1008643 (2020).

46. 46.

Letko, M., Seifert, S. N., Olival, K. J., Plowright, R. K. & Munster, V. J. Bat-borne virus diversity, spillover and emergence. *Nat. Rev. Microbiol.* **18**, 461–471 (2020).

47. 47.

Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**, 562–569 (2020).

48. 48.

Chase, J. M., Blowes, S. A., Knight, T. M., Gerstner, K. & May, F. Ecosystem decay exacerbates biodiversity loss with habitat loss. *Nature* **584**, 238–243 (2020).

49. 49.

Thompson, A. A. & Peersen, O. B. Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. *EMBO J.* **23**, 3462–3471 (2004).

50. 50.

Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

51. 51.

Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

52. 52.

Schatz, M. C. CloudBurst: highly sensitive read mapping with MapReduce. *Bioinformatics* **25**, 1363–1369 (2009).

53. 53.

NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018).

54. 54.

Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

55. 55.

Morgulis, A., Gertz, E. M., Schäffer, A. A. & Agarwala, R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol.* **13**, 1028–1040 (2006).

56. 56.

Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962 (2016).

57. 57.

Courtot, M., Gupta, D., Liyanage, I., Xu, F. & Burdett, T. BioSamples database: FAIRer samples metadata to accelerate research data management. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkab1046> (2021).

58. 58.

Levi, K., Rynge, M., Abeysinghe, E. & Edwards, R. A. Searching the Sequence Read Archive using Jetstream and Wrangler. In *Proc. Practice and Experience on Advanced Research Computing* 1–7 (Association for Computing Machinery, 2021).

59. 59.

Hunt, M. et al. IVA: accurate de novo assembly of RNA virus genomes. *Bioinformatics* **31**, 2374–2376 (2015).

60. 60.

Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).

61. 61.

Bushmanova, E., Antipov, D., Lapidus, A. & Prjibelski, A. D. rnaSPAdes: a de novo transcriptome assembler and its application to

RNA-seq data. *GigaScience* **8**, giz100 (2019).

62. 62.

Antipov, D., Raiko, M., Lapidus, A. & Pevzner, P. A. metaviralSPAdes: assembly of viruses from metagenomic data. *Bioinformatics* **36**, 4126–4129 (2020).

63. 63.

Meleshko, D. et al. BiosyntheticSPAdes: reconstructing biosynthetic gene clusters from assembly graphs. *Genome Res.* **29**, 1352–1362 (2019).

64. 64.

Pfam team. *Pfam SARS-CoV-2 Special Update (part 2)* <https://xfam.wordpress.com/2020/04/06/pfam-sars-cov-2-special-update-part-2/> (2020).

65. 65.

Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).

66. 66.

Edgar, R. C. UCHIME2: improved chimera prediction for amplicon sequencing. Preprint at <https://doi.org/10.1101/074252> (2016).

67. 67.

Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).

68. 68.

Thiel, V. et al. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**, 2305–2315 (2003).

69. 69.

Altman, T. *DARTH Coronavirus Annotation Pipeline*
<https://bitbucket.org/tomeraltman/DARTH/src/master/> (2020).

70. 70.

Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).

71. 71.

Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).

72. 72.

Schäffer, A. A. et al. VADR: validation and annotation of virus sequence submissions to GenBank. *BMC Bioinformatics* **21**, 211 (2020).

73. 73.

Nawrocki, E. *Coronavirus Annotation using VADR*
<https://github.com/nawrockie/VADR/wiki/Coronavirus-annotation#build> (2020).

74. 74.

Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).

75. 75.

Rfam team. *Rfam Coronavirus Special Release*
<https://xfam.wordpress.com/2020/04/27/rfam-coronavirus-release/>

(2020).

76. 76.

Rho, M., Tang, H. & Ye, Y. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* **38**, e191 (2010).

77. 77.

Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

78. 78.

Buels, R. et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* **17**, 66 (2016).

79. 79.

Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant review with the Integrative Genomics Viewer. *Cancer Res.* **77**, e31–e34 (2017).

80. 80.

Eilbeck, K. et al. The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.* **6**, R44 (2005).

81. 81.

Li, D. et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**, 3–11 (2016).

82. 82.

Hyatt, D., LoCascio, P. F., Hauser, L. J. & Uberbacher, E. C. Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **28**, 2223–2230 (2012).

83. 83.

The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

84. 84.

Altman, T., Travers, M., Kothari, A., Caspi, R. & Karp, P. D. A systematic comparison of the MetaCyc and KEGG pathway databases. *BMC Bioinformatics* **14**, 112 (2013).

85. 85.

Li, W., Fu, L., Niu, B., Wu, S. & Wooley, J. Ultrafast clustering algorithms for metagenomic sequence analysis. *Brief. Bioinformatics* **13**, 656–668 (2012).

86. 86.

Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**, 2490–2492 (2018).

87. 87.

Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

88. 88.

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).

89. 89.

Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2020).

90. 90.

Schoch, C. L. et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* **2020**, baaa062 (2020).

91. 91.

Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

92. 92.

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).

93. 93.

Czech, L., Barbera, P. & Stamatakis, A. Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics* **36**, 3263–3265 (2020).

94. 94.

Morel, B., Kozlov, A. M. & Stamatakis, A. ParGenes: a tool for massively parallel model selection and phylogenetic tree inference on thousands of genes. *Bioinformatics* **35**, 1771–1773 (2018).

95. 95.

Darriba, D. et al. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **37**, 291–294 (2019).

96. 96.

Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791 (1985).

97. 97.

Lemoine, F. et al. Renewing Felsenstein’s phylogenetic bootstrap in the era of big data. *Nature* **556**, 452–456 (2018).

98. 98.

Crisci, M. A. et al. Wide distribution of alternatively coded Lak megaphages in animal microbiomes. Preprint at <https://www.biorxiv.org/content/10.1101/2021.01.08.425732v1> (2021).

99. 99.

Rapid reconstitution of the fecal microbiome after extended diet-induced changes indicates a stable gut microbiome in healthy adult dogs. *Appl. Environ. Microbiol.* **86**, e00562-20 (2020).

Acknowledgements

The Serratus project is an initiative of the hackseqRNA genomics hackathon (<https://www.hackseq.com>). We thank the many contributors for code snippets and bioinformatic discussion (E. Erhan, J. Chu, S. Jackman, I. Birol, K. Wellman, O. Fornes, C. Xu, M. Huss, K. Ha, M. Krzywinski, E. Nawrocki, R. McLaughlin, C. Morgan-Lang, C. Blumberg and the J. Brister laboratory); A. Rodrigues, S. McMillan, V. Wu, C. Kennett, K. Chao, and N. Pereyaslavsky for AWS support; the J. Joy laboratory, G. Mordecai, J. Taylor, S. Roux, N. Kyrpides, E. Jan, T. Reddy, L. Bergner, R. Orton and D. Streicker for virology discussions; and H.-G. Drost and D. Weigel for supporting the adoption of DIAMOND v2 for Serratus protein alignments as part of an extended feature request. We are grateful to the entire team managing the NCBI SRA and the biology community for data sharing, with particular thanks to the E. Brodie, E. Lilleskov and E. Young laboratories.

T.A. thanks Advanced Research Computing resource at the University of British Columbia and B.B. thanks the Max Plank Society for financial support. P.B. was financially supported by the Klaus Tschira Foundation; R.C. by ANR Transipedia, Inception and PRAIRIE grants (PIA/ANR16-CONV-0005, ANR-18-CE45-0020, ANR-19-P3IA-0001); and M.d.l.P. by the Ministerio de Economía y Competitividad of Spain and FEDER grants (BFU2017-87370-P and PID2020-116008GB-I00). A.K. and D.M. were supported by the Russian Science Foundation (grant 19-14-00172) and computation was carried out in part by Resource Centre ‘Computer Centre of SPbU’. A.K. and D.M. are grateful to Saint Petersburg State University for the overall support of this work. Project support and computing resources were provided by the University of British Columbia Community Health and Wellbeing Cloud Innovation Centre, powered by AWS.

Author information

Author notes

1. These authors contributed equally: Robert C. Edgar, Jeff Taylor, Victor Lin, Tomer Altman, Pierre Barbera, Dmitry Meleshko, Dan Lohr, Gherman Novakovsky, Benjamin Buchfink, Basem Al-Shayeb, Jillian F. Banfield, Marcos de la Peña, Anton Korobeynikov, Rayan Chikhi, Artem Babaian

Affiliations

1. Independent researcher, Corte Madera, CA, USA

Robert C. Edgar

2. Independent researcher, Vancouver, British Columbia, Canada

Jeff Taylor & Artem Babaian

3. Independent researcher, Seattle, WA, USA

Victor Lin

4. Altman Analytics, San Francisco, CA, USA

Tomer Altman

5. Computational Molecular Evolution Group, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

Pierre Barbera

6. Center for Algorithmic Biotechnology, St Petersburg State University, St Petersburg, Russia

Dmitry Meleshko & Anton Korobeynikov

7. Tri-Institutional PhD Program in Computational Biology and Medicine, Weill Cornell Medical College, New York, NY, USA

Dmitry Meleshko

8. Unaffiliated, Atlanta, GA, USA

Dan Lohr

9. Bioinformatics Graduate Program, University of British Columbia, Vancouver, British Columbia, Canada

Gherman Novakovsky

10. Computational Biology Group, Max Planck Institute for Biology, Tübingen, Germany

Benjamin Buchfink

11. Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA

Basem Al-Shayeb

12. Department of Earth and Planetary Science, University of California,
Berkeley, Berkeley, CA, USA

Jillian F. Banfield

13. Instituto de Biología Molecular y Celular de Plantas, Universidad
Politécnica de Valencia–CSIC, Valencia, Spain

Marcos de la Peña

14. Department of Statistical Modelling, St Petersburg State University, St
Petersburg, Russia

Anton Korobeynikov

15. G5 Sequence Bioinformatics, Department of Computational Biology,
Institut Pasteur, Paris, France

Rayan Chikhi

Contributions

All authors contributed equally to this work. A.B. conceived and led the study. A.B. and J.T. designed and implemented the Serratus architecture. A.B. and R.C.E. constructed the virus pangenes and RdRP query. R.C.E. developed the SerraTax and Summarizer modules. P.B. developed the SerraPlace tree placement and taxonomy prediction code and calculated maximum likelihood trees. T.A. developed the DARTH annotation pipeline and submitted the annotated genomes to ENA. D.M. and A.K. developed the coronaSPAdes assembler. R.C. implemented the assembly pipeline, and deployed the assembly and annotation pipeline. B.B. optimized the DIAMOND algorithm for RdRP search. A.B., V.L. and D.L. designed and developed <https://serratus.io> and the SQL server. A.B. and G.N. developed the Tantalus R package. A.B., R.C.E., T.A., P.B., D.M., M.d.l.P., A.K. and R.C. analysed the coronavirus, RdRP and delta virus data. B.A.-S. and J.F.B. designed the phage panproteome, assembled phage genomes and conducted phage phylogenetic analyses. All authors contributed to data interpretation and writing the manuscript.

Corresponding author

Correspondence to [Artem Babaian](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks C. Titus Brown, Alice McHardy 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 Overview of the Serratus infrastructure.](#)

a Schematic and data workflow (**b**) as described in the methods for sequence alignment. **c** The align module accepts either a nucleotide or protein sequence query. **d** A nucleotide alignment completion rate for Serratus shows stable and linear performance to complete 1.29 million SRA accessions in a 24-hour period and the **e** cost breakdown for this run. Compute costs between modules are an approximate comparison of CPU requirements of each step. The total average cost per completed SRA accession was US\$0.0062 for nucleotide search or US\$0.0042 for translated-nucleotide search. **f** Tukey boxplot of biological cross-validation to measure alignment sensitivity for bowtie2 (nucleotide search),

DIAMOND¹⁴ (translated nucleotide search) or 32-mer for exact search. In brief, two RdRP sequences sharing the nominal amino acid identities form a “pair”. 100 bp reads were simulated from the coding sequence of one pair and mapped onto the second pair, with the fraction of reads mapped reported. A fraction of 0.5 indicates that half the simulated reads at the given RdRP percent identity are mappable and thus detectable (see Methods). For each of the 12 percent identity categories, n = 10 biologically independent RdRP pairs were analysed.

Extended Data Fig. 2 Analysis of palmprint contigs recovered by Serratus.

a Length distribution of amino acid sequences in the rdp1 query (upper histogram) and microassembled contigs (lower histogram, length=nucleotides/3). **b** Distribution of Palmscan confidence scores. **c** Observations of the 10 most frequent “super-motifs” (six well-conserved residues marked with asterisk) reported by Palmscan. **d** Kernel distribution and mean (white cross) of coverage vs. abundance (number of runs where a given palmprint is observed), showing that palmprints have similar underlying coverage distributions at all abundances. **e** Preston plot of distinct palmprints vs. abundance exhibiting similar, approximately log-log-linear relationships to totals at end-of-year 2015 to 2019 and final totals at approx. end of 2020 (all). **f** Preston plot of number of distinct palmprints observed in a given run vs. number of runs with 95% confidence interval. **g** Numbers of singletons and second observations (confirmations) at the end of each year showing that the growth in singletons is matched by a comparable growth in confirmations. **h** Kingdom predicted by Virsorter2 for RdRP+ contigs (by Palmscan) obtained by full assembly of 880 randomly chosen RdRP+ runs. **i** Number of palmprints in each phylum assigned by taxonomy (known) or predicted (novel). **j** Number of OTUs as a function of clustering identity.

Extended Data Fig. 3 Distribution of select RNA virus families.

a Histograms of datasets matching select RNA viral family by translated-nucleotide search against RdRP query, binned by the average amino acid identity. Score (gradient colouring) function approximates pangenome/gene

coverage (see methods) used for manual inspection and to prioritize assembly. Interactive and queryable versions of these plots for extended virus families are available at <https://serratus.io/explorer>. **b** Relationship between the nucleotide pangenome score function and the subsequent assembly success (defined by the presence of an RdRP+ contig) measured from 52,772 libraries with reads aligning to *Coronaviridae*. **c** Histogram of all detected sOTUs classified to *Riboviria* order (>40% amino acid identity to a named species) with unclassified sOTUs not shown. Segmented bars (left) show the fraction of sOTUs with similarity to known sOTU, binned into intervals 90+ (>=90%, ~species), 75+ (75% to 90%, ~genus), 50+ (50% to 75%, ~family), and <50% (~novel family). Complete multiple sequence alignments and tree files for per-order and per-family trees is available at <https://serratus.io/trees>.

Extended Data Fig. 4 Genome organization of Coronaviridae and neighbours.

a Length distribution for 11,120 assembled contigs classified as CoV-positive, showing a peak around the typical CoV genome length, 4,179 (37.58%) of contigs also contained a match for RdRP. **b** Phylogram shown in Figure 3 showing the *Mesoniviridae*, *Tobaniviridae*, and *Roniviridae* outgroups. **c** Triangular matrix showing median RdRP sequence identities between selected *Nidovirales* and group-E sequences. **d** Hidden Markov Model (HMM) protein domain matches from the RdRp in exemplar sequences (contigs or GenBank sequences), grouped by genus. Novel sOTUs identified in this analysis indicated by a coloured circle.

Extended Data Fig. 5 Newly characterized delta virus and delta-virus-like genomes.

Structure and organization of selected examples from the 14 delta virus-, 39 epsilon virus- and 311 zeta virus-like genomes identified in our study. **a** Similar to human delta virus (HDV), delta virus-like genomes from vertebrates (PmacDV SRR7910143; MmonDV SRR2136906; TgutDV SRR5001850; IchiDV SRR8954566 and BblaDV SRR8242383) and environmental datasets (SRR7286070 and SRR6943136) share similar

predicted stable rod-like folding, a predicted ORF coding for the delta antigen (δ Ag) and a delta ribozyme (dvrbz) on each polarity. Folding of the circular DNA virus Porcine Circovirus 2 (PCV2) and a shuffled MmonDV sequence are shown as negative controls. **b** Epsilon virus-like genomes detected in invertebrates (SulaEV SRR8739608; GsulEV SRR7170939 and BaerEV SRR12300397) and environmental datasets (SRR8840728 and SRR6943136) show similar structure and organization to delta viruses, with one or two predicted ORFs (epsilon antigen or Ag) and two hammerhead ribozymes (hhrbz) in equivalent genomic regions. **c** Zeta virus-like genomes detected in invertebrate (*Ocassitermes sp.* ZVs SRR8924823) and environmental datasets (SRR7286070, SRR6943136, SRR8840728, SRR6201737, SRR5864109 and SRR12063536) are smaller than delta and epsilon agents. Up to 90% of the zeta genomes have sizes multiple of 3 and predicted ORFs without stop codons, capable to encode endless tandem-repeated zeta antigens in both polarities (ζ Ag+ and ζ Ag- shown as yellow and red arrows, respectively). Both genomic zeta polarities keep hhrbz (shown as arrows overlapping the ORFs) similar to the epsilon ribozymes (Extended Fig 6). Larger zeta virus-like genomes (>651 nt) were less abundant (7% of all zeta genomes) and frequently show stop codons, or their sizes are not multiple of 3.

Extended Data Fig. 6 Evolutionary history of delta-virus-like agents.

a Consensus structures (weighted nucleotide conservation threshold of 90%) of delta virus ribozymes, including the 14 genomes described in this work. **b** Consensus structures of the two hammerhead ribozyme families (type III and extended-type III³⁹) detected in epsilon and zeta agents. Most positions of epsilon and zeta motifs are sequence conserved for each ribozyme family. **c** MSA of the predicted antigen (N-term domain) from delta and epsilon agents (genomes detected in this study are indicated with a red asterisk). The antiparallel coiled-coil of the HDV is delimited with a grey box, and conserved residues involved in hydrophobic interactions are shown at the bottom⁴⁰, supporting a highly divergent connection between delta and epsilon genomes. **d** Human HDV delta virus is known to contain a viroid-like domain related to the *Pospiviroidae* family of plant viroids. Both families of agents conserve a tertiary structure reminiscent of the E-loop 5S

rRNA (nucleotides in green) and are replicated by the RNA Pol II of the host⁴¹. *Pospiviroids*, despite lacking hhrbz, share with zeta genomes a small rod structure, and in some cases, the presence of predicted endless tandem-repeat ORFs, most notably in both polarities of numerous variants of the Hop Stunt Viroid (HSVd). Whereas viroids have been historically regarded as non-protein-coding RNAs, our reported observations warrant further investigation.

Extended Data Fig. 7 Huge phage and Lak phage detail.

Expanded view of maximum likelihood terminase large subunit protein phylogenetic trees for **(a)** the expansion of the Kabirphage clade by newly recovered sequences from different animal types (coloured dots). Red branches are public data recovered by Serratus, black branches indicate the previously reported genomes from⁴². **b** Publicly available Lak phage genomes⁹⁸ with sequences of two newly reconstructed complete Lak megaphage genomes. These are the first reported Lak megaphages from dogs (assembled from faecal sample metagenome reads from Allaway et al.⁹⁹). The genomes have identical terminase sequences (at the nucleotide level) although the dogs were in different housing areas and were sampled at different times (D. Allaway, personal communication).

Supplementary information

Supplementary Table 1

SRA run queries and CoV assembly table: Queries and accessions from this study. **a** SRA queries to retrieve collections of datasets. **b** Run accessions, assembly statistics and select metadata for the 11,120 runs for which *Coronaviridae*, or *Coronaviridae*-like sequences were assembled. **c** Assignment of assembled runs to operational taxonomic units (OTUs) based on 97% nucleotide identity of the RNA dependent RNA polymerase (RdRP) domain. **d** Assignment of GenBank records to RdRP OTUs. **e** Sub-palmprint RdRP fragments which are possibly novel Coronaviruses (50–90% amino acid identity to a known CoV). Non-overlapping fragments from the same library, or identical overlapping fragments from different

libraries are estimated to be the same virus. **f** Assignment of expected viral host for GenBank records based on Sequence Read Archive and JGI GOLD metadata^{1,30}. **g** Taxonomic source for RdRP containing assemblies.

Reporting Summary

Peer Review File

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Edgar, R.C., Taylor, J., Lin, V. *et al.* Petabase-scale sequence alignment catalyses viral discovery. *Nature* **602**, 142–147 (2022).
<https://doi.org/10.1038/s41586-021-04332-2>

- Received: 10 August 2020
- Accepted: 10 December 2021
- Published: 26 January 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04332-2>

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/s41586-021-04332-2>

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

- Article
- Open Access
- [Published: 07 December 2021](#)

Signature of long-lived memory CD8⁺ T cells in acute SARS-CoV-2 infection

- [Sarah Adamo](#) [ORCID: orcid.org/0000-0002-6161-3156](#)¹,
- [Jan Michler](#)²,
- [Yves Zurbuchen](#) [ORCID: orcid.org/0000-0001-5387-9950](#)¹,
- [Carlo Cervia](#) [ORCID: orcid.org/0000-0001-7120-8739](#)¹,
- [Patrick Taeschler](#)¹,
- [Miro E. Raeber](#) [ORCID: orcid.org/0000-0003-2609-0246](#)¹,
- [Simona Baghai Sain](#)²,
- [Jakob Nilsson](#)¹,
- [Andreas E. Moor](#) [ORCID: orcid.org/0000-0001-8715-8449](#)² &
- [Onur Boyman](#) [ORCID: orcid.org/0000-0001-8279-5545](#)^{1,3}

Nature volume 602, pages 148–155 (2022)

- 29k Accesses
- 1 Citations
- 938 Altmetric
- [Metrics details](#)

Subjects

- [Immunological memory](#)
- [Infection](#)

Abstract

Immunological memory is a hallmark of adaptive immunity and facilitates an accelerated and enhanced immune response upon re-infection with the same

pathogen^{1,2}. Since the outbreak of the ongoing COVID-19 pandemic, a key question has focused on which SARS-CoV-2-specific T cells stimulated during acute infection give rise to long-lived memory T cells³. Here, using spectral flow cytometry combined with cellular indexing of transcriptomes and T cell receptor sequencing, we longitudinally characterized individual SARS-CoV-2-specific CD8⁺ T cells of patients with COVID-19 from acute infection to 1 year into recovery and found a distinct signature identifying long-lived memory CD8⁺ T cells. SARS-CoV-2-specific memory CD8⁺ T cells persisting 1 year after acute infection express CD45RA, IL-7 receptor- α and T cell factor 1, but they maintain low expression of CCR7, thus resembling CD45RA⁺ effector memory T cells. Tracking individual clones of SARS-CoV-2-specific CD8⁺ T cells, we reveal that an interferon signature marks clones that give rise to long-lived cells, whereas prolonged proliferation and mechanistic target of rapamycin signalling are associated with clonal disappearance from the blood. Collectively, we describe a transcriptional signature that marks long-lived, circulating human memory CD8⁺ T cells following an acute viral infection.

[Download PDF](#)

Main

The coronavirus disease 2019 (COVID-19) pandemic has taken an extraordinary toll on global health and economy, affecting billions of lives all over the world. The ongoing vaccination efforts appear to curtail the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and prevent severe disease, even as new virus variants emerge^{4,5}. Yet, prevailing questions concern whether and how exposure to SARS-CoV-2 by infection or immunization might result in long-term protective immunity.

On encountering their cognate antigen on antigen-presenting cells, antigen-specific CD8⁺ T cells proliferate and differentiate into effector cells aimed at controlling the pathogen by killing virus-infected host cells. Following virus elimination, 90–95% of effector T cells undergo apoptosis, whereas some antigen-specific T cells survive to become long-lived memory T cells that are able to protect the host from re-infection with the same pathogen^{2,6}.

While antigen-specific effector T cell responses are generated during acute SARS-CoV-2 infection^{7,8,9,10,11,12} and persist for several months^{13,14,15,16,17}, little is known about changes in memory phenotypes over time. Previous studies using live-attenuated virus vaccines in healthy donors^{18,19,20,21} have described phenotypical trajectories of human antigen-specific T cell populations. However, it is unknown whether infection with a natural virus generates comparable memory T cell responses in humans, as

infection route, viral load, inflammation and various host-related factors are likely to affect T cell responses and memory formation. Moreover, phenotypical and transcriptional trajectories at the single T cell receptor (TCR) level and the factors instructing individual effector T cell clones on their development to long-lived memory T cells have not been investigated in humans.

Phenotype of SARS-CoV-2⁺ CD8⁺ T cells

To assess the dynamics of antigen-specific T cells in COVID-19, we recruited 175 patients with real-time PCR (RT-PCR)-confirmed COVID-19, sampled during their symptomatic acute phase and followed up 6 months and 1 year after acute infection (Fig. 1a). We conducted human leukocyte antigen (HLA) typing on all patients and healthy controls and selected individuals carrying the *HLA-A*01:01*, *HLA-A*11:01* or *HLA-A*24:02* alleles for this study ($n = 47$ patients and $n = 13$ healthy controls; characteristics are included in Extended Data Table 1). In these individuals, SARS-CoV-2-specific CD8⁺ T cells were detected by using HLA-A*01:01, HLA-A*11:01 and HLA-A*24:02 major histocompatibility complex class I (MHC-I) dextramers¹², hereafter termed CoV2-Dex (Fig. 1b, Extended Data Fig. 1a, b), and validated by using HLA-A*01:01 and HLA-A*11:01 MHC-I pentamers²², hereafter termed CoV2-Pent (Extended Data Fig. 1c, d). Healthy controls were seronegative for SARS-CoV-2 spike-specific IgG and IgA (Extended Data Fig. 1e).

Fig. 1: Characteristics of antigen-specific CD8⁺ T cells during acute and memory phases of SARS-CoV-2 infection.

 **figure 1**

a, Overview of study design. PBMC, peripheral blood mononuclear cell. **b**, Representative plots of CoV2-Dex staining. PE, Phycoerythrin. Numbers in the plots indicate percentage of parent population. **c**, Frequency of CoV2-Dex⁺ cells in healthy donors and patients with COVID-19 during acute infection and 6 months and 1 year after infection. Each dot represents an independent donor at the indicated timepoint ($n = 10$ healthy, $n = 37$ acute, $n = 32$ 6 months, $n = 29$ 1 year after infection). P values are shown. **d**, Linear regression of frequency of CoV2-Dex⁺ cells 6 months after infection as a function of CoV2-Dex⁺ cell frequencies during acute infection ($n = 11$). The P value was calculated with t -statistic. **e**, Uniform manifold approximation and projection (UMAP) plots of marker expression for up to 2,000 CD8⁺ T cells from each sample collected during acute infection ($n = 37$) analysed by spectral flow cytometry.

Regions with high marker expression appear in red. An overlay of CoV2-Dex⁺ cells (red) and total CD8⁺ T cells (grey) is shown in the top left. **f**, Representative histograms showing expression of selected markers on CoV2-Dex⁻ and CoV2-Dex⁺ cells. **g**, Frequency of Ki-67⁺, HLA-DR⁺, granzyme B⁺, CX3CR1⁺ and CD127⁺ cells in CoV2-Dex⁻ (grey) and CoV2-Dex⁺ cells during acute infection and 6 months and 1 year after infection. Analysis was conducted on paired samples from acute infection versus 6 months and/or 1 year after infection ($n = 28$ acute, $n = 24$ 6 months, $n = 29$ 1 year). The grey lines connect individual donors sampled at different timepoints. P values were calculated using a Wilcoxon–Mann–Whitney test in **c** and **g** and corrected for multiple comparisons in **g**. All tests were performed two-sided.

SARS-CoV-2-specific CD8⁺ T cells were found in most patients carrying an *HLA-A*01:01* or *HLA-A*11:01* allele during acute infection and 6 months later (Fig. [1c](#)). Moreover, we detected SARS-CoV-2-specific CD8⁺ T cells 1 year after acute SARS-CoV-2 infection (Fig. [1b,c](#)). Staining with HLA-A*24:02 CoV2-Dex (which carried a spike-derived peptide, QYIKWPWYI) showed much higher background staining in some healthy donors (Extended Data Fig. [2a](#)), possibly due to cross-reactivity. In individuals carrying *HLA-A*24:02* alleles, we did not observe increased frequencies of CoV2-Dex⁺ cells during acute infection compared to healthy donors (Extended Data Fig. [2b](#)), contrary to patients with *HLA-A*01:01* or *HLA-A*11:01* alleles (Fig. [1c](#)). This finding might indicate that T cells specific for this spike epitope did not undergo marked expansion during SARS-CoV-2 infection. We could not determine whether HLA-A*24:02 CoV2-Dex⁺ cells had an activated or proliferating phenotype due to low cell numbers. Furthermore, we noted a lower reactivity to HLA-A*11:01 dextramers than to HLA-A*01:01 dextramers (Extended Data Fig. [2c](#)) during acute infection, which persisted 6 months after infection. These data suggest heterogeneity in effector and memory T cell responses based on HLA type and specific epitopes, although they need careful interpretation due to limited patient numbers.

In patients with an *HLA-A*01:01* or *HLA-A*11:01* allele, the frequency of CoV2-Dex⁺ cells during acute infection correlated with the frequency of specific cells at the memory phase (Fig. [1d](#)). In acute infection, flow cytometry analysis of CoV2-Dex⁺ cells revealed a circumscribed phenotype of activated cells, dominated by high abundance of Ki-67 and HLA-DR (Fig. [1e,f](#), Extended Data Fig. [2d,e](#)). CoV2-Dex⁺ cells also expressed granzyme B and tended to have higher expression of the terminal differentiation marker CX3CR1, whereas surface CD127 (also known as IL-7 receptor- α) was markedly downregulated (Fig. [1f,g](#)). At the 6-month and 1-year timepoints, frequencies of Ki-67⁺ and HLA-DR⁺ CoV2-Dex⁺ cells declined and the frequency of CD127⁺ cells increased (Fig. [1f,g](#)), indicating a transition from effector to memory state^{[6,23](#)}. We did not observe phenotypical differences between HLA-A*01:01 and HLA-A*11:01 (Extended Data Fig. [2f](#)).

Notably, disease severity seemed to positively correlate with the extent of CD8⁺ T cell responses during acute infection, as well as frequencies of CoV2-Dex⁺ cells 6 months and 1 year after infection, although expansion of CoV2-Dex⁺ cells was also evident in patients with mild disease (Extended Data Fig. 3a). During acute infection, both proliferation and activation were only minimally affected by disease severity in the CoV2-Dex⁺ compartment, whereas a relevant difference was observed in CoV2-Dex⁻ cells (Extended Data Fig. 3b). This discrepancy could be due to higher bystander activation in severe disease or higher abundance of undetected SARS-CoV-2-reactive T cells. Patients with severe disease showed higher expression of granzyme B and CX3CR1 on CoV2-Dex⁺ cells, possibly indicating a different T cell differentiation program during the acute phase of severe COVID-19. These differences were no longer evident 6 months and 1 year after infection (Extended Data Fig. 3c, d).

Transcriptome of SARS-CoV-2⁺ CD8⁺ clones

To examine the transcriptional phenotype of individual SARS-CoV-2-specific CD8⁺ T cells, we sorted CoV2-Dex⁺CD8⁺ T cells and CoV2-Dex⁻CD8⁺ T cells, mixed them at a 1:10 ratio, and performed single-cell RNA sequencing (scRNA-seq) on a subgroup of patients ($n = 20$ acute and $n = 19$ 6-month timepoint). We classified sequenced cells as CoV2-Dex⁻ or CoV2-Dex⁺ based on their dCODE Dextramer unique molecular identifier counts (Methods, Extended Data Fig. 4a) and positivity for a single SARS-CoV-2 epitope (Extended Data Fig. 4b). Unbiased clustering revealed 12 distinct CD8⁺ T cell clusters (Fig. 2a), none of which was dominated by a single patient (Extended Data Fig. 4c). Some clusters showed nearly complete segregation between the acute and memory phases (Extended Data Fig. 5a, b). In line with our flow cytometry data (Fig. 1e, Extended Data Fig. 5c), CoV2-Dex⁺CD8⁺ T cells showed a rather segregated transcriptional makeup during acute infection, whereas their transcriptional state was more heterogeneous 6 months after infection (Fig. 2b). Comparing the contribution of CoV2-Dex⁺ cells to different clusters, we observed that clusters 1, 2 and 12 dominated the CoV2-Dex⁺CD8⁺ T cell response in the acute phase, whereas clusters 3, 6 and 11 became prominent in the recovery phase (Fig. 2c, Extended Data Fig. 5d). While clusters 1, 2 and 12 corresponded to cytotoxic, activated and proliferating cells, respectively, cluster 3 showed a signature marked by enrichment of NF-κB and Jun/Fos signalling, cluster 6 displayed an oxidative phosphorylation signature, and cluster 11 showed a dual signature marked by enrichment of interferon (IFN) response genes and genes encoding the effector cytokines IFNγ, tumour necrosis factor (TNF) and lymphotoxin-α (LTα) (Fig. 2d, e). Similarly, among genes with significantly higher expression in CoV2-Dex⁺ cells from the acute phase versus the recovery phase, we found genes related to cytotoxicity (*GZMA*, *GZMK* and *PFNI*), activation (*HLA-DRA*, *CD38* and *PDCD5*) and proliferation (*MKI67*, *MCM7* and *NUDC1*), along with IFN response genes (*IFI6*,

MX1, *IFI27L2* and *IFI44L*) (Extended Data Fig. 5e). *SELL* (which encodes CD62L) appeared to be enriched in cells retrieved during the acute phase rather than the recovery phase (Extended Data Fig. 5e).

Fig. 2: Transcriptional makeup of SARS-CoV-2-specific CD8⁺ T cell clones.

 figure 2



a, Single-cell transcriptomes of CD8⁺ T cells displayed by UMAP. Seurat-based clustering of 14,853 cells, coloured based on cluster ID. **b**, UMAP as in **a**; CoV2-Dex⁺ cells from the acute infection (red) and 6 months after infection (blue) are highlighted.

c, Cluster composition of CoV2-Dex⁺CD8⁺ T cells in acute infection versus 6 months after infection. **d**, Average expression (colour scale) and the percentage of expressing cells (size scale) of selected genes in indicated clusters. **e**, Schematic summary of the main clusters differentially represented in acute infection and 6 months after infection. **f**, Clonotype distribution in CoV2-Dex⁺ T cell clones (at least one CoV2-Dex⁺ cell per clone) for each of the four epitopes assessed. The number of T cell clones specific for the indicated epitopes at acute infection (top) and 6 months after infection (bottom) is provided within the circle. **g**, Alluvial plot showing relative representation of single clones present during acute infection and 6 months after infection ($n = 41$). **h**, UMAP as in **a**; cells from individual CoV2-Dex⁺CD8⁺ T cell clones in acute infection (red) and 6 months after infection (blue) are highlighted. **i**, Gene signature scores of individual CD8⁺ T cell clones in acute infection (acute gene signature, left) versus 6 months after infection (recovery gene signature, right) ($n = 41$). **j**, Expression of *MKI67* (left) and *HLA-DRB5* (right) in individual CD8⁺ T cell clones in acute infection versus 6 months after infection ($n = 41$). *P* values were calculated using a Wilcoxon signed-rank test in **i** and **j**.

To identify phenotypical trajectories in individual antigen-specific T cell clones, we performed TCR sequencing of CoV2-Dex⁺ cells, which revealed several antigen-specific CD8⁺ T cell clones for each epitope investigated (Fig. 2f). Clones were considered antigen-specific if any of the clonal cells were CoV2-Dex⁺ (data available as Supplementary Dataset 1), and clones that were CoV2-Dex⁺ in the acute phase were considered CoV2-Dex⁺ independently of CoV2-Dex staining at six months after infection, and vice versa. The number of clones detected during convalescence was markedly lower than that detected during the acute phase of infection (Fig. 2f). In most cases, but not all, dominant clones in the acute phase corresponded to the largest clones found in the recovery phase (Fig. 2g). The phenotypical changes in acute infection versus the recovery phase on the general CoV2-Dex⁺ population were reflected in individual T cell clones. Thus, analysis of individual CoV2-Dex⁺ clones showed multiple clones containing cells from clusters 1, 2 or 12 during acute infection and cells from clusters 3, 6 and 11 during recovery (Fig. 2h, Extended Data Fig. 6). To better compare gene expression in acute infection versus recovery across all clones, we compiled an ‘acute gene signature’ comprising *NKG7*, *PRF1*, *GZMB*, *CENPU*, *CENPF* and *MKI67*, and a ‘recovery gene signature’ comprising *TNF*, *IFIT2*, *IFIT3*, *MT-CO1*, *MT-CO2* and *MT-ATP6*. We observed a significant decrease in acute gene signature transcripts in individual T cell clones from the acute phase to the recovery phase, which was paralleled by an increase in the recovery gene signature (Fig. 2i). Accordingly, individual T cell clones showed a decrease in *MKI67* and *HLA-DRB5* expression between the acute phase and the recovery phase (Fig. 2j).

Memory paths of SARS-CoV-2⁺CD8⁺ cells

To better understand the phenotypical memory trajectories of antigen-specific CD8⁺ T cells following a naturally occurring acute virus infection, we followed CoV2-Dex⁺ cells longitudinally, at both the population level and the clonal level. In the acute phase, CoV2-Dex⁺ cells showed mostly an effector/effector memory ($T_{\text{effector}}/T_{\text{EM}}$) phenotype, whereas frequencies of naive (T_{naive}) cells were lower in CoV2-Dex⁺ than in CoV2-Dex⁻ CD8⁺ T cells (Fig. 3a, b, Extended Data Fig. 7a).

Fig. 3: Transition of antigen-specific CD8⁺ T cells to TCF1⁺CD45RA⁺ effector memory cells at 1 year.

 figure 3

a, Representative plots of CD45RA and CCR7 staining on CoV2-Dex⁻ and CoV2-Dex⁺ cells during acute infection and 6 months and 1 year after infection. Numbers in the plots indicate percentage of parent population. **b**, Percentages of T_{naive} , T_{SCM} , T_{CM} , $T_{\text{effector}}/T_{\text{EM}}$ and T_{EMRA} cells in CoV2-Dex⁻ and CoV2-Dex⁺ cells during acute

infection ($n = 28$). **c**, Percentages of T_{naive} , T_{SCM} , T_{CM} , $T_{\text{effector}}/T_{\text{EM}}$ and T_{EMRA} CoV2-Dex $^{+}$ cells in acute infection and 6 months and 1 year after infection ($n = 28$ acute, $n = 24$ 6 months, $n = 29$ 1 year). The grey lines connect individual donors sampled at different timepoints. P values are also shown. **d**, Expression of CD45RA (left) and CCR7 (right) determined by TotalSeq in individual CD8 $^{+}$ T cell clones in acute infection versus 6 months after infection ($n = 41$). **e–g**, Geometric mean fluorescence intensity (gMFI) of selected markers on $T_{\text{effector}}/T_{\text{EM}}$ and T_{EMRA} CoV2-Dex $^{+}$ cells in acute infection (**e**), and 6 months (**f**) and 1 year (**g**) after infection. Phenotypes were evaluated only in patients with more than 5 $T_{\text{effector}}/T_{\text{EM}}$ and T_{EMRA} CoV2-Dex $^{+}$ cells per sample ($n = 24$ acute, $n = 24$ 6 months, $n = 26$ 1 year). **h**, Percentages of TCF1 $^{+}$, T-BET $^{+}$, EOMES $^{+}$ and TOX $^{+}$ CoV2-Dex $^{-}$ and CoV2-Dex $^{+}$ cells during acute infection ($n = 28$). **i**, Percentages of TCF1 $^{+}$, T-BET $^{+}$, EOMES $^{+}$ and TOX $^{+}$ CoV2-Dex $^{+}$ cells in acute infection, and 6 months and 1 year after infection ($n = 28$ acute, $n = 24$ 6 months, $n = 29$ 1 year). P values were calculated using a Wilcoxon signed-rank test in **b**, **d–g**, and a Wilcoxon–Mann–Whitney test with a correction for multiple comparisons using the Holm method in **c** and **i**. All tests were performed two-sided.

These data were confirmed in CoV2-Pent $^{+}$ cells (Extended Data Fig. [7b, c](#)). At 6 months and 1 year after infection, we observed a progressive switch from a $T_{\text{effector}}/T_{\text{EM}}$ phenotype to a terminally differentiated T_{EM} cell re-expressing CD45RA (T_{EMRA}) phenotype; thus, 1 year after infection, most CoV2-Dex $^{+}$ cells were of a T_{EMRA} phenotype (Fig. [3c](#), Extended Data Fig. [7d, e](#)). Furthermore, we observed progressive enrichment in stem cell memory T (T_{SCM}) cells, particularly at the 1-year timepoint (Fig. [3c](#), Extended Data Fig. [7e](#)). Of note, the increase in T_{SCM} cell enrichment was accompanied by an enrichment in T_{naive} cells 1 year after infection (Fig. [3c](#)), possibly indicating that memory cells can reacquire CD45RA and CCR7 also in the absence of CD95 expression. We did not observe differences in memory phenotypes based on HLA, except for a tendency towards more central memory (T_{CM}) cells in HLA-A*11:01 1 year after infection (Extended Data Fig. [7f–h](#)). Conversely, memory phenotypes were strongly influenced by disease severity (Extended Data Fig. [7i–k](#)). Patients with severe disease had fewer CoV2-Dex $^{+}$ T_{naive} cells 6 months and 1 year after infection, tended to have fewer T_{SCM} cells and showed predominantly CoV2-Dex $^{+}$ T_{EMRA} cells 1 year after infection. When examining individual T cell clones, we observed an increase in CD45RA expression and a concomitant decrease in CCR7 protein expression determined by TotalSeq from acute infection to 6 months after infection (Fig. [3d](#)), thus confirming an enrichment of a T_{EMRA} phenotype also on a single TCR level.

Subsequently, we assessed whether the T_{effector}/T_{EM} and T_{EMRA} phenotypes were associated with specific T cell markers, suggesting distinct differentiation states. Indeed, in CoV2-Dex⁺ cells, we observed several differences between the T_{effector}/T_{EM} and T_{EMRA} populations. CoV2-Dex⁺ T_{EM} cells showed higher expression of Ki-67 and HLA-DR, whereas they had lower abundance of CX3CR1 already during acute infection (Fig. 3e). Notably, we observed the same phenotypical differences between CoV2-Dex⁺ T_{EM} and CoV2-Dex⁺ T_{EMRA} cells 6 months and 1 year after infection (Fig. 3f,g).

As T cell phenotypes are driven by specific transcription factors, we assessed the expression of T cell factor 1 (TCF1), T-box expressed in T cells (T-BET), eomesodermin (EOMES) and thymocyte selection-associated high-mobility group box (TOX), which are transcription factors known to have important roles in T cell differentiation^{24,25,26,27}. CoV2-Dex⁺ cells downregulated TCF1 expression during the acute phase, which was progressively restored at subsequent timepoints (Fig. 3h,i, Extended Data Fig. 8a–c). Conversely, the expression of T-BET was increased in the acute phase (Fig. 3h, Extended Data Fig. 8a) and progressively decreased 6 months and 1 year after infection (Extended Data Fig. 8b,c). A difference in T-BET expression between T_{effector}/T_{EM} and T_{EMRA} CoV2-Dex⁺ cells was not evident, except for a tendency 1 year after infection. However, T_{EMRA} CoV2-Dex⁺ cells expressed lower levels of TCF1 and TOX in the memory phase and lower levels of EOMES at all timepoints (Extended Data Fig. 8d).

Signatures of CD8⁺ memory precursors

Next, we sought to identify the factors present during acute infection that instruct T cell clones towards a memory fate. We compared clones detected in the peripheral blood in both the acute and the convalescent phases (termed persistent) to those that were only seen in the acute phase and became undetectable in the convalescent phase (non-persistent) (Fig. 4a). Not all clones detected at 6 months after infection were present in the acute phase, probably reflecting a limitation of detection (Extended Data Fig. 9a). Generally, clone size correlated positively with persistence (Fig. 4b). Cells of persistent clones showed a different transcriptional makeup in the acute phase when compared to cells of non-persistent clones (Fig. 4c), which also resulted in a different distribution in the previously identified CD8⁺ T cell clusters (Extended Data Fig. 9b). This effect was robustly seen in different clones and was not due to a few hyper-expanded clones (Extended Data Fig. 9c). Gene set enrichment analysis revealed distinct signatures in persistent versus non-persistent clones. Genes involved in IFN γ and IFN α responses and TNF signalling were enriched in cells from persistent clones, whereas mechanistic target of rapamycin (mTOR) signalling and mitosis-related genes

were enriched in cells from non-persistent clones (Fig. 4d). By comparing differentially regulated genes between cells from persistent and non-persistent clones, we observed genes associated with activation (*HLA-DQ α 1* and *HLA-DP α 1*), terminal differentiation (*KLRG1*) and cytotoxicity (*GZMM* and *NKG7*), as well as certain IFN-induced (*B2M* and *HLA-C*) and TNF-induced (*CCL4*) genes to be enriched in persisters, along with CD45RA protein expression determined by TotalSeq (Fig. 4e). Conversely, cells from non-persistent clones showed higher expression of CTLA4, TIM3 (encoded by *HAVCR2*), Ki-67 (encoded by *MKI67*) and the mTOR-induced gene *CORO1A* (Fig. 4e). The same differences in gene expression could be observed at the clonal level; thus, genes were upregulated or downregulated accordingly in non-persistent individual clones compared to persistent clones (Fig. 4f), as exemplified in two selected clones of a patient (Fig. 4g). Non-persisters showed higher expression of *SELL* during the acute phase (Fig. 4e, f). We also observed differential TCR-V β usage between persistent and non-persistent clones (Fig. 4e).

Fig. 4: Transcriptional signature of antigen-specific CD8 $^{+}$ T cell clones persisting at 6 months.

 **figure 4**

a, Proportion of CoV2-Dex⁺ CD8⁺ T cell clones present during acute infection that were also detectable 6 months after infection. **b**, Clone size of persisting versus non-persisting CoV2-Dex⁺ CD8⁺ T cell clones ($n = 41$ persistent, $n = 139$ non-persistent). **c**, UMAP plot of persistent (red) versus non-persistent (green) CoV2-Dex⁺CD8⁺ T cell clones detected during acute infection. **d**, Gene set enrichment analysis showing enrichment of genes associated with cytokine signalling in persistent clones and mTOR signalling and proliferation in non-persistent CoV2-Dex⁺ T cell clones. Red dashed lines indicate minimal and maximal cumulative enrichment values. *P* value calculation was performed as detailed in the Method section. **e**, Expression of selected genes and CCR7 and CD45RA protein determined by Totalseq for persistent versus non-persistent CoV2-Dex⁺ T cell clones. *P* values were calculated using a Wilcoxon signed-rank test; a Bonferroni correction was applied for multiple

comparisons. **f**, Expression level of selected genes in persistent versus non-persistent individual T cell clones; each dot represents one clone. **g**, Expression level of selected genes in cells from a single non-persistent clone compared to cells from a single persistent T cell clone; each dot represents one cell ($n = 5$ CASSQVIGNQPQHF, $n = 16$ CASSAPGPLTTQYF). In **f**, **g**, the white diamonds indicate median expression. For **b**, **f**, **g**, P values were calculated using a Wilcoxon–Mann–Whitney test. All tests were performed two-sided.

Discussion

In this study, we address outstanding questions related to CD8⁺ T cell memory upon acute SARS-CoV-2 infection in humans by longitudinally following individual memory CD8⁺ T cell clones. Phenotypically, we find a transition of CD8⁺ T cells from T_{effector}/T_{EM} cells to T_{EMRA} cells with progressive enrichment of TCF1⁺ cells, which is paralleled by a modest enrichment in T_{SCM} cells. While two previous papers have reported a high prevalence of T_{EMRA} cells among SARS-CoV-2-specific CD8⁺ T cells^{13,28}, our study provides a description of progressive enrichment in this specific phenotype during the memory phase, at both the clonal level and the population level, revealing that CD8⁺ T_{EMRA} cells might constitute the main circulating memory subset following an acute viral infection in humans.

Our data provide a different and more detailed view of individual antigen-specific human memory CD8⁺ T cells than the one observed in the tetramer-positive memory CD8⁺ T cell population in individuals vaccinated against yellow fever virus¹⁹, where the prevalent subset observed was T_{SCM}. As that study dealt with human memory CD8⁺ T cells examined several years after vaccination, T_{EMRA} and T_{SCM} cells could be part of the same phenotypical trajectory, with progressive enrichment in T_{SCM} cells over time due to differentiation or competitive advantage. We favour an alternative hypothesis based on the phenotypical differences between memory cells in mild and severe COVID-19 that we observed, suggesting that other factors—such as antigen availability, type of antigen-presenting cells stimulated and cytokine milieu—might influence the type of memory formed, with increased T_{EMRA} cell differentiation upon severe disease versus prevalent T_{SCM} cell formation in mild disease and upon inoculation with live-attenuated virus vaccines.

We also observed enrichment of CD62L expression in CoV2-Dex⁺ cells during the acute phase rather than the recovery phase and in non-persistent clones disappearing from the circulation rather than in clones giving rise to circulating memory T cells. Whereas CD62L expression is a hallmark of recirculating T_{CM} cells²⁹, which are the

prevalent memory cells in lymph and secondary lymphoid organs³⁰, CD62L is not typically associated with the T_{EMRA} phenotype³¹. This might explain our findings, as CD62L expression appears not to be part of the memory differentiation trajectory observed²⁰. Whether a fraction of T cell clones detected in peripheral blood during the acute phase contributes to different memory pools in secondary lymphoid organs cannot be excluded in our present study.

Understanding how the immune system maintains the balance between effector response and memory formation could provide insights on why some infections result in robust and long-lasting T cell memory, whereas others fail to do so. Our study helps to unravel the complexity of these processes by finding a transcriptional signature at the level of T cell clones that correlates with the acquisition of long-lived, circulating memory T cells. We find that a strongly proliferative phenotype is associated with clonal contraction and disappearance. Furthermore mTOR signalling, probably stimulated by TCR engagement, appears to instruct the fate of short-lived effector cells, similar to previous results in mice³² and in vitro studies³³. Conversely, cytokine signalling marks cells destined to become long-lived, circulating memory cells, in agreement with previous studies showing the importance of type I IFN for memory generation³⁴. As we sampled SARS-CoV-2-specific T cells from the peripheral blood, we can only infer persistence of CD8⁺ T cell clones in this compartment. Importantly, memory phenotypes and the factors instructing their differentiation might vary in different immunological contexts, such as the lymph node or tissue. Similarly, as we could examine specific cells only from donors with *HLA-A*01:01* and *HLA-A*11:01* alleles and, to some extent, donors with an *HLA-A*24:02* allele, further studies will be needed to compare our findings in other HLA types.

Collectively, our data demonstrate the formation of memory CD8⁺ T cells to be dependent on a delicate balance between cytokine and TCR signalling during acute infection, which in turn influences outcomes of long-lived, circulating memory T cells in humans.

Methods

Human participants and patient characteristics

Following written informed consent, adult patients with symptomatic, RT-qPCR-confirmed SARS-CoV-2 infection were recruited in the Canton of Zurich, Switzerland, between 2 April and 24 September 2020. The study was approved by the Cantonal Ethics Committee of Zurich (BASEC 2016-01440). Patients ($n = 175$) donated peripheral blood at the time of inclusion into the study, and 116 and 90 patients donated peripheral blood approximately 6 months and 1 year after infection,

respectively. Standardized clinical data were collected for all included patients and disease severity was assessed, as previously described for this cohort^{35,36,37}. Peripheral blood mononuclear cells (PBMCs) and serum were bio-banked, as previously described^{35,36,37}. Following HLA class I typing, patients carrying an *HLA-A*01:01*, *HLA-A*11:01* and/or an *HLA-A*24:02* allele with sufficient bio-banked samples at two different timepoints were selected for the study ($n = 47$). Thirteen healthy donors carrying *HLA-A*01:01*, *HLA-A*11:01* and/or an *HLA-A*24:02* allele were included for comparison. SARS-CoV-2-specific CD8⁺ T cells were detected with MHC-I dextramers and pentamers in 42 and 12 patients, respectively.

IgA and IgG immunoassays

Spike S1-specific IgA and IgG antibodies were assessed with a commercial ELISA kit (SARS-CoV-2 IgA and IgG immunoassay, Euroimmun), as previously described³⁵. OD ratios higher than 2.0 and 1.1 were considered positive for serum IgA and IgG, respectively.

Dextramer and pentamer staining

PBMCs (4×10^6) per patient were incubated with Human TruStain FcX blocking reagent (422302, BioLegend) for 10 min at 4 °C. After washing, cells were incubated with MHC-I dextramers (see Supplementary Table 1) in the presence of l-biotin and herring sperm DNA according to the manufacturer's instructions, for 10 min at room temperature. Two peptides presented on HLA-A*01:01 dextramers (FTSDYYQLY from ORF3a and TTDPNFLGRY from ORF1ab), two peptides presented on HLA-A*11:01 dextramers (ATEGALNTPK and KTFPPTEP from nucleocapsid protein) and one peptide presented on HLA-A*24:02 dextramers (QYIKWPWYI from spike protein) were included. For MHC-I pentamer staining, cells were incubated for 10 min at 37 °C with pentamers (Extended Data Table 1). One peptide presented on HLA-A*01:01 pentamers (FTSDYYQLY from ORF3a) and one peptide presented on HLA-A*11:01 pentamers (KTFPPTEP from nucleocapsid protein) were included. Frozen PBMCs were used throughout the study.

Spectral flow cytometry

After dextramer or pentamer staining, a concentrated surface staining antibody mix (Supplementary Tables 2, 3) was applied without washing and samples were incubated at room temperature for further 20 min. After four rounds of washing, cells were resuspended in a fixation permeabilization solution (eBioscience Foxp3/transcription factor staining buffer) and incubated for 60 min at room temperature. After washing, an antibody mix for intracellular staining (Supplementary Tables 2, 3) was added and

cells were incubated for 30 min at room temperature. After washing, samples were acquired on a Cytek Aurora spectral flow cytometer using the SpectroFlo software. Quality control for the cytometer was performed daily. Data were analysed with FlowJo (version 10.7.1) and OMIQ (www.omiq.ai). Phenotypes were evaluated only in patients with more than 5 CoV2-Dex⁺ cells per sample ($n = 28$ acute, $n = 24$ 6 months after infection, $n = 29$ 1 year after infection) or more than 5 CoV2-Pent⁺ cells per sample ($n = 7$ acute, $n = 9$ 6 months after infection). Correlation between frequency of CoV2-Dex⁺ cells during acute infection and frequency of CoV2-Dex⁺ cells 6 months after infection was assessed only for patients sampled at least 14 days after symptom onset ($n = 11$).

Sample sets of sorted and unsorted cells and healthy controls

A concentrated antibody mix containing TotalSeq antibodies (see Supplementary Table 4 for a complete list) was applied after dextramer staining without washing and cells were incubated at 4 °C for 30 min. After four rounds of washing, cells were resuspended in PBS with 2% FBS and 2 mM EDTA and sorted with a BD Aria cell sorter. For each patient, CoV2-Dex⁻ and CoV2-Dex⁺ cells were sorted approximately in a 10:1 ratio. All CoV2-Dex⁺ cells from each sample were sorted, the corresponding amount of CoV2-Dex⁻ cells was calculated and sorted in the same tube. Cells from ten patients at the same timepoint were pooled together, generating four individual sample sets in total: (1) patients CoV2_001–CoV2_010, acute; (2) patients CoV2_001–CoV2_010, 6 months after infection; (3) patients CoV2_011–CoV2_020, acute; and (4) patients CoV2_011–CoV2_20, 6 months after infection. Two additional sample sets were generated using 5,000 unsorted PBMCs from each patient's sample: (5) patients CoV2_001–CoV2_010, 6 months after infection unsorted; and (6) patients CoV2_011–CoV2_020, 6 months after infection unsorted. Finally, using PBMCs from four healthy donors, we generated sample set (7) by sorting and pooling 2,000 CD8⁺ T cells from each healthy donor sample.

scRNA-seq library preparation and sequencing

Cells of sample sets 1–7 were analysed by scRNA-seq utilizing the 5' Single Cell GEX and VDJ v1.1 platforms (10x Genomics). Each sample set was processed individually. Cell suspensions were pelleted, resuspended and loaded into the Chromium Chip following the manufacturer's instructions. Fourteen cycles of initial cDNA amplification were used for all sets and single-cell sequencing libraries for whole-transcriptome analysis (GEX), TCR profiling (VDJ), and combined cell-surface protein and dCODE Dextramers detection (ADT) were generated. Final libraries were quantified using a Qubit Fluorometer, pooled in a ratio of 5:1:1 (GEX:VDJ:ADT) and

sequenced on a NovaSeq 6000 system with the following cycle configuration: read 1: 28 bp; index read 1: 10 bp; read 2: 101 bp.

Single-cell transcriptome analysis

Raw scRNA-seq FASTQ files were aligned to the human GRCh38 genome with Cell Ranger version 5.0.0 with default settings for the ‘cellranger multi’ pipeline (10x Genomics). The reference genome was downloaded from the 10x Genomics website (<https://cf.10xgenomics.com/supp/cell-exp/refdata-gex-GRCh38-2020-A.tar.gz>) and built as per official release notes (https://support.10xgenomics.com/single-cell-gene-expression/software/release-notes/build#GRCh38_2020A). Every sample set was analysed with the ‘cellranger multi’ pipeline, which allows to process together the paired GEX, ADT and VDJ libraries for each set. Downstream analysis was conducted in R version 4.1.0 with the package Seurat version 4.0.3 (ref. ³⁸). Cells with fewer than 200 or more than 2,500 detected genes and cells with more than 10% detected mitochondrial genes were excluded from the analysis.

To investigate possible patient biases, we demultiplexed cells from patient pools 1–6 based on genetic variants detected within the scRNA-seq reads. For this, we used the tool souporcell version 2 (ref. ³⁹). To cluster cells based on their patient-specific genetic variants, we merged sample sets 1, 2 and 5 (comprising sorted cells from both timepoints of patients CoV2_001–CoV2_010 and unsorted cells of the same patients) and sets 3, 4 and 6 (comprising cells from both timepoints of patients CoV2_011–CoV2_020 and unsorted cells of the same patients). Then, we executed the souporcell pipeline with option $k = 10$ (the number of clusters to be determined) for each of the two merged sample sets. This analysis allowed us to classify 88% of the cells passing the filtering steps from above into 20 genotype-driven ‘patient’ clusters.

After log normalization and variable feature calculation, independent datasets were integrated using Seurat’s anchoring-based integration method. Data scaling, principal component analysis, clustering and UMAP visualizations were performed on the integrated dataset using 15 principal components and a resolution of 0.5 for the shared nearest-neighbour clustering algorithm. To define distinct biological features of cell clusters, differential gene expression analyses were done on assay ‘RNA’ of the integrated dataset. FindAllMarkers was executed with logfc.threshold and min.pct cut-offs set to 0.25. For the analysis of clusters, FindMarkers was used with default settings for the comparison of persistent and non-persistent clones. For the differential expression analysis of manually selected genes and cell-surface proteins (CD45RA and CCR7), logfc.threshold and min.pct cut-offs were set to 0.

For gene set enrichment analysis, the FindMarkers function from Seurat was first used for the differential expression of genes between cells belonging to persistent and non-

persistent clones (using the default Wilcoxon rank-sum test, with options ‘min.pct=0.1, logfc.threshold = -Inf’, to account also for small expression changes, as long as the genes were expressed in at least 10% of cells of at least one group). The resulting 4,701 genes were pre-ranked in decreasing order by the negative logarithm of their *P* value, multiplied for the sign of their average log-fold change (in R, ‘-log(*P_val*)*sign(avg_log₂FC)'). Gene set enrichment analysis⁴⁰ was performed on this pre-ranked list using the R package FGSEA (<https://github.com/ctlab/fgsea/>)⁴¹. We used the FGSEA-simple procedure with 100,000 permutations and the hallmark gene sets for *Homo sapiens* from the Molecular Signatures Database (<https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>, made accessible in R by the package msigdbr; <https://github.com/cran/msigdbr>) and set the seed value ('set.seed(42)' in R) before execution to make the results reproducible. For significance testing, the function fgsea::fgsea() was used, which performs a *P* value estimation based on an adaptive multi-level split Monte-Carlo scheme. A multiple hypothesis correction procedure was applied to get adjusted *P* values. The results were filtered for gene sets that were significantly enriched with adjusted *P* < 0.1.

TCR profiling

Paired chain TCR sequences were obtained through targeted amplification of full-length V(D)J segments during library preparation. Sequence assembly and clonotype calling was done through cellranger's immune profiling pipeline (cellranger multi). TCR profiling on filtered contig annotations was done using R package scRepertoire version 1.1.4 (ref. ⁴²), which assigns TCR nucleotide and amino acid sequences together with clonal frequency counts and a clonotype classification to each cell. The function combineTCR was executed with filterMulti=T to isolate the top two expressed chains in cell barcodes with multiple chains. Clonotypes were called based on the amino acid sequence of the CDR3 region of TCR α and TCR β chains. For cells of which only one of the two chains could be identified, the available chain was used. Clone calling was done for each sample set independently before integration.

SARS-CoV-2 peptide-loaded dextramer binding of CD8⁺ T cells

To identify SARS-CoV-2-specific CD8⁺ T cells, we used dCODE Dextramers loaded with viral peptides presented on MHC-I molecules as described above. To assess unspecific binding, a negative control dextramer (peptide STEGGGLAY presented on HLA-A*01:01) and a general negative control dextramer were included. After analysis of the flow cytometry data, we noticed strong background staining of dextramer HLA-A*24:02 (peptide QYIKWPWYI) in samples of healthy donors, indicating unspecific dextramer binding. Thus, we excluded all sequencing counts from this dextramer in the downstream analysis. For other dextramers, cells were considered CoV2-Dex⁺

when the unique molecular identifier (UMI) count of a CoV2-dextramer was more than ten and more than five times higher than the UMI count of the negative control in the same cell. Cells that were positive for more than one dextramer according to this classification (less than 0.2% of all cells with known TCR) were excluded from the analysis. A TCR clone was considered SARS-CoV-2-specific when at least one cell of the clone was CoV2-Dex⁺.

Statistics

Wilcoxon–Mann–Whitney test was used for comparisons of two independent groups. Wilcoxon signed-rank test was used for paired testing. *P* values were adjusted for multiple comparisons with the Holm method. A linear regression model was used to quantify the relationship between variables. Significance was assessed by non-parametric methods unless otherwise specified. All tests were performed two sided. Analyses were performed with R (version 4.0.0 or 4.1.0).

Reporting summary

Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this paper.

Data availability

The sequencing dataset generated in this study has been deposited at [zenodo.org](#) and is available at <https://zenodo.org/record/5770747>. Flow cytometry datasets are available from the corresponding author on reasonable request.

Code availability

The code generated during the current study is available at <https://github.com/Moors-Code/SARS-CoV-2-Tcell-Boyman-collaboration>.

References

1. 1.

Crotty, S. & Ahmed, R. Immunological memory in humans. *Semin. Immunol.* **16**, 197–203 (2004).

2. 2.

Sallusto, F., Lanzavecchia, A., Araki, K. & Ahmed, R. From vaccines to memory and back. *Immunity* **33**, 451–463 (2010).

3. 3.

Saad-roy, C. M. et al. Immune life history, vaccination, and the dynamics of SARS-CoV-2 over the next 5 years. *Science* **818**, 811–818 (2020).

4. 4.

Jalkanen, P. et al. COVID-19 mRNA vaccine induced antibody responses against three SARS-CoV-2 variants. *Nat. Commun.* **12**, 3991 (2021).

5. 5.

Chemaitelly, H. et al. mRNA-1273 COVID-19 vaccine effectiveness against the B.1.1.7 and B.1.351 variants and severe COVID-19 disease in Qatar. *Nat. Med.* **27**, 1614–1621 (2021).

6. 6.

Raeber, M. E., Zurbuchen, Y., Impellizzieri, D. & Boyman, O. The role of cytokines in T-cell memory in health and disease. *Immunol. Rev.* **283**, 176–193 (2018).

7. 7.

Weiskopf, D. et al. Phenotype and kinetics of SARS-CoV-2-specific T cells in COVID-19 patients with acute respiratory distress syndrome. *Sci. Immunol.* **5**, eabd2071 (2020).

8. 8.

Grifoni, A. et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489–1501.e15 (2020).

9. 9.

Braun, J. et al. SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* **587**, 270–274 (2020).

10. 10.

Le Bert, N. et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **584**, 457–462 (2020).

11. 11.

Le Bert, N. et al. Highly functional virus-specific cellular immune response in asymptomatic SARS-CoV-2 infection. *J. Exp. Med.* **218**, e20202617 (2021).

12. 12.

Saini, S. K. et al. SARS-CoV-2 genome-wide T cell epitope mapping reveals immunodominance and substantial CD8⁺ T cell activation in COVID-19 patients. *Sci. Immunol.* **6**, eabf7550 (2021).

13. 13.

Dan, J. M. et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* **371**, eabf4063 (2021).

14. 14.

Zuo, J. et al. Robust SARS-CoV-2-specific T cell immunity is maintained at 6 months following primary infection. *Nat. Immunol.* **22**, 620–626 (2021).

15. 15.

Bonifacius, A. et al. COVID-19 immune signatures reveal stable antiviral T cell function despite declining humoral responses. *Immunity* **54**, 340–354 (2021).

16. 16.

Hou, H. et al. Immunologic memory to SARS-CoV-2 in convalescent COVID-19 patients at one-year post-infection. *J. Allergy Clin. Immunol.* **148**, 1481–1492.e2 (2021).

17. 17.

Minervina, A. A. et al. Longitudinal high-throughput tcr repertoire profiling reveals the dynamics of T-cell memory formation after mild COVID-19 infection. *eLife* **10**, e63502 (2021).

18. 18.

Akondy, R. S. et al. The yellow fever virus vaccine induces a broad and polyfunctional human memory CD8⁺ T cell response. *J. Immunol.* **183**, 7919–7930 (2009).

19. 19.

Akondy, R. S. et al. Origin and differentiation of human memory CD8 T cells after vaccination. *Nature* **552**, 362–367 (2017).

20. 20.

Mold, J. E. et al. Divergent clonal differentiation trajectories establish CD8⁺ memory T cell heterogeneity during acute viral infections in humans. *Cell Rep.* **35**, 109174 (2021).

21. 21.

Graham, N. et al. Rapid induction and maintenance of virus-specific CD8⁺ T_{EMRA} and CD4⁺ T_{EM} cells following protective vaccination against dengue virus challenge in humans. *Front. Immunol.* **11**, 479 (2020).

22. 22.

Peng, Y. et al. Broad and strong memory CD4⁺ and CD8⁺ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nat. Immunol.* **21**, 1336–1345 (2020).

23. 23.

Surh, C. D. & Sprent, J. Homeostasis of naive and memory T cells. *Immunity* **29**, 848–862 (2008).

24. 24.

Escobar, G., Mangani, D. & Anderson, A. C. T cell factor 1: a master regulator of the T cell response in disease. *Sci. Immunol.* **5**, eabb9726 (2020).

25. 25.

Intlekofer, A. M. et al. Effector and memory CD8⁺ T cell fate coupled by T-bet and eomesodermin. *Nat. Immunol.* **6**, 1236–1244 (2005).

26. 26.

Khan, O. et al. TOX transcriptionally and epigenetically programs CD8⁺ T cell exhaustion. *Nature* **571**, 211–218 (2019).

27. 27.

Wieland, D. et al. TCF1⁺ hepatitis C virus-specific CD8⁺ T cells are maintained after cessation of chronic antigen stimulation. *Nat. Commun.* **8**, 15050 (2017).

28. 28.

Jung, J. H. et al. SARS-CoV-2-specific T cell memory is sustained in COVID-19 convalescent patients for 10 months with successful development of stem cell-like memory T cells. *Nat. Commun.* **12**, 4043 (2021).

29. 29.

Geginat, J., Lanzavecchia, A. & Sallusto, F. Proliferation and differentiation potential of human CD8⁺ memory T-cell subsets in response to antigen or homeostatic cytokines. *Blood* **101**, 4260–4266 (2003).

30. 30.

Buggert, M. et al. The identity of human tissue-emigrant CD8⁺ T cells. *Cell* **183**, 1946–1961.e15 (2020).

31. 31.

Sallusto, F., Lenig, D., Forster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708–712 (1999).

32. 32.

Araki, K. et al. mTOR regulates memory CD8 T-cell differentiation. *Nature* **460**, 108–112 (2009).

33. 33.

Langenkamp, A. et al. T cell priming by dendritic cells: thresholds for proliferation, differentiation and death and intraclonal functional diversification. *Eur. J. Immunol.* **32**, 2046–2054 (2002).

34. 34.

Kolumam, G. A., Thomas, S., Thompson, L. J., Sprent, J. & Murali-Krishna, K. Type I interferons act directly on CD8 T cells to allow clonal expansion and memory formation in response to viral infection. *J. Exp. Med.* **202**, 637–650 (2005).

35. 35.

Cervia, C. et al. Systemic and mucosal antibody responses specific to SARS-CoV-2 during mild versus severe COVID-19. *J. Allergy Clin. Immunol.* **147**, 545–557 (2021).

36. 36.

Chevrier, S. et al. A distinct innate immune signature marks progression from mild to severe COVID-19. *Cell Rep. Med.* **2**, 100166 (2021).

37. 37.

Adamo, S. et al. Profound dysregulation of T cell homeostasis and function in patients with severe COVID-19. *Allergy* **76**, 2866–2881 (2021).

38. 38.

Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

39. 39.

Heaton, H. et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020).

40. 40.

Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

41. 41.

Korotkevich, G. et al. Fast gene set enrichment analysis. Preprint at <https://doi.org/10.1101/060012> (2021).

42. 42.

Borcherding, N., Bormann, N. L. & Kraus, G. scRepertoire: an R-based toolkit for single-cell immune receptor analysis. *F1000Res.* **9**, 47 (2020).

Acknowledgements

We thank S. Hasler for her assistance with patient recruitment and coordination; E. Baechli, A. Rudiger, M. Stüssi-Helbling, L. C. Huber and D. J. Schaer for their support in patient recruitment; and the members of the Boyman laboratory for helpful discussions. Graphical representations were generated with BioRender.com. This work was funded by the Swiss National Science Foundation (4078P0-198431 to O.B. and J.N.; and 310030-172978 and 310030-200669 to O.B.), the Clinical Research Priority Program of the University of Zurich for CRPP CYTImm-Z (to O.B.), an Innovation grant of University Hospital Zurich (to O.B.), the Pandemic Fund of University of Zurich (to O.B.), and the Botnar Research Centre for Child Health (COVID-19 FTC to A.E.M.). S.A., C.C. and Y.Z. received Swiss Academy of Medical Sciences fellowships (323530-177975, 323530-191220 and 323530-191230, respectively), S.A. received a Forschungskredit Candoc grant from the University of Zurich (FK-20-022) and M.E.R. received a Young Talents in Clinical Research Project Grant by the Swiss Academy of Medical Sciences and G. & J. Bangerter-Rhyner Foundation (YTCR 08/20).

Author information

Affiliations

1. Department of Immunology, University Hospital Zurich, Zurich, Switzerland

Sarah Adamo, Yves Zurbuchen, Carlo Cervia, Patrick Taeschler, Miro E. Raeber, Jakob Nilsson & Onur Boyman

2. Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland

Jan Michler, Simona Baghai Sain & Andreas E. Moor

3. Faculty of Medicine, University of Zurich, Zurich, Switzerland

Onur Boyman

Contributions

S.A. designed and performed flow cytometry and scRNA-seq experiments, analysed and interpreted data. J.M. performed scRNA-seq experiments, analysed and interpreted data. Y.Z. contributed to flow cytometry and scRNA-seq experiments, patient recruitment and data collection. C.C. and P.T. contributed to patient recruitment and data collection. M.E.R. contributed to patient recruitment and clinical management. S.B.S. analysed scRNA-seq data. J.N. contributed to study design and patient recruitment. A.E.M. designed experiments and interpreted data. O.B. conceived the project, designed experiments and interpreted data. S.A. and O.B. wrote the manuscript with contribution by J.M. and S.B.S. All authors edited and approved the final draft of the article.

Corresponding author

Correspondence to [Onur Boyman](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Katherine Kedzierska 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 Gating strategy for antigen-specific CD8⁺ T cells and SARS-CoV2-specific antibodies in healthy donors and COVID-19 patients.](#)

a, Gating strategy for CoV2-Dex⁺ cells. **b**, Representative plot of CoV2-Dex staining for HLA-A*11:01 and HLA-A*24:02. **c**, Gating strategy for CoV2-Pent⁺ cells. **d**,

Linear regression of frequency of CoV2-Dex⁺ cells as a function of frequency of CoV2-Pent⁺ cells ($n = 7$). The p-value was calculated with t-statistic. **e**, SARS-CoV2-specific IgA (left) and IgG (right) of healthy donors compared to COVID-19 patients during acute infection, six months and one year after infection ($n = 13$ healthy, $n = 46$ acute, $n = 41$ six months, $n = 30$ one year). Dashed lines represent diagnostic cut-off values, i.e., 2.0 and 1.1 for IgA and IgG, respectively. P-values were calculated with a Mann-Whitney-Wilcoxon test. All tests were performed two-sided.

Extended Data Fig. 2 HLA-A*24:02 dextramer staining and CoV2-Dex⁺ cell frequency and phenotype based on HLA.

a, Representative plots of CoV2-Dex (HLA-A*24:02) staining in healthy donors. **b**, Percentage of CoV2-Dex⁺ cells in HLA-A*24:02 healthy donors and COVID-19 patients in acute infection and six months and one year after infection. Dots represent independent donors at each timepoint ($n = 3$ healthy, $n = 6$ acute, $n = 5$ six months, $n = 3$ one year). Gray lines connect individual donors sampled at different timepoints. **c**, Percentage of CoV2-Dex⁺ cells in HLA-A*01:01 and HLA-A*11:01 in acute infection and six months and one year after infection (HLA-A*01:01 $n = 26$ and HLA-A*11:01 $n = 10$ acute, HLA-A*01:01 $n = 21$ and HLA-A*11:01 $n = 9$ six months, HLA-A*01:01 $n = 20$ and HLA-A*11:01 $n = 9$ one year). **d**, Gating strategy for phenotypical analysis of CoV2-Dex⁺ compared to CoV2-Dex⁻ cells. **e**, Gating strategy for phenotypical analysis of CoV2-Pent⁺ compared to CoV2-Pent⁻ cells and frequency of Ki-67⁺, HLA-DR⁺ and CD127 levels in CoV2-Pent⁻ (gray) and CoV2-Pent⁺ (red) cells in acute COVID-19 ($n = 7$). **f**, Frequency of Ki-67⁺, HLA-DR⁺, granzyme B⁺, CX3CR1⁺ and CD127⁺ cells in CoV2-Dex⁺ cells in patients with an HLA-A*01:01 versus HLA-A*11:01 allele (HLA-A*01:01 $n = 22$ and HLA-A*11:01 $n = 6$) in the acute phase. **b**, **c**, and **f**, P-values were calculated with a Mann-Whitney-Wilcoxon test. **e**, P-values were calculated with a Wilcoxon signed-rank test. All tests were performed two-sided.

Extended Data Fig. 3 Frequency and phenotype of antigen-specific CD8⁺ T cells relative to disease severity.

a, Frequency of CoV2-Dex⁺ CD8⁺ T cells in healthy controls and patients with mild and severe disease during acute COVID-19 six months and one year after infection ($n = 10$ healthy; acute: $n = 21$ mild, $n = 15$ severe; six months: $n = 18$ mild, $n = 16$ severe; one year: $n = 19$ mild, $n = 10$ severe). **b-d**, Frequency of Ki-67⁺, HLA-DR⁺, granzyme B⁺, CX3CR1⁺ and CD127⁺ in CoV2Dex⁻ and CoV2-Dex⁺ cells in patients with mild versus severe disease (**b**) during acute infection, (**c**) six months and (**d**) one year after infection (acute: $n = 16$ mild, $n = 12$ severe; six months: $n = 11$ mild, $n = 16$ severe; one

year: $n = 19$ mild, $n = 10$ severe). P-values were calculated with a Mann-Whitney-Wilcoxon test and adjusted for multiple comparisons using the Holm method. All tests were performed two-sided.

Extended Data Fig. 4 Definition of CoV2-Dex⁺ cells and single patient contribution to individual clusters.

a, Unique molecular identifier (UMI) counts for CoV2-Dex HLA-A*11:01 (ATEGALNTPK) versus UMI counts for negative control dextramer; cells defined as CoV2-Dex HLA-A*11:01 (ATEGALNTPK)⁺ are depicted in blue. **b**, (UMI) counts for CoV2-Dex HLA-A*11:01 (ATEGALNTPK) versus UMI counts for CoV2-Dex HLA-A*11:01 (KTFPPTEPK). **c**, Uniform manifold approximation and projection (UMAP) plot colored by patient ID (left) and cluster distribution for single patients (right).

Extended Data Fig. 5 Characteristics of antigen-specific CD8⁺ T cells six months after infection.

a, UMAP plot of transcriptomes from CD8⁺ T cells of healthy individuals (gray), COVID-19 patients in acute infection (red), and six months after infection (blue). **b**, Cluster contribution to total CD8⁺ T cells of healthy subjects, acute infection, and six months after infection. **c**, UMAP plots of marker expression for up to 2,000 CD8⁺ T cells from each sample collected six months after infection analyzed by spectral flow cytometry. Regions with high expression of specific markers appear red. Overlay of CoV2-Dex⁺ cells (red) and total CD8⁺ T cells (gray) is shown on the upper left. **d**, Percentages of CoV2-Dex⁺ cells contributing to indicated clusters during acute infection or six months after infection. Percentages are calculated on total cells per cluster per timepoint. **e**, Expression of selected genes (italicized), as well as CCR7 and CD45RA proteins determined by TotalSeqTM, for CoV2-Dex⁺ cells detected in the acute versus recovery phase. P-values were calculated using a Mann-Whitney-Wilcoxon test and a Bonferroni correction was applied for multiple comparisons.

Extended Data Fig. 6 Transcriptome of SARS-CoV-2-specific CD8⁺ T cell clones.

UMAP plots highlighting single-cell transcriptomes belonging to individual CoV2-Dex⁺ T cell clones in acute infection (red) and six months after infection (blue).

Extended Data Fig. 7 Memory phenotypes in CoV2-Dex[±] and CoV2-Pent[±] cells.

a, b, Gating strategy for identifying naïve, stem cell memory (T_{SCM}), central memory (T_{CM}), effector/effector memory ($T_{effector}/T_{EM}$), and effector memory T cells re-expressing CD45RA (T_{EMRA}) in (a) CoV2-Dex⁻ and CoV2-Dex⁺ cells or (b) CoV2-Pent⁻ and CoV2-Pent⁺ cells. **c**, Percentages of naïve, T_{CM} , $T_{effector}/T_{EM}$, and T_{EMRA} cells among CoV2-Pent⁻ (gray) and CoV2-Pent⁺ (red) cells ($n = 7$). **d, e**, Percentages of naïve, T_{CM} , $T_{effector}/T_{EM}$, and T_{EMRA} cells among CoV2-Dex⁻ (gray) and CoV2-Dex⁺ (red) cells (**d**) six months and (**e**) one year after infection ($n = 24$ six months, $n = 29$ one year). **f–h**, Percentages of naïve, T_{SCM} , T_{CM} , $T_{effector}/T_{EM}$, and T_{EMRA} in CoV2-Dex⁺ cells in patients with HLA-A*01:01 versus HLA-A*11:01 allele (**f**) during acute infection, (**g**) six months and (**h**) one year after infection (acute: $n = 22$ HLA-A*01:01 and $n = 6$ HLA-A*11:01; six months: $n = 19$ HLA-A*01:01 and $n = 5$ HLA-A*11:01; one year: $n = 20$ HLA-A*01:01 and $n = 9$ HLA-A*11:01). **i–k**, Percentages of naïve, T_{SCM} , T_{CM} , $T_{effector}/T_{EM}$, and T_{EMRA} in CoV2-Dex⁺ cells in patients with mild versus severe disease (**i**) during acute infection, (**j**) six months and (**k**) one year after infection (acute: $n = 16$ mild, $n = 12$ severe; six months: $n = 11$ mild, $n = 16$ severe; one year: $n = 19$ mild, $n = 10$ severe). P-values were calculated with a Wilcoxon signed-rank test in **c–e** and with a Mann-Whitney-Wilcoxon test in **f–k**. All tests were performed two-sided.

Extended Data Fig. 8 Gating strategy and pseudolongitudinal time course of transcription factor expression.

a, Gating strategy TCF1⁺, T-BET⁺, EOMES⁺, and TOX⁺ cells in CoV2-Dex⁻ and CoV2-Dex⁺ cells. **b**, Percentages of TCF1⁺, T-BET⁺, EOMES⁺, and TOX⁺ cells in CoV2-Dex⁻ and CoV2-Dex⁺ cells (left) six months and (right) one year after infection ($n = 24$ six months, $n = 29$ one year). **c**, Linear regression of frequencies of TCF1⁺, T-BET⁺, EOMES⁺, and TOX⁺ cells in CoV2-Dex⁺ cells as a function of time since symptom onset ($n = 81$). P-values were calculated with t-statistic. **d**, Expression of transcription factors in $T_{effector}/T_{EM}$ and T_{EMRA} in CoV2-Dex⁺ cells in acute infection and six months and one year after infection ($n = 24$ acute, $n = 24$ six months, $n = 26$ one year). P-values were calculated using a Wilcoxon signed-rank test in **b** and **d**. All tests were performed two-sided.

Extended Data Fig. 9 Clonality and phenotype of persistent versus non-persistent CoV2-Dex[±] T cell clones.

a, Venn diagram showing overlapping clones during acute infection and at six months after infection. **b**, Cluster composition of persistent versus non-persistent CD8⁺ T cell clones. **c**, UMAP plots showing top ten persistent CoV2-Dex⁺ CD8⁺ T cell clones and top ten non-persistent CoV2-Dex⁺ CD8⁺ T cell clones.

Extended Data Table 1 Clinical and demographic characteristics of healthy subjects and COVID-19 patients

Supplementary information

Reporting Summary

Supplementary Table 1

SARS-CoV-2-specific dextramers (Dex) and pentamers (Pent).

Supplementary Table 2

Fluorophore-marked reagents used in spectral flow cytometry (dextramer staining).

Supplementary Table 3

Fluorophore-marked reagents used in spectral flow cytometry (pentamer staining).

Supplementary Table 4

Fluorophore-marked reagents used in cell sorting.

Supplementary Dataset 1

Percentages of CoV2Dex⁺ cells in CoV2Dex⁺ clones.

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

Adamo, S., Michler, J., Zurbuchen, Y. *et al.* Signature of long-lived memory CD8⁺ T cells in acute SARS-CoV-2 infection. *Nature* **602**, 148–155 (2022).
<https://doi.org/10.1038/s41586-021-04280-x>

- Received: 22 July 2021
- Accepted: 24 November 2021
- Published: 07 December 2021
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04280-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

Further reading

- [The T cell immune response against SARS-CoV-2](#)
 - Paul Moss

- Article
- [Published: 30 November 2021](#)

An autoimmune stem-like CD8 T cell population drives type 1 diabetes

- [Sofia V. Gearty](#) [ORCID: orcid.org/0000-0001-8633-3663](#)^{1,2},
- [Friederike Dündar](#) [ORCID: orcid.org/0000-0002-2301-112X](#)^{3,4},
- [Paul Zumbo](#)^{3,4},
- [Gabriel Espinosa-Carrasco](#)¹,
- [Mojdeh Shakiba](#)¹,
- [Francisco J. Sanchez-Rivera](#)⁵,
- [Nicholas D. Soccia](#)⁶,
- [Prerak Trivedi](#)¹,
- [Scott W. Lowe](#) [ORCID: orcid.org/0000-0002-5284-9650](#)⁵,
- [Peter Lauer](#)⁷,
- [Neeman Mohibullah](#)⁸,
- [Agnes Viale](#)⁸,
- [Teresa P. DiLorenzo](#)^{9,10,11},
- [Doron Betel](#) [ORCID: orcid.org/0000-0002-8006-7752](#)^{4,12,13} &
- [Andrea Schietinger](#) [ORCID: orcid.org/0000-0003-3644-1687](#)^{1,2}

[Nature](#) volume 602, pages 156–161 (2022)

- 16k Accesses
- 1 Citations
- 191 Altmetric

- [Metrics details](#)

Subjects

- [Type 1 diabetes](#)

Abstract

CD8 T cell-mediated autoimmune diseases result from the breakdown of self-tolerance mechanisms in autoreactive CD8 T cells¹. How autoimmune T cell populations arise and are sustained, and the molecular programmes defining the autoimmune T cell state, are unknown. In type 1 diabetes, β-cell-specific CD8 T cells destroy insulin-producing β-cells. Here we followed the fate of β-cell-specific CD8 T cells in non-obese diabetic mice throughout the course of type 1 diabetes. We identified a stem-like autoimmune progenitor population in the pancreatic draining lymph node (pLN), which self-renews and gives rise to pLN autoimmune mediators. pLN autoimmune mediators migrate to the pancreas, where they differentiate further and destroy β-cells. Whereas transplantation of as few as 20 autoimmune progenitors induced type 1 diabetes, as many as 100,000 pancreatic autoimmune mediators did not. Pancreatic autoimmune mediators are short-lived, and stem-like autoimmune progenitors must continuously seed the pancreas to sustain β-cell destruction. Single-cell RNA sequencing and clonal analysis revealed that autoimmune CD8 T cells represent unique T cell differentiation states and identified features driving the transition from autoimmune progenitor to autoimmune mediator. Strategies aimed at targeting the stem-like autoimmune progenitor pool could emerge as novel and powerful immunotherapeutic interventions for type 1 diabetes.

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: Phenotype, function and heterogeneity of IGRP-specific CD8 T cells.

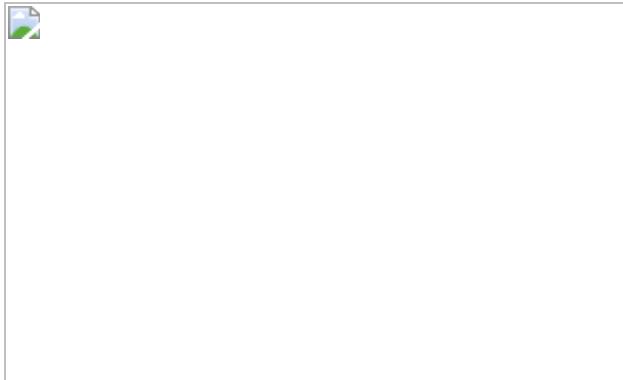


Fig. 2: Genome-wide transcriptome analysis of IGRP-specific CD8 T cell populations from pancreatic lymph node and pancreas.

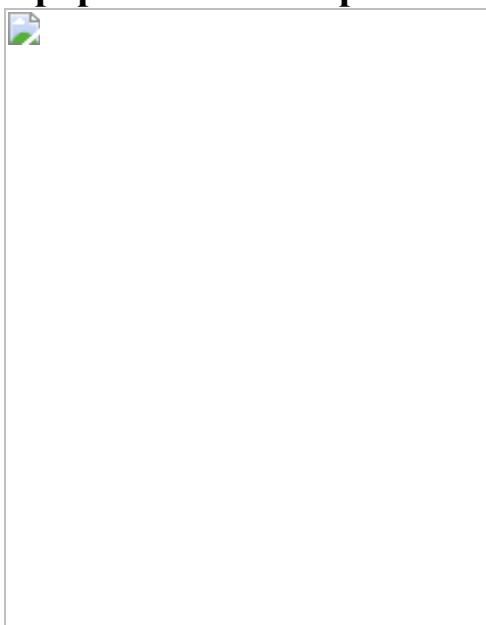


Fig. 3: IGRP-specific pLN TCF1^{hi} CD8 T cells represent a stem-like progenitor population that self-renews and gives rise to short-lived AM in the pancreas.

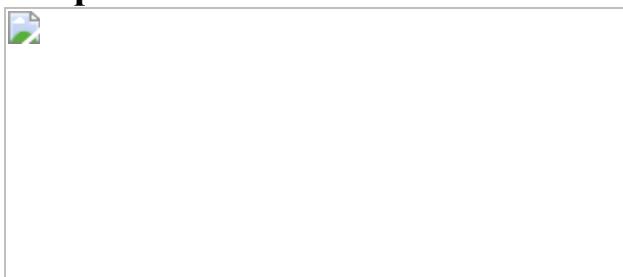
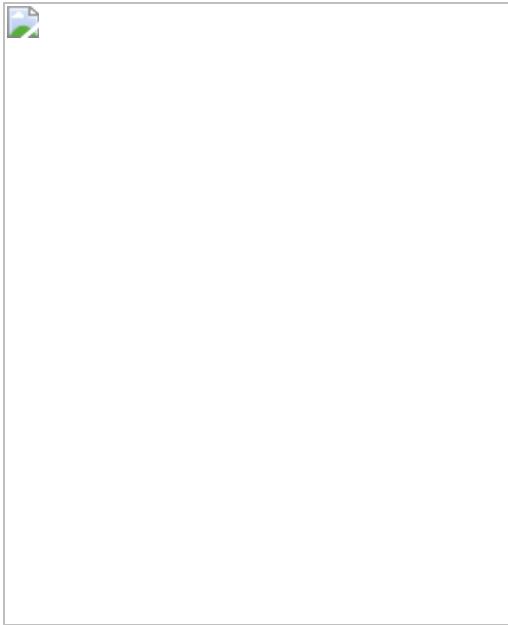


Fig. 4: Single-cell transcriptomic and T cell receptor analyses of IGRP-specific CD8 T cell populations.



Data availability

The RNA-seq data have been deposited in the Gene Expression Omnibus under accession [GSE151652](#). Comparisons to publicly available published data used the following datasets: GSE119943²⁴, GSE130130²⁵ and GSE70813⁵⁶. Additional information and materials will be made available upon request. [Source data](#) are provided with this paper.

Code availability

All custom scripts used in this study are deposited at
<https://github.com/abcwcm/GeartySchietinger> and
<https://github.com/soccin/ScreenSEQ/tree/feature/simpleScripts>.

References

1. 1.

Bluestone, J. A., Bour-Jordan, H., Cheng, M. & Anderson, M. T cells in the control of organ-specific autoimmunity. *J. Clin. Invest.* **125**, 2250–2260 (2015).

2. 2.

Kaech, S. M. & Cui, W. Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat. Rev. Immunol.* **12**, 749–761 (2012).

3. 3.

Philip, M. & Schietinger, A. CD8⁺ T cell differentiation and dysfunction in cancer. *Nat. Rev. Immunol.* <https://doi.org/10.1038/s41577-021-00574-3> (2021).

4. 4.

Bluestone, J. A., Herold, K. & Eisenbarth, G. Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature* **464**, 1293–1300 (2010).

5. 5.

Herold, K. C., Vignali, D. A., Cooke, A. & Bluestone, J. A. Type 1 diabetes: translating mechanistic observations into effective clinical outcomes. *Nat. Rev. Immunol.* **13**, 243–256 (2013).

6. 6.

Anderson, M. S. & Bluestone, J. A. The NOD mouse: a model of immune dysregulation. *Annu. Rev. Immunol.* **23**, 447–485 (2005).

7. 7.

Unanue, E. R. Antigen presentation in the autoimmune diabetes of the NOD mouse. *Annu. Rev. Immunol.* **32**, 579–608 (2014).

8. 8.

Tsai, S., Shameli, A. & Santamaria, P. CD8⁺ T cells in type 1 diabetes. *Adv. Immunol.* **100**, 79–124 (2008).

9. 9.

Zakharov, P. N., Hu, H., Wan, X. & Unanue, E. R. Single-cell RNA sequencing of murine islets shows high cellular complexity at all stages of autoimmune diabetes. *J. Exp. Med.* **217**, e20192362 (2020).

10. 10.

Coppieters, K. T. et al. Demonstration of islet-autoreactive CD8 T cells in insulitic lesions from recent onset and long-term type 1 diabetes patients. *J. Exp. Med.* **209**, 51–60 (2012).

11. 11.

Ferris, S. T. et al. A minor subset of Batf3-dependent antigen-presenting cells in islets of Langerhans is essential for the development of autoimmune diabetes. *Immunity* **41**, 657–669 (2014).

12. 12.

Lieberman, S. M. et al. Identification of the beta cell antigen targeted by a prevalent population of pathogenic CD8⁺ T cells in autoimmune diabetes. *Proc. Natl Acad. Sci. USA* **100**, 8384–8388 (2003).

13. 13.

Escobar, G., Mangani, D. & Anderson, A. C. T cell factor 1: a master regulator of the T cell response in disease. *Sci Immunol* **5**, eabb9726 (2020).

14. 14.

van Loosdregt, J. & Coffer, P. J. The role of WNT signaling in mature T cells: T cell factor is coming home. *J. Immunol.* **201**, 2193–2200 (2018).

15. 15.

Reya, T. et al. A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423**, 409–414 (2003).

16. 16.

Weissman, I. L., Anderson, D. J. & Gage, F. Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. *Annu. Rev. Cell Dev. Biol.* **17**, 387–403 (2001).

17. 17.

Christianson, S. W., Shultz, L. D. & Leiter, E. H. Adoptive transfer of diabetes into immunodeficient NOD-scid/scid mice. Relative contributions of CD4⁺ and CD8⁺ T-cells from diabetic versus prediabetic NOD.NON-Thy-1a donors. *Diabetes* **42**, 44–55 (1993).

18. 18.

Verdaguer, J. et al. Spontaneous autoimmune diabetes in monoclonal T cell nonobese diabetic mice. *J. Exp. Med.* **186**, 1663–1676 (1997).

19. 19.

Graef, P. et al. Serial transfer of single-cell-derived immunocompetence reveals stemness of CD8⁺ central memory T cells. *Immunity* **41**, 116–126 (2014).

20. 20.

Lapidot, T. et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645–648 (1994).

21. 21.

Maki, T., Gottschalk, R., Ogawa, N. & Monaco, A. P. Prevention and cure of autoimmune diabetes in nonobese diabetic mice by continuous administration of FTY720. *Transplantation* **79**, 1051–1055 (2005).

22. 22.

Brinkmann, V. et al. Fingolimod (FTY720): discovery and development of an oral drug to treat multiple sclerosis. *Nat. Rev. Drug Discov.* **9**, 883–897 (2010).

23. 23.

Ehlers, M. R. & Rigby, M. R. Targeting memory T cells in type 1 diabetes. *Curr. Diab. Rep.* **15**, 84 (2015).

24. 24.

Yao, C. et al. Single-cell RNA-seq reveals TOX as a key regulator of CD8⁺ T cell persistence in chronic infection. *Nat. Immunol.* **20**, 890–901 (2019).

25. 25.

Schauder, D. M. et al. E2A-regulated epigenetic landscape promotes memory CD8 T cell differentiation. *Proc. Natl Acad. Sci. USA* **118**, e2013452118 (2021).

26. 26.

Abdelsamed, H. A. et al. Beta cell-specific CD8⁺ T cells maintain stem cell memory-associated epigenetic programs during type 1 diabetes. *Nat. Immunol.* **21**, 578–587 (2020).

27. 27.

Culina, S. et al. Islet-reactive CD8⁺ T cell frequencies in the pancreas, but not in blood, distinguish type 1 diabetic patients from healthy donors. *Sci Immunol* **3**, eaao4013 (2018).

28. 28.

Pelletier, D. & Hafler, D. A. Fingolimod for multiple sclerosis. *N. Engl. J. Med.* **366**, 339–347 (2012).

29. 29.

Herold, K. C. et al. An anti-CD3 antibody, teplizumab, in relatives at risk for type 1 diabetes. *N. Engl. J. Med.* **381**, 603–613 (2019).

30. 30.

Sims, E. K. et al. Teplizumab improves and stabilizes beta cell function in antibody-positive high-risk individuals. *Sci. Transl. Med.* **13**, eabc8980 (2021).

31. 31.

Magnuson, A. M. et al. Population dynamics of islet-infiltrating cells in autoimmune diabetes. *Proc. Natl Acad. Sci. USA* **112**, 1511–1516 (2015).

32. 32.

Thivolet, C., Bendelac, A., Bedossa, P., Bach, J. F. & Carnaud, C. CD8⁺ T cell homing to the pancreas in the nonobese diabetic mouse is CD4⁺ T cell-dependent. *J. Immunol.* **146**, 85–88 (1991).

33. 33.

Sinnathamby, G. et al. Priming and activation of human ovarian and breast cancer-specific CD8⁺ T cells by polyvalent *Listeria monocytogenes*-based vaccines. *J. Immunother.* **32**, 856–869 (2009).

34. 34.

Chee, J. et al. Effector-memory T cells develop in islets and report islet pathology in type 1 diabetes. *J. Immunol.* **192**, 572–580 (2014).

35. 35.

Li, D. S., Yuan, Y. H., Tu, H. J., Liang, Q. L. & Dai, L. J. A protocol for islet isolation from mouse pancreas. *Nat. Protoc.* **4**, 1649–1652 (2009).

36. 36.

Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

37. 37.

Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

38. 38.

Harrow, J. et al. GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res.* **22**, 1760–1774 (2012).

39. 39.

Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

40. 40.

Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

41. 41.

DEGreport: Report of DEG analysis (2020).

42. 42.

Sergushichev, A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. Preprint at <https://doi.org/10.1101/060012> (2016).

43. 43.

Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

44. 44.

Amezquita, R. A. et al. Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* **17**, 137–145 (2020).

45. 45.

McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186 (2017).

46. 46.

Kim, H. J., Lin, Y., Geddes, T. A., Yang, J. Y. H. & Yang, P. CiteFuse enables multi-modal analysis of CITE-seq data. *Bioinformatics* **36**, 4137–4143 (2020).

47. 47.

Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75 (2016).

48. 48.

Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).

49. 49.

Lun, A. T., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122 (2016).

50. 50.

Angerer, P. et al. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243 (2016).

51. 51.

Ji, Z. & Ji, H. TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* **44**, e117 (2016).

52. 52.

Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).

53. 53.

Marini, F. & Binder, H. pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components. *BMC Bioinf.* **20**, 331 (2019).

54. 54.

Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720 (2015).

55. 55.

Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).

56. 56.

Mackay, L. K. et al. Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science* **352**, 459–463 (2016).

Acknowledgements

We thank the members of the Schietinger laboratory, M. Li (MSKCC) and S. Reiner (Columbia University) for helpful discussions; K. Herold (Yale University) for reviewing data and discussion; H. Appleby and E. Bruno for technical support; S. Miakicheva for help with experiments; the MSKCC Flow Cytometry Core, especially R. Gardner; the Integrated Genomics Operation Core, especially A. Farina; M. Philip (Vanderbilt University) for critical feedback on the manuscript; and S. Monette (Tri-Institutional Laboratory of Comparative Pathology, MSKCC) for pathological evaluation of insulitis. This work was supported by the following funding sources: Medical Scientist Training Programme grant NIH T32GM007739 to the Weill Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program (S.V.G.), NIH NIDDK F30DK122691 (S.V.G.), NIH NCI DP2CA225212 (A.S.), Lloyd Old STAR Award of the Cancer Research Institute (A.S.), Josie Robertson Foundation (A.S), the MSKCC Core Grant P30CA008748, Weill Cornell Medicine Core Laboratories Center (P.Z., F.D. and D.B.), NIH NIAID R01AI123730 (T.P.D.). F.J.S.-R. was partially supported by the MSKCC TROT programme (5T32CA160001), a GMTEC Postdoctoral Researcher Innovation Grant and is a HHMI Hanna Gray Fellow. The MSKCC Integrated Genomics Operation Core was supported by Cycle for Survival and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology.

Author information

Affiliations

1. Immunology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Sofia V. Gearty, Gabriel Espinosa-Carrasco, Mojdeh Shakiba, Prerak Trivedi & Andrea Schietinger

2. Immunology and Microbial Pathogenesis Program, Weill Cornell Medicine, New York, NY, USA

Sofia V. Gearty & Andrea Schietinger

3. Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA

Friederike Dündar & Paul Zumbo

4. Applied Bioinformatics Core, Weill Cornell Medicine, New York, NY, USA

Friederike Dündar, Paul Zumbo & Doron Betel

5. Cancer Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Francisco J. Sanchez-Rivera & Scott W. Lowe

6. Bioinformatics Core, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Nicholas D. Socci

7. Aduro Biotech, Berkeley, CA, USA

Peter Lauer

8. Integrated Genomics Operation Core, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Neeman Mohibullah & Agnes Viale

9. Department of Microbiology and Immunology & Division of Endocrinology, Department of Medicine, Albert Einstein College of Medicine, Bronx, NY, USA

Teresa P. DiLorenzo

10. Einstein-Mount Sinai Diabetes Research Center, Albert Einstein College of Medicine, Bronx, NY, USA

Teresa P. DiLorenzo

11. The Fleischer Institute for Diabetes and Metabolism, Albert Einstein College of Medicine, Bronx, NY, USA

Teresa P. DiLorenzo

12. Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA

Doron Betel

13. Division of Hematology and Medical Oncology, Department of Medicine, Weill Cornell Medicine, New York, NY, USA

Doron Betel

Contributions

S.V.G. and A.S. conceived and designed the study and analysed and interpreted data. F.D., P.Z. and D.B. performed RNA-seq, scRNA-seq, scTCR-seq and CITE-seq analyses. N.D.S. performed genetic fate-mapping analyses. M.S. and G.E.-C. carried out experiments. F.J.S.-R. and S.W.L. designed barcodes for genetic fate mapping. P.L. provided the *Listeria* strain. N.M. and A.V. provided advice on scRNA-seq, scTCR-seq and CITE-seq. P.T. provided advice on study design. T.P.D. provided critical advice on study design and T cell epitopes in NOD mice. S.V.G. and A.S. wrote the manuscript, with all authors contributing to the writing and providing feedback.

Corresponding authors

Correspondence to [Doron Betel](#) or [Andrea Schietinger](#).

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review information

Nature thanks Stephen Turner 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 Phenotypic characterization of IGRP-specific CD8 T cells in NOD mice.

a, Depletion of CD8 T cells prevents T1D. 12-week-old female WT NOD mice were treated with anti-CD8 α depleting antibody (clone 53-6.7, 250ug/mouse intraperitoneally (i.p.) twice weekly) for 4 months. Control mice received PBS i.p. n = 4 per group. Statistical testing was conducted via log-rank (Mantel-Cox) test. **b**, Absolute numbers (top) or frequency (bottom) of IGRP-specific (NRP-V7 tetramer+) CD8+ T cells in pancreatic lymph node (pLN; blue) and pancreas (red) at 5-9, 10-14, 15-20, and 21-30 week-old NOD mice, respectively. For absolute numbers: pLN: n = 6 (5-9wk), n = 4 (10-14wk), n = 4 (15-20wk), n = 3 (21-30wk). pancreas: n = 3 (5-9wk), n = 4 (10-14wk), n = 4 (15-20wk), n = 3 (21-30wk). For frequencies: pLN: n = 10 (5-9wk), n = 8 (10-14wk), n = 17 (15-20wk), n = 16 (21-30wk). Pancreas: n = 5 (5-9wk), n = 7 (10-14wk), n = 15 (15-20wk), n = 15 (21-30wk). Each symbol represents an individual mouse. Data are mean \pm SEM. Statistical testing via unpaired two-tailed Student's *t*-test. ns, not significant. **c**, Expression of CD44, PD1, CD38, CD39, LAG3, and TIGIT of IGRP-specific CD8+ T cells from pLN (blue) and pancreas (PA; red). Naïve T cells (NA; NRP-V7-CD44-CD8+ NOD splenocytes from the same mouse) are shown as controls (grey); for CD38: NRP-V7 CD8+ NOD

splenocytes. Representative of 2-3 independent experiments, n=4-5. **d**, Expression of transcription factors EOMES, TBET, BLIMP1 and TOX in IGRP-specific CD8+ (NRP-V7 tet+) T cells isolated from pLN (blue) and pancreas (PA; red); Naïve T cells (NA; NRP-V7 tet- CD44- CD8+ splenocytes) are shown as controls (grey). Data are normalized to NA mean fluorescence intensity (MFI). EOMES and TOX, pooled data from 2 independent experiments, n = 9 (EOMES) and n = 7 (TOX). TBET and BLIMP1, data representative of 2 independent experiments, n=3. Statistical testing via paired two-tailed Student's *t*-test. Data are mean ± SEM.

[Source data](#)

Extended Data Fig. 2 Phenotypic characterization of IGRP-specific and Insulin B (InsB)-specific CD8+ T cells in various tissues of NOD mice.

a, Expression of TCF1 versus TOX in IGRP-specific CD8+ T cells from pancreatic lymph node (pLN) and pancreas; Naïve (NA) T cells are shown as controls. Inset numbers show percentage of cells per gate. Data is representative of 2-3 independent experiments, n = 3-6. **b**, Bifurcation of IGRP-specific CD8 T cells into TCF1^{hi} and TCF1^{lo} T cell populations is not found iLN or spleen. Expression profiles of TCF1, CD44, CD39, PD1, and TOX of IGRP-specific CD8 T cells isolated from inguinal LN (iLN; purple) and spleen (green) of NOD mice; naïve T cells are shown as controls (NA). Inset numbers show percentage for each gate. Right panel, percentage of TCF1^{lo}CD39+ cells of NRP-V7 tet+ CD44+ CD8+ T cells in iLN and spleen (iLN, n=5; spleen, n = 3). Statistical testing via unpaired two-tailed Student's *t*-test. Data are mean ± SEM. Representative of 2 (iLN) and 5 (spleen) independent experiments. **c**, IGRP-specific and Insulin B (InsB)-specific CD8 T cells bifurcate into phenotypically similar populations in the pancreatic lymph node. Left, NRP-V7+ and InsB+ CD8 T cell populations in pancreas; FACS plot is gated on live CD8+ T cells. Right, expression of TCF1, CD39, PD1, and CD44 in pLN TCF1^{hi} and TCF1^{lo} NRP-V7+ (grey) and InsB+ (blue) CD8 T cells in pancreatic lymph node (top), and NRP-V7+ (grey) and InsB+ (red) CD8 T cells in pancreas (bottom). pLN cells pooled from multiple mice. Representative of 2-3

independent experiments with n = 3-14 (pLN) and n = 6-9 (pancreas). **d**, Expression levels of CD44, TCF1, CD39, PD1, and CD62L of TCF1^{hi} and TCF1^{lo} IGRP-specific (NRP-V7 tet+) and Insulin B (InsB tet+) CD8 T cells from pLN and pancreas. For pLN, each symbol represents 3 pooled mice; for pancreas, each symbol represents one individual mouse. Data are mean ± SEM. Statistical testing conducted via unpaired two-tailed Student's *t*-test. **e**, TCF1 and PD1 expression profiles of NRP-V7+ and polyclonal CD44+CD39+CD8+ T cells from pLN and pancreas. pLN cells pooled from multiple mice.

[Source data](#)

Extended Data Fig. 3 Transcriptome analysis of IGRP-specific CD8 T cells from pancreatic lymph node and pancreas.

a, MA plot of RNA-seq of pLN TCF1^{hi} and pLN TCF1^{lo} IGRP-specific CD8 T cells. Significantly differentially expressed genes are colored in light or dark blue and labeled with *. Selected genes are highlighted. **b**, Gene ontology (GO) term analysis of differentially expressed genes in pLN RNA-seq data set. GeneRatio>20/X. *P*-adj<0.01. Statistical testing conducted via hypergeometric test with Benjamini-Hochberg correction. **c**, GSEA of pLN TCF1^{hi} and pLN TCF1^{lo} cells including Hallmark and KEGG gene sets. **d**, Over-representation of KEGG and Hallmark pathways and GO terms in groups 1 and 4. *P*-adj<0.05.

[Source data](#)

Extended Data Fig. 4 IGRP-specific autoimmune CD8 T cell populations are phenotypically and transcriptionally distinct from IGRP-specific effector and memory CD8 T cells generated during acute infections.

a, Experimental scheme showing generation of NRP-V7 tet+ effector and memory CD8 T cells during an acute Listeria infection in comparison to NRP-V7 tet+ autoimmune CD8 T cells in NOD mice. A Listeria monocytogenes strain was generated expressing the NRP-V7 mimotope

(*L.m.* NRP-V7); host mice (Balb/c or NOD) were infected with *L.m.* NRP-V7 and effector and memory T cells from indicated tissues were analyzed at indicated time points using NRP-V7 tetramer. Effector and memory T cells were compared with NRP-V7+ autoimmune T cell populations from NOD mice and tissues as indicated. **b**, Top, identification of NRP-V7+ CD8 T cells in pancreas (effectors (Eff); black) at day 6-7 post infection of Balb/c mice or autoimmune NRP-V7+ CD8 T cells in pancreas of uninfected NOD mice (AI; red). Right, frequency of NRP-V7+ of CD8 T cells. Expression levels of KLRG1, CD127, TBET, and TCF1 in Eff and AI NRP-V7+ CD8 T cells from pancreas of Balb/c and NOD, respectively, are shown and quantified. Eff, n=4 (%NRP-V7+, %KLRG1+, %CD127+, TCF1 MFI) or n = 6 (TBET MFI); AI, n=3 (%NRP-V7+, %KLRG1+, %CD127+, TBET MFI) or n = 4 (TCF1 MFI). **c**, Expression of KLRG1 and CD127 of NRP-V7+ effectors from spleens of *L.m.*-infected Balb/c mice. **b, c**, Representative of 3 independent experiments. **d**, Frequency of KLRG1+ of NRP-V7+ CD8 T cells from pancreas (left) and spleen (right) of *L.m.*-infected Balb/c and NOD mice at day 6 post *L.m.* NRP-V7 infection compared to autoimmune (AI) NRP-V7+ CD8 T cells from NOD mice. Eff, n=3; NOD Eff, n = 4; NOD, n = 3 (pancreas), n=4 (spleen). Representative of 1 experiment. **e**, Top, identification of NRP-V7+ CD8 T cells in pancreatic lymph node (pLN) (memory (Mem); black) at day 21-28 post infection and autoimmune (AI) NRP-V7+ CD8 T cells in pLN of uninfected NOD mice (blue). Frequency of NRP-V7+ of CD8 T cells quantified on right. Expression profiles and frequencies of KLRG1, CD127, TCF1, CD62L, CD122, and CXCR3 of Mem and AI NRP-V7+ CD8 T cells from pLN of Balb/c and NOD, respectively. Mem, n = 7 (%NRP-V7+), n=6 (%TCF1+CD44+), n = 9 (%CD62L+), n = 3 (%CD127+, CD122 MFI, CXCR3 MFI); AI, n = 3 (NRP-V7+), n=7 (%TCF1+CD44+), n = 11 (%CD62L+), n=4 (%CD127+, CD122 MFI, CXCR3 MFI), n = 3 (%NRP-V7+). Representative of 3 independent experiments; %CD62L+ shows 3 combined experiments. Statistical testing conducted by Student's two-tailed unpaired *t*-test. Data are mean ± SEM. **f**, Gene set enrichment analysis of RNA-seq data from IGRP-specific pLN TCF1^{hi} CD8 T cells (see Fig. 2) compared to central memory CD8 T cells generated during acute LCMV infection; dataset from Mackay et al, 2016, GSE70813. NES, normalized enrichment score.

Source data

Extended Data Fig. 5 Longitudinal analysis of IGRP-specific 8.3 CD8 T cells post transfer into NOD Scid mice.

a, NRP-V7+ tetramer staining of CD8+ splenocytes from 8.3 TCR transgenic mouse. Representative of at least 3 independent experiments, n = 2. **b**, Expression of TCF1, CD44, CD39, PD1, and TOX of naïve 8.3 CD8 T cells pre-transfer (Day 0). **c**, Expression of indicated molecules in pLN and pancreas of NOD Scid hosts at day 2, 5, 12, and 20 post transfer (note: transferred donor T cells were not detected in pancreas before day 12). **b, c**, Representative of 2 (day 2, day 5, and 12) or 3 (day 20) independent experiments with n=2 (day 2, day 5, and 12) or n = 6 (day 20) replicates. **d**, Expression of indicated molecules in inguinal lymph nodes (iLN) and spleens of NOD Scid hosts at day 20 post transfer. Representative of 2 independent experiments with n=2 replicates.

Extended Data Fig. 6 Analysis of transferred TCF1^{hi} or TCF1^{lo} CD8 T cells in 2nd NOD Scid hosts.

a–c, 5,000 pLN TCF1^{hi} or 5,000 pancreatic TCF1^{lo} 8.3 CD45.2+ CD8 T cells (isolated from 1st NOD Scid hosts) were transferred to 2nd NOD Scid hosts. **a**, Top, blood glucose curve of 2nd NOD Scid hosts receiving 5,000 pLN TCF1^{hi} (blue, n = 5) or pancreatic TCF1^{lo} (red, n = 4) 8.3 CD8 T cells. Each line represents an individual mouse. Bottom, percentage of non-diabetic mice over time in indicated cohorts. Statistical testing conducted via two-sided log-rank (Mantel-Cox) test. Representative of at least three independent experiments with n = 3-6 mice per group; mice were monitored for 50 days post transfer. **b**, FACS plots show donor TCF1^{hi} or TCF1^{lo} 8.3 found in pLN and pancreas 7 days (TCF1^{hi}, n = 2; TCF1^{lo}, n = 3) and 19-21 days (TCF1^{hi}, n = 5; TCF1^{lo}, n = 6) post transfer into NOD Scid 2nd hosts. **c**, Absolute numbers of donor T cells recovered in indicated tissues of 2nd NOD Scid hosts 19-21 days post transfer. Top, recovery of 5,000 8.3 TCF1^{hi} (blue) or TCF1^{lo} (red) donor CD8 T cells. Bottom, recovery of 1x10⁶ polyclonal CD4 T cells co-transferred with 5,000 TCF1^{hi} or TCF1^{lo}

CD8 T cells. Total donor T cells per spleen, pair of pLN or inguinal LN (iLN), or per 100mg pancreas. Data are mean \pm SEM. Statistical testing conducted via unpaired two-tailed Student's *t*-test. **d**, Flow cytometric analysis of active caspases 3 and 7 in IGRP-specific CD8+ T cells isolated from pLN and pancreas. Shown is mean fluorescence intensity (MFI) normalized to MFI of naïve splenocytes. Each symbol represents an individual mouse, lines indicate populations from the same mouse. Pooled data from 2 independent experiments, n = 7 total. Statistical testing conducted by paired Student's two-tailed *t*-test. **e–g** (related to Fig. 3f–h), analysis at day 31 post FTY720 treatment. **e**, number of 8.3 cells per 100mg pancreas relative to the average number of cells per 100mg pancreas at day 15 (left); flow cytometric analysis of apoptosis using apotransfer detection reagent (see Methods) (PBS, n=4; FTY720, n = 3) (right); **f**, Absolute cell counts per pLN. **g**, Flow cytometric analysis of apoptosis of 8.3 donor T cells in spleens using apotransfer detection reagent. **f** and **g**, n = 4 per group. Representative of 1 (**e**, **g**) or 2 (**f**) independent experiments. Each symbol represents an individual mouse. Statistical testing was conducted by unpaired two-tailed Student's *t*-tests. Data are mean \pm SEM.

[Source data](#)

Extended Data Fig. 7 Stem-like functionality of pLN TCF1^{hi} CD8 T cells and generation of differentiated TCF1^{lo} CD8 T cells are independent of lymphopenic conditions.

a, Top, experimental scheme of 8.3 adoptive T cell transfer into NOD Scid (1st host) and wild-type (WT) NOD hosts (2nd host). Bottom, TCF1 and CD39 expression profiles of endogenous NRP-V7 tet+ CD8+ T cells (top) or transferred donor pLN TCF1^{hi} 8.3 CD8+ T cells (bottom) isolated from pLN and pancreas from same WT NOD 2nd hosts 35 days post transfer. Flow plots are representative of n=6 mice. **b**, Top, experimental scheme of 8.3 adoptive T cell transfer into WT NOD (1st host) and NOD Scid hosts (2nd host). Bottom, Ly108 and CD39 expression profiles of donor 8.3 CD8 T cells in WT NOD 1st hosts isolated from pLN and pancreas 29 days post transfer. Representative of 3 independent experiments with n = 10-15 NOD 1st hosts, n = 5-10 pLN pooled for sort. Right, TCF1 and CD39 expression

profiles of donor pLN TCF1^{hi} cells isolated from pLN and pancreas of NOD Scid 2nd hosts 41 days post transfer. Representative of 2 independent experiments with n = 3 mice per group. **c**, Blood glucose curves of NOD Scid 2nd hosts receiving 200-500 pLN TCF1^{hi} (blue) or 1,000 pancreatic TCF1^{lo} (red) 8.3 CD8 T cells. Each line represents an individual mouse. Representative of 2 independent experiments with n = 3 mice per group.

[Source data](#)

Extended Data Fig. 8 Single cell (sc) RNA-seq of pLN IGRP-specific (NRP-V7+) CD8 T cells reveals intermediate cell states during differentiation of pLN TCF1^{hi} to pLN TCF1^{lo} cells.

a, Numbers of T cells in each cluster classified as TCF1^{hi}CD62L^{hi}, TCF1^{lo}, and intermediate cells in pLN based on scRNA-seq data in individual mice (n=29). **b**, Smoothed conditional means of expression of selected genes across individual cells ordered by pseudotime; gray shading indicates 95% confidence interval and the bar on top represents the cluster color code. **c**, Diffusion maps colored by normalized expression values of individual genes. **d**, Heatmap of pLN scRNA-seq expression (row-based z-score of normalized log counts) for cluster marker genes (false discovery rate (FDR) < 5%). Cells were ordered based on pseudotime. Red bars on top indicate normalized protein expression values as determined by CITE-seq. **e**, Left, Expression profiles of TCF1, CD62L, and CD39 in pLN IGRP-specific CD8 T cells by flow cytometric analysis; Right, expression of CD62L and CD39 via CITE-seq antibody-derived tags. **f**, Overrepresented Gene Ontology Biological Process terms based on the marker genes of each cluster.

Extended Data Fig. 9 scTCR-Seq and genetic barcoding experiments reveal derivation of IGRP-specific pancreatic T cells from pLN TCF1^{hi} population.

a, Alluvial plots tracking the size of non shared (yellow) and shared clonotypes in pLN TCF1^{hi}, pLN TCF1^{lo}, and pancreas (PA) populations

from n = 21 mice. Each non-yellow stratum represents a clonotype that is found across multiple populations. **b**, Number of pancreatic T cells from individual mice that contain clonotypes that were either found in pLN TCF1^{hi} (red) or not found in pLN TCF1^{hi} (orange) population. Each row represents one individual mouse. **c**, Number of clonotypes across all mice that are shared with at least one other population (blue) or non-shared (yellow) in pLN TCF1^{hi} and pLN TCF1^{lo} populations. **d**, Fraction of clonotypes in pLN TCF1^{hi} and pLN TCF1^{lo} populations from individual mice that are shared with at least one other population (blue) or non-shared (yellow). Each row represents one mouse. **e**, Bar plot (top) and Venn diagram (bottom) displaying the different populations from which a given clonotype was recovered. The numbers are based on all clonotypes from all mice for which paired information from pLN and pancreas was available. **f**, Top, pie charts represent number of cells in each population containing clonotypes from pLN TCF1^{hi} that are shared with at least one other population (n = 127 clonotypes, identified in Extended Data Fig. 9c). Bottom, median number of cells per clonotype in each population split by shared and non shared clonotypes. Each dot represents the cell count from a single mouse. For box plots, center indicates median, lower and upper hinges indicate 25th and 75th percentiles, respectively, and whiskers extend up to 1.5xIQR (inter-quartile range) from the smallest/highest values; points shown individually fall outside that range. Notches extend 1.58xIQR/sqrt(n), roughly indicating 95% confidence interval for comparing medians. **g**, Cells from one example shared clonotype are shown within the coordinates of the scRNA-seq based diffusion map. Individual cells are colored according to their cluster membership. **h**, Left, Alluvial plots from genetic barcoding experiment show barcodes detected in pLN TCF1^{hi} 8.3 and pancreatic 8.3 for n = 2 mice; Right, Venn diagrams show numbers of barcodes detected in pLN TCF1^{hi} 8.3 (to achieve 99% barcode coverage) and expanded in pancreas (>1% of barcodes detected).

[Source data](#)

[**Extended Data Fig. 10 Autoimmune NOD pLN IGRP-specific CD8 T cells in T1D are distinct from CD8 T cells generated during acute and chronic infections.**](#)

scRNA-seq datasets from pLN NRP-V7+ CD8 T cells (Fig. 4) were compared to scRNA-seq datasets of memory precursor effector cells (MPEC) and central memory CD8 T cells (T_{CM}) generated during acute LCMV infection (day 7 and day 129 respectively), as well as to progenitor exhausted CD8 T cells (T_{PEX}) from day 7 during chronic (clone 13) LCMV infection. **a**, Principal component analysis (PCA) of pseudo-bulk RNA-seq samples representing $TCF1^{hi}CD62L^{hi}$, $TCF1^{hi}CD62L^{lo}$ (intermediate), and $TCF1^{lo}CD62L^{lo}$ NOD pLN NRP-V7+ CD8 T cells from 3 technical replicates from Gearty et al (Fig. 4; T1D), T_{CM} cells from Schauder et al (PNAS 2021; GSE130130; acute LCMV, day 129) and MPEC as well as T_{PEX} as defined by Yao *et al* (Nat. Immunology, 2019; GSE119943; acute and chronic LCMV, day 7). The bar charts to the left and on the bottom of the PCA plot represent the genes that are most strongly (anti)correlated with a given principal component (PC). **b**, Dendrogram representing the results of hierarchical clustering based on the pseudo-bulk RNA-seq samples. **c**, Heatmap displaying row-normalized expression values of individual cells for genes identified as marker genes when comparing the single cells of the different populations ($TCF1^{hi}CD62L^{hi}$, intermediate, and $TCF1^{lo}$, MPEC, T_{PEX} , T_{CM}) to each other. **d**, Top, GO terms (biological processes) that are enriched based on genes that are specifically overexpressed in the different populations compared to all others. Bottom, the network-like representations display the genes (small circles) that belong to the respective GO terms (large circles), highlighting genes shared among individual GO terms and upregulated in indicated cell populations.

Supplementary information

Supplementary Figure 1

Representative flow cytometry gating strategies.

Reporting Summary

Source data

[**Source Data Fig. 1**](#)

[**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. 6**](#)

[**Source Data Extended Data Fig. 7**](#)

[**Source Data Extended Data Fig. 9**](#)

Rights and permissions

[**Reprints and Permissions**](#)

About this article

Cite this article

Gearty, S.V., Dündar, F., Zumbo, P. *et al.* An autoimmune stem-like CD8 T cell population drives type 1 diabetes. *Nature* **602**, 156–161 (2022).
<https://doi.org/10.1038/s41586-021-04248-x>

- Received: 20 July 2020
- Accepted: 15 November 2021

- Published: 30 November 2021
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04248-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

Further reading

- [**A subset of immune-system T cells branded as seeds for type 1 diabetes**](#)

- Stephen J. Turner
- Nicole L. La Gruta

Nature (2022)

[**A subset of immune-system T cells branded as seeds for type 1 diabetes**](#)

- Stephen J. Turner
- Nicole L. La Gruta

News & Views 25 Jan 2022

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

- Article
- [Published: 20 January 2022](#)

Life histories of myeloproliferative neoplasms inferred from phylogenies

- [Nicholas Williams](#) ORCID: [orcid.org/0000-0003-3989-9167¹](https://orcid.org/0000-0003-3989-9167),
- [Joe Lee^{1,2}](#),
- [Emily Mitchell^{1,2,3,4}](#),
- [Luiza Moore](#) ORCID: [orcid.org/0000-0001-5315-516X¹](https://orcid.org/0000-0001-5315-516X),
- [E. Joanna Baxter³](#),
- [James Hewinson¹](#),
- [Kevin J. Dawson¹](#),
- [Andrew Menzies¹](#),
- [Anna L. Godfrey⁴](#),
- [Anthony R. Green^{2,3,4 na1}](#),
- [Peter J. Campbell](#) ORCID: [orcid.org/0000-0002-3921-0510^{1,2,3 na1}](https://orcid.org/0000-0002-3921-0510) &
- [Jyoti Nangalia](#) ORCID: [orcid.org/0000-0001-7122-4608^{1,2,3,4 na1}](https://orcid.org/0000-0001-7122-4608)

[Nature](#) volume 602, pages 162–168 (2022)

- 6789 Accesses
- 93 Altmetric
- [Metrics details](#)

Subjects

- [Cancer genomics](#)

- [Myeloproliferative disease](#)

Abstract

Mutations in cancer-associated genes drive tumour outgrowth, but our knowledge of the timing of driver mutations and subsequent clonal dynamics is limited^{1,2,3}. Here, using whole-genome sequencing of 1,013 clonal haematopoietic colonies from 12 patients with myeloproliferative neoplasms, we identified 580,133 somatic mutations to reconstruct haematopoietic phylogenies and determine clonal histories. Driver mutations were estimated to occur early in life, including the in utero period. *JAK2^{V617F}* was estimated to have been acquired by 33 weeks of gestation to 10.8 years of age in 5 patients in whom *JAK2^{V617F}* was the first event. *DNMT3A* mutations were acquired by 8 weeks of gestation to 7.6 years of age in 4 patients, and a *PPM1D* mutation was acquired by 5.8 years of age. Additional genomic events occurred before or following *JAK2^{V617F}* acquisition and as independent clonal expansions. Sequential driver mutation acquisition was separated by decades across life, often outcompeting ancestral clones. The mean latency between *JAK2^{V617F}* acquisition and diagnosis was 30 years (range 11–54 years). Estimated historical rates of clonal expansion varied substantially (3% to 190% per year), increased with additional driver mutations, and predicted latency to diagnosis. Our study suggests that early driver mutation acquisition and life-long growth and evolution underlie adult myeloproliferative neoplasms, raising opportunities for earlier intervention and a new model for cancer development.

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: Patient cohort and experimental design.

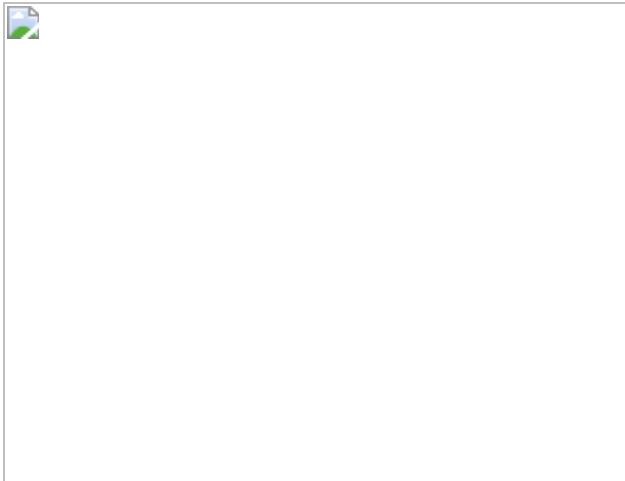


Fig. 2: Phylogenetic histories of three patients with MPN driven by $JAK2^{V617F}$.

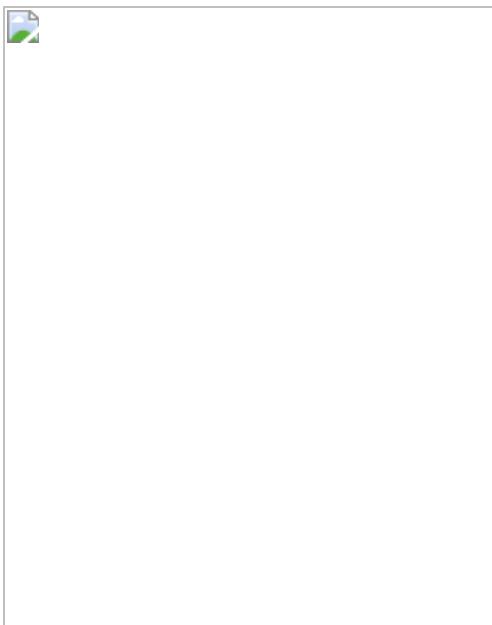


Fig. 3: Phylogenetic histories of 7 patients with $JAK2^{V617F}$ -mutated MPN and clonal evolution.



Fig. 4: Phylogenetic trees, mutation burdens and telomere lengths.



Fig. 5: Clonal fitness and latency to diagnosis.

Data availability

Whole-genome sequencing data in the form of BAM files across all samples reported in this study have been deposited in the European Genome–Phenome Archive (<https://www.ebi.ac.uk/ega/home>) with accession codes [EGAD00001007714](#) (whole-genome sequencing colonies) and [EGAD00001007715](#) (targeted-recapture sequencing). Per patient VCF files containing information on somatic mutations identified are available on Mendeley (doi: 10.17632/hrmxybrd2n.1) .

Code availability

Single-nucleotide substitutions (SNV) were called using the cancer variants through expectation maximization (CaVEMan) algorithm, version 1.13.14 (<https://github.com/cancerit/CaVEMan>). Small insertions and deletions were called using the Pindel algorithm as implemented in the cgpPindel workflow, version 3.2.0 (<https://github.com/cancerit/cgpPindel>). Copy number variants were called using the ASCAT algorithm as implemented in the ascatNgs workflow, version 3.2.0 (<https://github.com/cancerit/ascatNgs>). Mutational signatures analysis was performed using MutationalPatterns v1.10, available on Github (<https://github.com/UMCUGenetics/MutationalPatterns>) and SigProfiler (<https://github.com/AlexandrovLab>). Allele counts at SNV and indel sites were carried out using vafCorrect (<https://github.com/cancerit/vafCorrect>). Telomere lengths were estimated using telomerecat, version 3.2 (<https://github.com/cancerit/telomerecat>). Mutations were mapped to phylogenetic branches using Rtreemut developed for this study (<https://github.com/NickWilliamsSanger/treemut>). Temporal branch lengths and per driver mutation rates were inferred using rtreefit developed for this study (<https://github.com/NickWilliamsSanger/rtreefit>). Simulation of HSC populations and phylogenies with selection were carried out using rsimpop developed for this study (<https://github.com/NickWilliamsSanger/rsimpop>). Other analyses were carried out using custom R scripts available at https://github.com/NickWilliamsSanger/mpn_phylogenies_and_evolution.

References

1. 1.

ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).

2. 2.

Gerstung, M. et al. The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).

3. 3.

Mitchell, T. J. et al. Timing the landmark events in the evolution of clear cell renal cell cancer: TRACERx Renal. *Cell* **173**, 611–623.e17 (2018).

4. 4.

Welch, J. S. et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).

5. 5.

Alexandrov, L. B. et al. Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).

6. 6.

Radivoyevitch, T., Hlatky, L., Landaw, J. & Sachs, R. K. Quantitative modeling of chronic myeloid leukemia: insights from radiobiology. *Blood* **119**, 4363–4371 (2012).

7. 7.

Martincorena, I. et al. Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911–917 (2018).

8. 8.

Lee-Six, H. et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**, 532–537 (2019).

9. 9.

Brunner, S. F. et al. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* **574**, 538–542 (2019).

10. 10.

Moore, L. et al. The mutational landscape of normal human endometrial epithelium. *Nature* **580**, 640–646 (2020).

11. 11.

Martincorena, I. et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).

12. 12.

Yokoyama, A. et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312–317 (2019).

13. 13.

Genovese, G. et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).

14. 14.

Jaiswal, S. et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).

15. 15.

Xie, M. et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).

16. 16.

Young, A. L., Challen, G. A., Birmann, B. M. & Druley, T. E. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat. Commun.* **7**, 12484 (2016).

17. 17.

Abelson, S. et al. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature* **559**, 400–404 (2018).

18. 18.

Vainchenker, W. & Kralovics, R. Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. *Blood* **129**, 667–679 (2017).

19. 19.

Grinfeld, J. et al. Classification and personalized prognosis in myeloproliferative neoplasms. *N. Engl. J. Med.* **379**, 1416–1430 (2018).

20. 20.

Lee-Six, H. et al. Population dynamics of normal human blood inferred from somatic mutations. *Nature* **561**, 473–478 (2018).

21. 21.

Osorio, F. G. et al. Somatic mutations reveal lineage relationships and age-related mutagenesis in human hematopoiesis. *Cell Rep.* **25**, 2308–2316.e4 (2018).

22. 22.

Abascal, F. et al. Somatic mutation landscapes at single-molecule resolution. *Nature* **593**, 405–410 (2021).

23. 23.

Mitchell, E. et al. Clonal dynamics of haematopoiesis across the human lifespan. Preprint at <https://doi.org/10.1101/2021.08.16.456475> (2021).

24. 24.

Nangalia, J. et al. DNMT3A mutations occur early or late in patients with myeloproliferative neoplasms and mutation order influences phenotype. *Haematologica* **100**, 438–442 (2015).

25. 25.

Ortmann, C. A. et al. Effect of mutation order on myeloproliferative neoplasms. *N. Engl. J. Med.* **372**, 601–612 (2015).

26. 26.

Lundberg, P. et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood* **123**, 2220–2228 (2014).

27. 27.

Godfrey, A. L. et al. JAK2V617F homozygosity arises commonly and recurrently in PV and ET, but PV is characterized by expansion of a dominant homozygous subclone. *Blood* **120**, 2704–2707 (2012).

28. 28.

Kahn, J. D. et al. PPM1D-truncating mutations confer resistance to chemotherapy and sensitivity to PPM1D inhibition in hematopoietic cells. *Blood* **132**, 1095–1105 (2018).

29. 29.

Vaziri, H. et al. Evidence for a mitotic clock in human hematopoietic stem cells: loss of telomeric DNA with age. *Proc. Natl Acad. Sci. USA* **91**, 9857–9860 (1994).

30. 30.

Anand, S. et al. Effects of the JAK2 mutation on the hematopoietic stem and progenitor compartment in human myeloproliferative neoplasms. *Blood* **118**, 177–181 (2011).

31. 31.

Chapman, M. S. et al. Lineage tracing of human development through somatic mutations. *Nature* **595**, 85–90 (2021).

32. 32.

de Kanter, J. K. et al. Antiviral treatment causes a unique mutational signature in cancers of transplantation recipients. *Cell Stem Cell* **28**, 1726–1739 (2021).

33. 33.

Titmarsh, G. J. et al. How common are myeloproliferative neoplasms? A systematic review and meta-analysis. *Am. J. Hematol.* **89**, 581–587 (2014).

34. 34.

Mehta, J., Wang, H., Iqbal, S. U. & Mesa, R. Epidemiology of myeloproliferative neoplasms in the United States. *Leuk. Lymphoma* **55**, 595–600 (2014).

35. 35.

Van Egeren, D. et al. Reconstructing the lineage histories and differentiation trajectories of individual cancer cells in myeloproliferative neoplasms. *Cell Stem Cell* **28**, 514–523.e9 (2021).

36. 36.

McKerrell, T. et al. JAK2 V617F hematopoietic clones are present several years prior to MPN diagnosis and follow different expansion

kinetics. *Blood Adv.* **1**, 968–971 (2017).

37. 37.

Hirsch, P. et al. Clonal history of a cord blood donor cell leukemia with prenatal somatic JAK2 V617F mutation. *Leukemia* **30**, 1756–1759 (2016).

38. 38.

Wong, W. H. et al. Engraftment of rare, pathogenic donor hematopoietic mutations in unrelated hematopoietic stem cell transplantation. *Sci. Transl. Med.* **12**, eaax6249 (2020).

39. 39.

Olcaydu, D. et al. A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat. Genet.* **41**, 450–454 (2009).

40. 40.

Hinds, D. A. et al. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. *Blood* **128**, 1121–1128 (2016).

41. 41.

Fleischman, A. G. Inflammation as a driver of clonal evolution in myeloproliferative neoplasm. *Mediators Inflamm.* **2015**, 606819 (2015).

42. 42.

Bick, A. G. et al. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763–768 (2020).

43. 43.

Bao, E. L. et al. Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. *Nature* **586**, 769–775 (2020).

44. 44.

Watson, C. J. et al. The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science* **367**, 1449–1454 (2020).

45. 45.

Arber, D. A. et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405 (2016).

46. 46.

Nielsen, C., Birgens, H. S., Nordestgaard, B. G. & Bojesen, S. E. Diagnostic value of *JAK2* V617F somatic mutation for myeloproliferative cancer in 49 488 individuals from the general population. *Br. J. Haematol.* **160**, 70–79 (2013).

47. 47.

Kiladjian, J. J. et al. Pegylated interferon- α -2a induces complete hematologic and molecular responses with low toxicity in polycythemia vera. *Blood* **112**, 3065–3072 (2008).

48. 48.

Pieri, L. et al. *JAK2V617F* complete molecular remission in polycythemia vera/essential thrombocythemia patients treated with ruxolitinib. *Blood* **125**, 3352–3353 (2015).

49. 49.

Nangalia, J. et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated *JAK2*. *N. Engl. J. Med.* **369**, 2391–2405 (2013).

50. 50.

Ellis, P. et al. Reliable detection of somatic mutations in solid tissues by laser-capture microdissection and low-input DNA sequencing. *Nat. Protoc.* **16**, 841–871 (2021).

51. 51.

Jones, D. et al. cgpCaVEManWrapper: simple execution of CaVEMan in order to detect somatic single nucleotide variants in NGS data. *Curr. Protoc. Bioinforma.* **56**, 15.10.1–15.10.18 (2016).

52. 52.

Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).

53. 53.

Van Loo, P. et al. Allele-specific copy number analysis of tumors. *Proc. Natl Acad. Sci. USA* **107**, 16910–16915 (2010).

54. 54.

Hoang, D. T. et al. MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. *BMC Evol. Biol.* **18**, 11 (2018).

55. 55.

Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).

56. 56.

Tavaré, S. The linear birthdeath process: An inferential retrospective. *Adv. Appl. Probab.* **50**, 253–269 (2018).

Acknowledgements

We thank Cambridge Blood and Stem Cell Biobank, funded by the Cambridge Cancer Centre and Wellcome Trust Cambridge Stem Cell Institute, Wellcome Sanger CASM and DNA pipelines for their assistance; and S. Behjati and C. Harrison for valuable discussion. The study was supported by Cancer Research UK (J.N.), EHA Research Award (J.N.), MPN Research Foundation (J.N.) and the Wellcome Trust (P.J.C., A.R.G. and J.L.). Work in the A.R.G. laboratory is supported by the Wellcome Trust, Bloodwise, Cancer Research UK, the Kay Kendall Leukaemia Fund and the Leukaemia and Lymphoma Society of America. J.N. is a CRUK Clinician Scientist fellow. We thank the patients for their participation in the study.

Author information

Author notes

1. These authors jointly supervised this work: Anthony R. Green, Peter J. Campbell, Jyoti Nangalia

Affiliations

1. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK

Nicholas Williams, Joe Lee, Emily Mitchell, Luiza Moore, James Hewinson, Kevin J. Dawson, Andrew Menzies, Peter J. Campbell & Jyoti Nangalia

2. Wellcome–MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Cambridge, UK

Joe Lee, Emily Mitchell, Anthony R. Green, Peter J. Campbell & Jyoti Nangalia

3. Department of Haematology, University of Cambridge, Cambridge, UK

Emily Mitchell, E. Joanna Baxter, Anthony R. Green, Peter J. Campbell & Jyoti Nangalia

4. Department of Haematology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

Emily Mitchell, Anna L. Godfrey, Anthony R. Green & Jyoti Nangalia

Contributions

J.N., A.R.G. and P.J.C. conceived the study. N.W. performed genomic, phylogenetic and population dynamics analyses with J.N. J.L. assisted with signature, clinical and telomere analyses. E.M. provided genomic data and analyses for normal samples. L.M. assisted with low-input sequencing and mutation signature analysis. A.L.G. assisted with clinical correlation. J.N. and E.J.B. obtained samples. K.J.D. assisted with simulation inferences. A.M. and J.H assisted with computational and laboratory processing pipelines. J.N. directed the study and wrote the manuscript with input from co-authors. All authors reviewed and approved the manuscript.

Corresponding author

Correspondence to [Jyoti Nangalia](#).

Ethics declarations

Competing interests

A patent has been filed by the Wellcome Sanger Institute (inventors N.W. and J.N.; Application number PCT/EP2021/071952) covering somatic mutation identification in the context of tumour contamination of the matched germline sample.

Peer review information

Nature thanks Steven McCarroll, Seishi Ogawa and the other, anonymous reviewers 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 legend. Patient characteristics and somatic mutation fractions in haematopoietic colonies.](#)

a. Patient characteristics. PV, Polycythaemia vera; ET, Essential thrombocythaemia; MF, myelofibrosis; HC, Hydroxycarbamide; IFN, Interferon-alpha; FU, follow-up. *PV diagnosed on red cell mass study. **b.** The distribution of variant allele fractions (VAF) for point mutations pooled across colonies per patient. The mean VAF of individual colonies is shown as red dots. Only autosomal somatic mutations are shown, with those in regions with copy-number aberrations and loss-of-heterozygosity excluded. The plot shows that the colony VAFs are close to 0.5 for the majority.

[Extended Data Fig. 2 Legend: Parallel evolution within phylogenetic trees.](#)

a. Phylogenetic tree of PD5117 depicting 3 separate 9pUPD (UPD, uniparental disomy) acquisitions (blue branches), downstream of *JAK2^{V617F}* (red branch). Below the phylogenetic tree are three B-allele frequency plots showing the regions of 9pUPD in the different clades with vertical red lines showing the boundary of loss of heterozygosity. The event shown on the far right has a distinct breakpoint from the left two events. Blue and green vertical lines show somatic mutations (either prior or subsequent to the UPD event), suggesting that the 9pUPD event depicted in the middle plot occurred first as more mutations have had time to accrue

since the copy number aberration. **b.** Phylogenetic tree of PD5179 depicting two separate 1q+ (orange branches) and 9q- (blue branches) acquisitions. Left plot shows the aggregate VAF of germline single nucleotide polymorphisms (SNP) on Chr1 for samples in the 1q+ major clade versus 1q+ minor clade (left plot). SNPs at a VAF = 2/3 in one clade are at 1/3 in the minor clade, and vice-versa, confirming that different parental chromosomes are amplified in each clade. SNPs in the affected 9q- region also exhibit a clear pattern in VAF (right panel), with VAF = 0.5 for samples in the major 9q- clade but VAF = 0 or 1 for samples in the minor 9q- clade. A proposed model of chr9 copy number changes is shown in the upper right. **c.** Phylogenetic tree of PD4781 depicting two separate *JAK2*^{V617F} acquisitions (red branches) each followed by 9pUPD (blue branches). *JAK2*^{V617F} acquisition occurred on different parental alleles in each instance as SNPs on 9p that have a VAF ~1 for samples in the major *JAK2*-mutant clade (horizontal bar coloured red) have a VAF ~0 in samples from the minor *JAK2*-mutant clade (horizontal bar coloured blue) and vice-versa.

Extended Data Fig. 3 Legend: Mutation rates and burden following driver mutation acquisition.

a. Mutation rate estimates for wildtype and different mutant clades within patients. Mutation acquisition is modelled using Poisson modelling taking into account the timing of transition from wildtype to driver mutation acquisition within mutant clades and an excess mutation rate earlier in life ([Methods](#)). Patients and genotypes of clades are shown on the left together with colony number for each clade (N). Wildtype (WT) clades are shown in grey bars, *JAK2*-mutated clades are shown in red and other mutant clades are shown in yellow. The cohort wide estimate for the mutation rate in WT colonies is shown by the dotted black vertical line at the top. * $P < 0.05$, ** $P < 0.01$ (** also significant after multiple hypothesis testing; Bonferroni adjusted, two-sided test). Significantly different mutation rates between clades are highlighted only for those significant by both Poisson and Negative Binomial modelling of mutation rates ([Methods](#)). Average mutation burdens are shown to the right for the different timepoints of sampling. **b.** Non-parametric comparison of mutation burdens in wildtype versus mutant colonies using limma's rankSumTestWithCorrelation. This accounts for the non-independence of data in mutant colonies but does not

account for the timing of driver mutation acquisition. *indicates significance at $P < 0.05$ following Bonferroni multiple hypothesis correction.

Extended Data Fig. 4 Legend: Mutational signatures and telomeres.

a. Signature contributions of SBS1, SBS5, SBS19 and SBS32 on a per-patient/per-clade basis. Single base substitution mutational signature 5 (SBS5), thought to represent a time-dependent mutational process active in all tissues, was the predominant mutational process in colonies . **b.** The proportion of C>T transitions at CpG dinucleotides across WT, *JAK2*-mutated and colonies with other driver mutations. * $P < 0.05$, ** $P < 0.01$ (** also significant after multiple hypothesis testing; Chi-square test). **c.** The relationship between ‘sharedness’ (see [Methods](#)) and telomere length across all phylogenetic trees shows that telomeres shorten in line with increased phylogenetic ‘sharedness’ in keeping with the increased cell divisions during clonal expansion. **d.** The heritability of telomere length, that is, whether closely related colonies had more similar telomere lengths compared to more distantly related colonies, is assessed using Pagel’s Lambda and Blomberg’s K, with both values in the vicinity of 1 or above, suggesting that telomere length variation across colonies in a phylogenetic tree follows the expected covariance based on phylogenetic relationship. Power for PD5147 is limited because there is little difference in ‘sharedness’ in the mutant colonies. **e.** The modelled reduction in telomere length per additional stem cell division in *JAK2* mutant clades is shown per patient, with a cohort wide estimate of -57.4bp (-74.2, -40.59 95% CI). See Supplementary Note [7](#) for further interpretation.

Extended Data Fig. 5 legend: Time based trees and timing driver mutation acquisition.

a. Time-based phylogenetic trees. Different coloured branches identify separate clades alongside light blue wild type colonies. The vertical axis represents age post conception with treatment received alongside. Driver mutations are depicted in the middle of the branches but may have occurred

at any point between the start and end of the branches. Given the uncertainties in the exact ages at the starts and ends of the branches due to modelling branch lengths from mutation count data ([Methods](#)), the credibility intervals for the ends of the branches harbouring driver mutations are shown as black lines and also in b-c. **b.** Each horizontal grey box represents an individual patient from birth until the last colony sampling timepoint. The time before birth is represented on an expanded scale and is shaded pink. Within each grey box is shown the range of ages during which driver mutation and copy number aberrations are estimated to have occurred. The start and ends of each coloured box represent the median lower and upper bounds of time estimates corresponding to the start and end of the shared branches harbouring driver mutations. Thus, the upper bounds (right edge of the coloured boxes) represent the latest time by which mutation acquisition is estimated to have occurred from phylogenetic analysis. Black lines show the 95% credibility intervals for the start and end of the branches carrying the drivers. Mutation timings are inferred from a model where mutation accumulation within branches follows a Poisson distribution but were not substantially different when using a Negative Binomial model. Diamonds show age at diagnosis. **c.** Raw data from a-b is shown with 95% CI intervals around the estimated ages of the starts and ends of branches harbouring driver mutations for different patients, together with adjusted SNV counts for branches.

[Extended Data Fig. 6 legend: Estimates of clonal expansion rates in patients.](#)

a. The figures shows the smoothed posterior density distribution of the selection coefficient (proportion additional growth per year) vs driver timing for all analysed clades from population simulations and approximate Bayesian computation (ABC). Marginal distributions are also shown. The prior distribution for driver timing is clade dependent and is largely determined by the mutation count at the start and end of the associated branch. Both clonal fractions and lineages-through-time were used as summary statistics in the approximate Bayesian computation for estimates of selection. Main plots show driver mutations acquired after birth, and driver mutations pre-birth are shown within the black box, taking into account driver mutation acquisition during a time when the background

stem cell population size is modelled to be growing. **b.** Data from **a.** in tabular format. Here, selection coefficients have been converted to clonal expansion (median growth % per year, Selection). The ABC approach gives alternative estimates for ages of driver mutation acquisition as shown. N depicts the number of simulations per clade. Clones with sufficient immediate descendants (>5 coalescences) were included for estimates of selection. **c.** Comparison of estimates of selection of mutant clades (each labelled by patient ID and driver mutation) from ABC versus Phylofit. The grey lines show 95% credibility intervals for estimates from each approach. Correlation coefficient $r = 0.96$. Note, that the PD5182 and PD5847 *in-utero* DNMT3A expansions from panel **a.** are not shown because, only the ABC approach, and not Phylofit, allowed for modelling selection against a growing background population.

Extended Data Fig. 7 legend: Aberrant cell fractions in bulk blood samples and validation of selection estimates.

a. Plots showing aberrant cell fraction (ACF) in colonies and bulk longitudinal mature blood cell samples. Colony samples were derived from peripheral blood (red dots) or bone marrow (orange dots, in PD5182 and PD5847) mononuclear cells. Bulk mature blood cell samples comprised mostly peripheral blood granulocytes (black dots) and occasionally, bone marrow derived (grey dots) granulocytes (in PD5847, PD6629) or mononuclear cells (in PD5182), and whole blood (brown dots, in PD9478, PD6629). ACF in colonies is the clonal fraction proportion of all colonies. In bulk samples, ACF is calculated as twice the mean VAF of variants that map to the shared ancestral branch of the clone. The x-axis is patient age at sample timepoints. Lines depict the inferred ACF trajectories from the top 0.01% of simulations from approximate Bayesian computation. Black lines, median ACF; grey lines, 95% CI; dotted line, inferred future growth trajectory beyond the sampling time using the growth rate S and accounting for a sigmoid clonal trajectory as clonal dominance is approached. **b.** 95% confidence intervals for the difference in parent branch and aggregate descendant daughter branch ACFs from phylogenetic tree clades. Confidence intervals are calculated assuming a normal sampling distribution of aggregate mutant read fractions for each branch. Diamonds indicate those recapture samples closest to the colony sampling.* denotes

interferon treatment at time of sampling. **c.** Comparison of estimates of selection coefficients for clades with single driver mutations using Phylofit fitted using the branching pattern within the tree (lineage through time, LTT) and ACF (horizontal axis), versus selection coefficients estimates using just the branching pattern of the tree (LTT) and no ACF (vertical axis) to identify clades that show early rapid branching, but smaller than expected final clonal fractions. 95% credibility intervals for selection coefficients are shown as grey lines and the corresponding median estimates as black dots. Possible early faster expansion are seen in two *in utero* mutated-DNMT3A clades (PD5182 and PD5847) and the JAK2^{V617F} clade in PD5163 prior to Interferon therapy.

Supplementary information

Supplementary Information

This file contains Supplementary Notes 1–8, including 12 figures, Table 1 and references.

Reporting Summary

Peer Review File

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Williams, N., Lee, J., Mitchell, E. *et al.* Life histories of myeloproliferative neoplasms inferred from phylogenies. *Nature* **602**, 162–168 (2022).
<https://doi.org/10.1038/s41586-021-04312-6>

- Received: 16 October 2020
- Accepted: 06 December 2021
- Published: 20 January 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04312-6>

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/s41586-021-04312-6>

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

- Matters Arising
- [Published: 02 February 2022](#)

No freshwater-filled glacial Arctic Ocean

- [Robert F. Spielhagen](#) ORCID: [orcid.org/0000-0001-9740-667X¹](https://orcid.org/0000-0001-9740-667X),
- [Jan C. Scholten](#) ORCID: [orcid.org/0000-0002-6940-9053²](https://orcid.org/0000-0002-6940-9053),
- [Henning A. Bauch](#) ORCID: [orcid.org/0000-0003-4599-6916^{1,3}](https://orcid.org/0000-0003-4599-6916) &
- [Anton Eisenhauer¹](#)

[Nature](#) volume 602, pages E1–E3 (2022)

- 862 Accesses
- 1 Citations
- 12 Altmetric
- [Metrics details](#)

Subjects

- [Palaeoceanography](#)
- [Palaeoclimate](#)

[Matters Arising](#) to this article was published on 02 February 2022

The [Original Article](#) was published on 03 February 2021

arising from W. Geibert et al. *Nature* <https://doi.org/10.1038/s41586-021-03186-y> (2021)

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: Bathymetric map of the Nordic Seas.

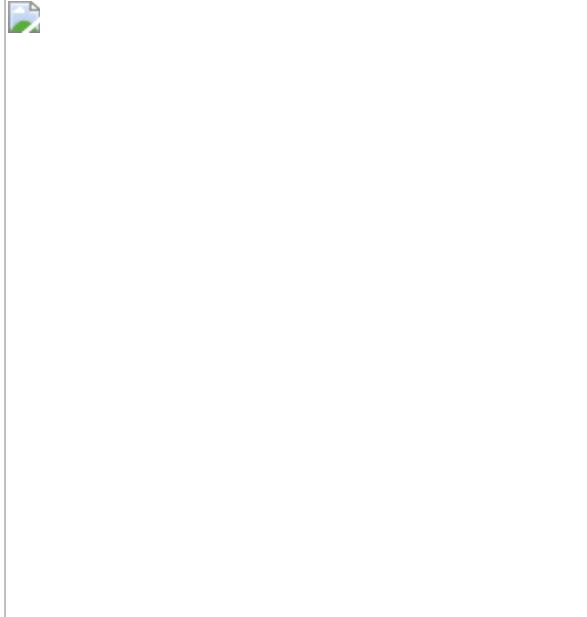
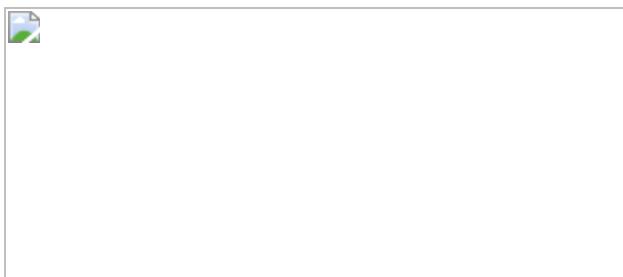


Fig. 2: Published stable oxygen isotope records from Fram Strait and the Nordic Seas sediment cores^{6,11,13,14}.



Data availability

All data shown in this contribution have been published previously and are fully referenced.

References

1. 1.

Geibert, W., Matthiessen, J., Stimac, I., Wollenburg, J. & Stein, R. Glacial episodes of a freshwater Arctic Ocean covered by a thick ice shelf. *Nature* **590**, 97–102 (2021).

2. 2.

Eisenhauer, A., et al. in *Geological History of the Polar Oceans: Arctic Versus Antarctic* (eds Bleil, U. & Thiede, J.) 475–487 (Kluwer Academic Publishers, 1990).

3. 3.

Scholten, J. C., Botz, R., Paetsch, H. & Stoffers, P. $^{230}\text{Th}_{\text{ex}}$ flux into Norwegian-Greenland Sea sediments: Evidence for lateral sediment transport during the past 300,000 years. *Earth Planet. Sci. Lett.* **121**, 111–124 (1994).

4. 4.

Scholten, J., Botz, R., Paetsch, H., Stoffers, P. & Weinelt, M. High resolution uranium-series dating of Norwegian-Greenland Sea sediments: ^{230}Th vs. ^{18}O . *Mar. Geol.* **121**, 77–85 (1994).

5. 5.

McManus, J. F., Anderson, R. F., Broecker, W. S., Fleisher, M. Q. & Higgins, S. M. Radiometrically determined sedimentary fluxes in the sub-polar North Atlantic during the last 140,000 years. *Earth Planet. Sci. Lett.* **155**, 29–43 (1998).

6. 6.

Spielhagen, R. F., et al Arctic Ocean deep-sea record of northern Eurasian ice sheet history. *Quat. Sci. Rev.* **23**, 1455–1483 (2004).

7. 7.

Kaste, J. M. & Baskaran, M. in *Handbook of Environmental Isotope Geochemistry*, Vol. 1 (ed. Baskaran, M.) 61–85 (Springer, 2012).

8. 8.

McHargue, L. R., Jull, A. J. T. & Cohen, A. Measurement of ^{10}Be from Lake Malawi (Africa) drill core sediments and implications for geochronology. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **303**, 110–119 (2011).

9. 9.

Sproson, A. D. et al. Beryllium isotopes in sediments from Lake Maruwan Oike and Lake Skallen, East Antarctica, reveal substantial glacial discharge during the late Holocene. *Quat. Sci. Rev.* **256**, 106841 (2021).

10. 10.

Lisiecki, L. E. & Stern, J. V. Regional and global benthic $\delta^{18}\text{O}$ stacks for the last glacial cycle. *Paleoceanogr.* **31**, 1368–1394 (2016).

11. 11.

Fronval, T. & Jansen, E. Eemian and early Weichselian (140–60 ka) paleoceanography and paleoclimate in the Nordic seas with comparisons to Holocene conditions. *Paleoceanogr.* **12**, 443–462 (1997).

12. 12.

Sarnthein, M., et al. Variations in Atlantic surface ocean paleoceanography, 50°–80°N: a time-slice record of the last 30,000 years. *Paleoceanogr.* **10**, 1063–1094 (1995).

13. 13.

Bauch, H. A., Erlenkeuser, H., Grootes, P. M. & Jouzel, J. Implications of stratigraphic and paleoclimatic records of the last interglaciation from the Nordic seas. *Quat. Res.* **46**, 260–269 (1996).

14. 14.

Didié, C. & Bauch, H. A. Implications of upper Quaternary stable isotope records of marine ostracodes and benthic foraminifers for paleoecological and paleoceanographical investigations. *AGU Monogr. Ser.* **131**, 279–299 (2002).

15. 15.

Hoffmann, S. S., McManus, J. F., Curry, W. B. & Brown-Leger, L. S. Persistent export of ^{231}Pa from the deep central Arctic Ocean over the past 35,000 years. *Nature* **497**, 603–606 (2013).

16. 16.

Teller, J. T., Leverington, D. W. & Mann, J. D. Freshwater outbursts to the oceans from glacial Lake Agassiz and their role in climate change during the last deglaciation. *Quat. Sci. Rev.* **21**, 879–887 (2002).

17. 17.

Tarasov, L. & Peltier, W. R. Arctic freshwater forcing of the Younger Dryas cold reversal. *Nature* **435**, 662–665 (2005).

18. 18.

Condron, A. & Winsor, P. Meltwater routing and the Younger Dryas. *Proc. Natl Acad. Sci.* **109**, 19928–19933 (2012).

19. 19.

Stein, R., et al The last deglaciation event in the eastern central Arctic Ocean. *Science* **264**, 692–696 (1994).

20. 20.

Author information

Affiliations

1. GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany

Robert F. Spielhagen, Henning A. Bauch & Anton Eisenhauer

2. Institute for Geosciences, Kiel University, Kiel, Germany

Jan C. Scholten

3. Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

Henning A. Bauch

Contributions

R.F.S. initiated the project. R.F.S. and J.C.S. collected the data from published sources and wrote the manuscript. All authors contributed to data interpretation and to refining the manuscript.

Corresponding author

Correspondence to [Robert F. Spielhagen](#).

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.

Supplementary information

Supplementary Information

This file contains Supplementary Figs. 1–3 and refs. 1–20. Supplementary Fig. 1 shows a simplified cross section over the Arctic Ocean and the Nordic Seas and consequences of the decay of a thick ice sheet covering this area. Supplementary Fig. 2 shows a map of the northern North Atlantic with core locations. Supplementary Fig. 3 shows oxygen isotope records of cores in the North Atlantic.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Spielhagen, R.F., Scholten, J.C., Bauch, H.A. *et al.* No freshwater-filled glacial Arctic Ocean. *Nature* **602**, E1–E3 (2022).
<https://doi.org/10.1038/s41586-021-04089-8>

- Received: 10 May 2021
- Accepted: 01 October 2021
- Published: 02 February 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04089-8>

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

Further reading

- [Reply to: No freshwater-filled glacial Arctic Ocean](#)

- Walter Geibert
- Jens Matthiessen
- Ruediger Stein

Nature (2022)

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04089-8>

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

- Matters Arising
- [Published: 02 February 2022](#)

Reply to: No freshwater-filled glacial Arctic Ocean

- [Walter Geibert](#) ORCID: [orcid.org/0000-0001-8646-2334¹](https://orcid.org/0000-0001-8646-2334),
- [Jens Matthiessen](#) ORCID: [orcid.org/0000-0002-6952-2494¹](https://orcid.org/0000-0002-6952-2494),
- [Jutta Wollenburg](#) ORCID: [orcid.org/0000-0002-8169-8310¹](https://orcid.org/0000-0002-8169-8310) &
- [Ruediger Stein](#)^{1,2}

Nature volume 602, pages E4–E6 (2022)

- 378 Accesses
- 1 Altmetric
- [Metrics details](#)

Subjects

- [Cryospheric science](#)
- [Palaeoceanography](#)
- [Palaeoclimate](#)

The [Original Article](#) was published on 02 February 2022

replying to R. F. Spielhagen et al. *Nature* <https://doi.org/10.1038/s41586-021-04089-8> (2021)

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.

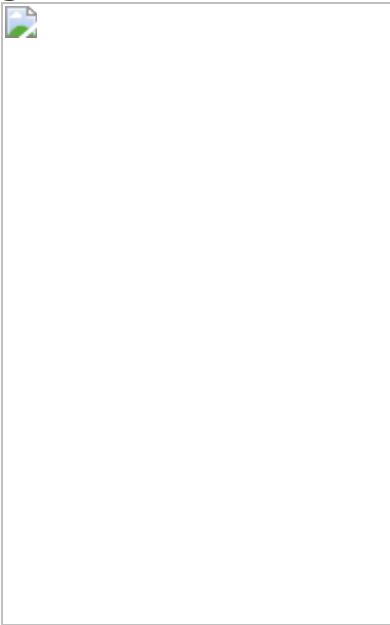
Additional access options:

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

Fig. 1: $^{230}\text{Th}_{\text{ex}}$ data for four cores from the Nordic Seas.



Fig. 2: $\delta^{18}\text{O}$ in diatoms versus $\delta^{18}\text{O}$ in foraminifera.



Data availability

All data shown here have been published previously and are fully referenced.

References

1. 1.

Spielhagen, R. F., Scholten, J. C., Bauch, H. A. & Eisenhauer, A. No freshwater-filled glacial Arctic Ocean. *Nature* <https://doi.org/10.1038/s41586-021-04089-8> (2021).

2. 2.

Geibert, W., Matthiessen, J., Stimac, I., Wollenburg, J. & Stein, R. Glacial episodes of a freshwater Arctic Ocean covered by a thick ice shelf. *Nature* **590**, 97–102 (2021).

3. 3.

Jakobsson, M. et al. Evidence for an ice shelf covering the central Arctic Ocean during the penultimate glaciation. *Nat. Commun.* **7**, 10365 (2016).

4. 4.

Blischke, A. et al. The Jan Mayen microcontinent's Cenozoic stratigraphic succession and structural evolution within the NE-Atlantic. *Mar. Pet. Geol.* **103**, 702–737 (2019).

5. 5.

Paetsch, H. *Sedimentation im Europäischen Nordmeer: Radioisotopische, Geochemische und Tonmineralogische Untersuchungen Spätquartärer Ablagerungen*. PhD thesis, Christian-Albrechts-Universität Kiel (1991).

6. 6.

Scholten, J., Botz, R., Paetsch, H., Stoffers, P. & Weinelt, M. High-resolution uranium-series dating of Norwegian-Greenland Sea sediments: ^{230}Th vs. $\delta^{18}\text{O}$ stratigraphy. *Mar. Geol.* **121**, 77–85 (1994).

7. 7.

Werner, F. & Winn, K. *Documentation of sediment core GIK17728-2* <https://doi.org/10.1594/PANGAEA.720640> (PANGAEA, 2009);

supplement to Grobe, H. et al. The GIK-Archive of sediment core radiographs with documentation. *Earth Syst. Sci. Data* **9**, 969–976 (2017).

8. 8.

Brown, E. T. et al. Beryllium isotope geochemistry in tropical river basins. *Geochim. Cosmochim. Acta* **56**, 1607–1624 (1992).

9. 9.

Aldahan, A., Haiping, Y. & Possnert, G. Distribution of beryllium between solution and minerals (biotite and albite) under atmospheric conditions and variable pH. *Chem. Geol.* **156**, 209–229 (1999).

10. 10.

Frank, M. Radiogenic isotopes: tracers of past ocean circulation and erosional input. *Rev. Geophys.* **40**, 1-1-1-38 (2002).

11. 11.

Maier, E. et al. North Pacific freshwater events linked to changes in glacial ocean circulation. *Nature* **559**, 241–245 (2018).

12. 12.

Greco, M., Meilland, J., Zamelczyk, K., Rasmussen, T. L. & Kucera, M. The effect of an experimental decrease in salinity on the viability of the subarctic planktonic foraminifera *Neogloboquadrina incompta*. *Polar Res.* **39**, <https://doi.org/10.33265/polar.v39.3842> (2020).

13. 13.

Kemp, A. E. S. et al. Production of giant marine diatoms and their export at oceanic frontal zones: implications for Si and C flux from stratified oceans. *Global Biogeochem. Cycles* **20**, <https://doi.org/10.1029/2006GB002698> (2006).

14. 14.

Rasmussen, T. L., Oppo, D. W., Thomsen, E. & Lehman, S. J. Deep sea records from the southeast Labrador Sea: ocean circulation changes and ice-rafting events during the last 160,000 years. *Paleoceanography* **18**, 1018–1029 (2003).

15. 15.

Böhm, E. et al. Strong and deep Atlantic meridional overturning circulation during the last glacial cycle. *Nature* **517**, 73–76 (2015).

16. 16.

Rasmussen, T. L. & Thomsen, E. The role of the North Atlantic Drift in the millennial timescale glacial climate fluctuations. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **210**, 101–116 (2004).

17. 17.

Wollenburg, J. E., Kuhnt, W. & Mackensen, A. Changes in Arctic Ocean paleoproductivity and hydrography during the last 145 kyr: the benthic foraminiferal record. *Paleoceanography* **16**, 65–77 (2001).

18. 18.

Seidenkrantz, M.-S. et al. Two-step deglaciation at the oxygen isotope stage 6/5E transition: the Zeifen-Kattegat climate oscillation. *Quat. Sci. Rev.* **15**, 63–75 (1996).

Author information

Affiliations

1. Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

Walter Geibert, Jens Matthiessen, Jutta Wollenburg & Ruediger Stein

2. MARUM – Center for Marine Environmental Sciences and Faculty of Geosciences, University of Bremen, Bremen, Germany

Ruediger Stein

Contributions

W.G. wrote this reply based on published data. All authors contributed to interpretation of data, selection of references and writing of the text. I. Stimac, who was an author of the original article, was not involved in this reply because her contribution to the original article (methods and results sections of the paper) was not discussed as part of the comment. She still supports the published article.

Corresponding author

Correspondence to [Walter Geibert](#).

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.

Supplementary information

Supplementary Information

This file contains the Supplementary Discussions, Fig. 1 and refs. 1–17.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Geibert, W., Matthiessen, J., Wollenburg, J. *et al.* Reply to: No freshwater-filled glacial Arctic Ocean. *Nature* **602**, E4–E6 (2022).
<https://doi.org/10.1038/s41586-021-04090-1>

- Published: 02 February 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04090-1>

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

[Glacial episodes of a freshwater Arctic Ocean covered by a thick ice shelf](#)

- Walter Geibert
- Jens Matthiessen
- Ruediger Stein

Article 03 Feb 2021

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04090-1>

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

- Matters Arising
- [Published: 02 February 2022](#)

Rainfall an unlikely factor in Kīlauea's 2018 rift eruption

- [Michael P. Poland](#) ORCID: orcid.org/0000-0001-5240-6123¹,
- [Shaul Hurwitz](#)²,
- [James P. Kauahikaua](#) ORCID: orcid.org/0000-0003-3777-503X³,
- [Emily K. Montgomery-Brown](#) ORCID: orcid.org/0000-0001-6787-2055¹,
- [Kyle R. Anderson](#) ORCID: orcid.org/0000-0001-8041-3996²,
- [Ingrid A. Johanson](#)³,
- [Matthew R. Patrick](#) ORCID: orcid.org/0000-0002-8042-6639³ &
- [Christina A. Neal](#)⁴

[Nature](#) volume 602, pages E7–E10 (2022)

- 259 Accesses
- 8 Altmetric
- [Metrics details](#)

Subjects

- [Natural hazards](#)
- [Volcanology](#)

[Matters Arising](#) to this article was published on 02 February 2022

The [Original Article](#) was published on 22 April 2020

arising from J. I. Farquharson & F. Amelung, *F. Nature*
<https://doi.org/10.1038/s41586-020-2172-5> (2020)

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: Monitoring data from Kīlauea Volcano spanning January–April 2018.

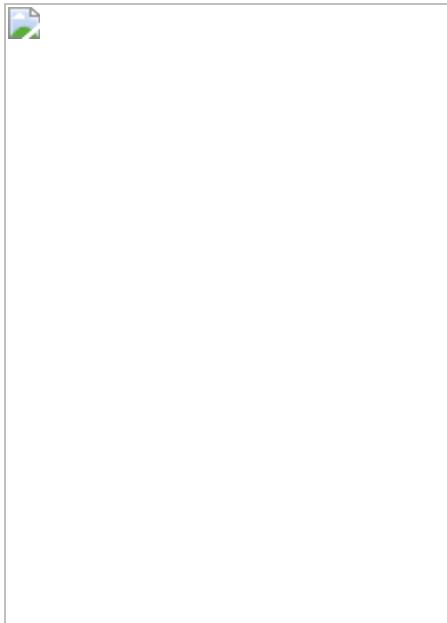
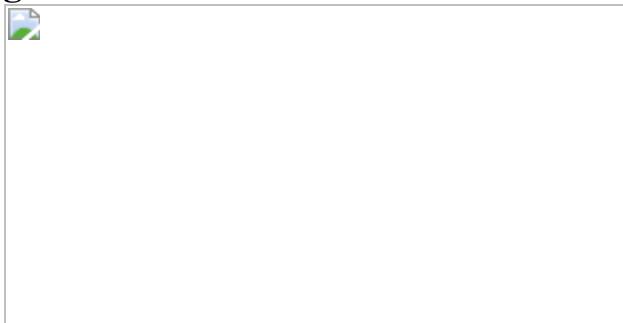


Fig. 2: Rainfall on the Island of Hawai'i.



Data availability

GNSS datasets analysed during the current study are available from the UNAVCO Geodetic Facility for the Advancement of Geosciences (GAGE) facility, <https://www.unavco.org/data/gps-gnss/gps-gnss.html>, and rainfall data are available from the National Oceanic and Atmospheric Administration National Centers for Environmental Information, <https://www.ncdc.noaa.gov/cdo-web/search>.

References

1. 1.
Farquharson, J. I. & Amelung, F. Extreme rainfall triggered the 2018 rift eruption at Kīlauea Volcano. *Nature* **580**, 491–495 (2020).
2. 2.
Neal, C. A. et al. The 2018 rift eruption and summit collapse of Kīlauea Volcano. *Science* **363**, 367–374 (2019).
3. 3.
Farquharson, J. I. & Amelung, F. Author Correction: Extreme rainfall triggered the 2018 rift eruption at Kīlauea Volcano. *Nature* **582**, E3 (2020).
4. 4.
Patrick, M. R. et al. The cascading origin of the 2018 Kīlauea eruption and implications for future forecasting. *Nat. Commun.* **11**, 5646 (2020).
5. 5.
Orr, T. R. et al. in *Hawaiian Volcanoes: From Source to Surface* (eds Carey, R. et al.) Vol. 208, Ch. 18, 393–420 (AGU, 2015).
6. 6.

Patrick, M., Orr, T., Anderson, K. R. & Swanson, D. Eruptions in sync: improved constraints on Kīlauea Volcano's hydraulic connection. *Earth Planet. Sci. Lett.* **507**, 50–61 (2019).

7. 7.

Hazard Notification System (HANS) for Volcanoes (USGS, accessed 3 March 2021); https://volcanoes.usgs.gov/vhp/archive_search.html

8. 8.

Patrick, M. R., Anderson, K. R., Poland, M. P., Orr, T. R. & Swanson, D. A. Lava lake level as a gauge of magma reservoir pressure and eruptive hazard. *Geology* **43**, 831–834 (2015).

9. 9.

Emter, D. in *Tidal Phenomena* (eds Wilhelm, H. et al.) 293–309 (Springer, 1997).

10. 10.

Manga, M. When it rains, lava pours. *Nature* **580**, 457–458 (2020).

11. 11.

Chen, Y., Ebert, E. E., Walsh, K. J. E. & Davidson, N. E. Evaluation of TRMM 3B42 precipitation estimates of tropical cyclone rainfall using PACRAIN data. *J. Geophys. Res.* **118**, 2184–2196 (2013).

12. 12.

Giambelluca, T. W. et al. Online rainfall atlas of Hawai'i. *Bull. Am. Meteorol. Soc.* **94**, 313–316 (2013).

13. 13.

Caruso, S. J. & Businger, S. Subtropical cyclogenesis over the central north Pacific. *Weather Forecast.* **21**, 192–205 (2006).

14. 14.

Nugent, A. D. et al. Fire and rain: the legacy of Hurricane Lane in Hawai'i. *Bull. Am. Meteorol. Soc.* **101**, E954–E957 (2020).

15. 15.

Climate Data Online (NOAA, accessed 3 March 2021);
<https://www.ncdc.noaa.gov/cdo-web/search>

16. 16.

Hurwitz, S. et al. in *The 2008-2018 Summit Lava Lake at Kīlauea Volcano, Hawai'i* (eds Patrick, M. et al.) Ch. F, US Geological Survey Professional Paper 1867 <https://doi.org/10.3133/pp1867F> (USGS, 2021).

17. 17.

Global Volcanism Program (Smithsonian Institution, accessed 3 March 2021); <https://volcano.si.edu/volcano.cfm?vn=332010&vtab=Eruptions>

18. 18.

Violette, S. et al. Can rainfall trigger volcanic eruptions? A mechanical stress model of an active volcano: ‘Piton de la Fournaise’, Reunion Island. *Terra Nova* **13**, 18–24 (2001).

19. 19.

Sawi, T. M. & Manga, M. Revisiting short-term earthquake triggered volcanism. *Bull. Volcanol.* **80**, 57 (2018).

Acknowledgements

We are grateful for comments from S. Ingebritsen and L. Mastin.

Author information

Affiliations

1. U.S. Geological Survey, Cascades Volcano Observatory, Vancouver, WA, USA

Michael P. Poland & Emily K. Montgomery-Brown

2. U.S. Geological Survey, California Volcano Observatory, Moffett Field, CA, USA

Shaul Hurwitz & Kyle R. Anderson

3. U.S. Geological Survey, Hawaiian Volcano Observatory, Hilo, HI, USA

James P. Kauahikaua, Ingrid A. Johanson & Matthew R. Patrick

4. U.S. Geological Survey, Volcano Science Center, Anchorage, AK, USA

Christina A. Neal

Contributions

M.P.P., E.K.M.-B., K.R.A., I.A.J. and M.R.P. contributed to magma pressure descriptions and calculations. S.H., J.P.K. and E.K.M.-B. contributed to rainfall calculations and pore pressure impacts. C.A.N. contributed hazard assessment and communication information. M.P.P. and M.R.P. contributed to statistical arguments. M.P.P. coordinated manuscript preparation, in which all authors engaged.

Corresponding author

Correspondence to [Michael P. Poland](#).

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.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Poland, M.P., Hurwitz, S., Kauahikaua, J.P. *et al.* Rainfall an unlikely factor in Kīlauea's 2018 rift eruption. *Nature* **602**, E7–E10 (2022).
<https://doi.org/10.1038/s41586-021-04163-1>

- Received: 05 June 2020
- Accepted: 19 October 2021
- Published: 02 February 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04163-1>

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/s41586-021-04163-1>

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

- Matters Arising
- [Published: 02 February 2022](#)

Reply to: Rainfall an unlikely factor in Kīlauea's 2018 rift eruption

- [Jamie I. Farquharson](#) ORCID: [orcid.org/0000-0003-4933-2607¹](https://orcid.org/0000-0003-4933-2607) &
- [Falk Amelung¹](#)

Nature volume **602**, pages E11–E14 (2022)

- 175 Accesses
- 1 Altmetric
- [Metrics details](#)

Subjects

- [Natural hazards](#)
- [Volcanology](#)

The [Original Article](#) was published on 02 February 2022

replying to M. P. Poland et al. *Nature* <https://doi.org/10.1038/s41586-020-2172-5> (2022)

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.

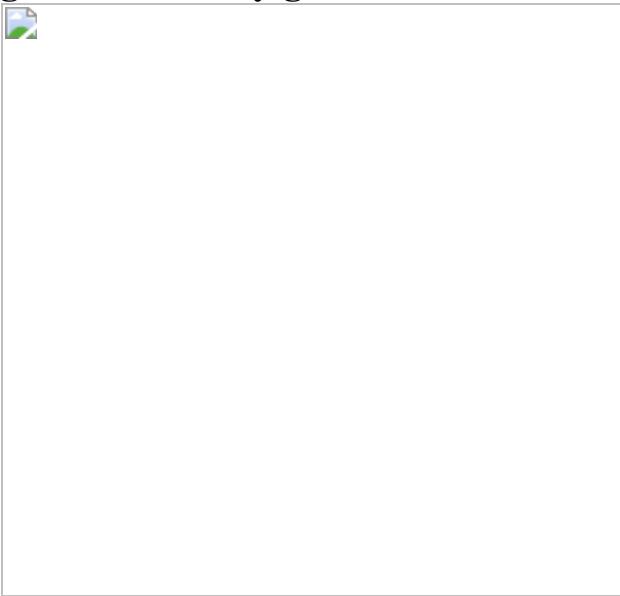
Additional access options:

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

Fig. 1: Pre-eruptive rainfall data.



Fig. 2: Precursory ground deformation at Kīlauea.



Data availability

All data are open source. Satellite-derived rainfall data (TRMM and GPM satellite data) are available from the NASA (National Aeronautics and Space Administration) EarthData Goddard Earth Sciences Data and Information Services Center portal (<https://doi.org/10.5067/TRMM/TMPA/3H/7>). Rainfall gauge data are available from the National Oceanic and Atmospheric Administration's National Centers for Environmental Information climate data portal (<https://www.ncdc.noaa.gov/cdo->

web/datasets/GHCND/stations/GHCND:USC00511303/detail). GPS data are available from the Nevada Geodetic Laboratory (<http://geodesy.unr.edu/NGLStationPages/stations/>). Sentinel-1 ascending- and descending-track SAR acquisitions were obtained through UNAVCO's Seamless SAR Archive (<https://doi.org/10.5194/isprsarchives-XL-1-65-2014>). Derived time series products of Kīlauea are available at <https://doi.org/10.5281/zenodo.3944709> and <https://doi.org/10.5281/zenodo.3957859>.

Code availability

Code required for data access, analysis and display is available, in Jupyter Notebook format, at
https://github.com/jifarquharson/Farquharson_Amelung_2020_Kilauea-Nature/blob/master/Farquharson_Amelung_Kilauea_Supplemental_2.ipynb (Fig. 1) and
https://github.com/jifarquharson/Farquharson_Amelung_2020_Kilauea-Nature/blob/master/Farquharson_Amelung_Kilauea_Supplemental_1.ipynb (Fig. 2).

References

1. 1.
Poland, M. P. et al. Rainfall an unlikely factor in Kīlauea's 2018 rift eruption. *Nature* <https://doi.org/10.1038/s41586-020-2172-5> (2022).
2. 2.
Farquharson, J. I. & Amelung, F. Extreme rainfall triggered the 2018 rift eruption at Kīlauea Volcano. *Nature* **580**, 491–495 (2020).
3. 3.
Farquharson, J. I. & Amelung, F. Author Correction: Extreme rainfall triggered the 2018 rift eruption at Kīlauea Volcano. *Nature* **582**, E3 (2020).

4. 4.

Dzurisin, D. Influence of fortnightly earth tides at Kilauea Volcano, Hawaii. *Geophys. Res. Lett.* **7**, 925–928 (1980).

5. 5.

Lipman, P. W., Lockwood, J. P., Okamura, R. T., Swanson, D. A. & Yamashita, K. M. *Ground Deformation Associated with the 1975 Magnitude-7.2 Earthquake and Resulting Changes in Activity of Kilauea Volcano, Hawaii* (1985).

6. 6.

Poland, M. P., Sutton, A. J. & Gerlach, T. M. Magma degassing triggered by static decompression at Kīlauea Volcano, Hawai‘i. *Geophys. Res. Lett.* **36**, L16306 (2009).

7. 7.

Orr, T. R., Thelen, W. A., Patrick, M. R., Swanson, D. A. & Wilson, D. C. Explosive eruptions triggered by rockfalls at Kīlauea volcano, Hawai‘i. *Geology* **41**, 207–210 (2013).

8. 8.

Volcano Hazards Program FAQs (USGS, 2011);
https://volcanoes.usgs.gov/vsc/file_mngr/file-153/FAQs.pdf

9. 9.

Oosterbaan, R. J. in **Vol.** **16** 175–224 (IILRI, 1994).

10. 10.

Chen, Y., Ebert, E. E., Walsh, K. J. E. & Davidson, N. E. Evaluation of TRMM 3B42 precipitation estimates of tropical cyclone rainfall using PACRAIN data. *J. Geophys. Res. Atmos.* **118**, 2184–2196 (2013).

11. 11.

Bagnardi, M. et al. Gravity changes and deformation at Kīlauea Volcano, Hawaii, associated with summit eruptive activity, 2009–2012. *J. Geophys. Res. Solid Earth* **119**, 7288–7305 (2014).

12. 12.

Patrick, M. R., Anderson, K. R., Poland, M. P., Orr, T. R. & Swanson, D. A. Lava lake level as a gauge of magma reservoir pressure and eruptive hazard. *Geology* **43**, 831–834 (2015).

13. 13.

Anderson, K. R., Poland, M. P., Johnson, J. H. & Miklius, A. in *Hawaiian Volcanoes* 229–250 (American Geophysical Union, 2015).

14. 14.

Parfitt, L. & Wilson, L. *Fundamentals of Physical Volcanology* (Wiley, 2009).

15. 15.

Wauthier, C., Roman, D. C. & Poland, M. P. Modulation of seismic activity in Kīlauea's upper East Rift Zone (Hawai'i) by summit pressurization. *Geology* **47**, 820–824 (2019).

16. 16.

Frazier, A. G. & Giambelluca, T. W. Spatial trend analysis of Hawaiian rainfall from 1920 to 2012. *Int. J. Climatol.* **37**, 2522–2531 (2017).

17. 17.

Global Volcanism Program, 2013. Volcanoes of the World Version 4.9.1 (accessed 17 September 2020);
<https://doi.org/10.5479/si.GVP.VOTW4-2013>.

18. 18.

Klein, F. W. Eruption forecasting at Kilauea Volcano, Hawaii. *J. Geophys. Res. Solid Earth* **89**, 3059–3073 (1984).

19. 19.

Díez, M., Femina, P. C. L., Connor, C. B., Strauch, W. & Tenorio, V. Evidence for static stress changes triggering the 1999 eruption of Cerro Negro Volcano, Nicaragua and regional aftershock sequences. *Geophys. Res. Lett.* **32**, L04309 (2005).

20. 20.

Hainzl, S., Kraft, T., Wassermann, J., Igel, H. & Schmedes, E. Evidence for rainfall-triggered earthquake activity. *Geophys. Res. Lett.* **33**, L19303 (2006).

21. 21.

Hayba, D. O. & Ingebritsen, S. E. Multiphase groundwater flow near cooling plutons. *J. Geophys. Res. Solid Earth* **102**, 12235–12252 (1997).

22. 22.

Hsieh, P. A. & Ingebritsen, S. E. Groundwater inflow toward a preheated volcanic conduit: application to the 2018 eruption at Kīlauea Volcano, Hawai’i. *J. Geophys. Res. Solid Earth* **124**, 1498–1506 (2019).

23. 23.

Heap, M. J. et al. Towards more realistic values of elastic moduli for volcano modelling. *J. Volcanol. Geotherm. Res.* **390**, 106684 (2020).

24. 24.

Farquharson, J., Heap, M. J., Baud, P., Reuschlé, T. & Varley, N. R. Pore pressure embrittlement in a volcanic edifice. *Bull. Volcanol.* **78**, 6 (2016).

25. 25.

Li, D., Wang, T., Cheng, T. & Sun, X. Static and dynamic tensile failure characteristics of rock based on splitting test of circular ring. *Trans. Nonferrous Met. Soc. China* **26**, 1912–1918 (2016).

Acknowledgements

Copernicus Sentinel-1 and Cosmo-Skymed SAR data are available thanks to the Group on Earth Observation's Geohazard Supersites and Natural Laboratory Initiative. This work was supported by funding from NASA's Interdisciplinary Research in Earth Science programme (grant number 80NSSC17K0028 P00003). Data processing was conducted using Stampede2 at the Texas Advanced Computing Center of the Extreme Science and Engineering Discovery Environment, supported by National Science Foundation grant number ACI-1548562, using the public domain InSAR Scientific Computing Environment software of the Jet Propulsion Laboratory. We thank B. Varugu for discussions.

Author information

Affiliations

1. Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL, USA

Jamie I. Farquharson & Falk Amelung

Contributions

J.I.F. processed the GPS and rainfall data, and plotted all data. F.A. processed the InSAR data. Both authors contributed to the writing.

Corresponding author

Correspondence to [Jamie I. Farquharson](#).

Ethics declarations

Competing interests

This work was supported by funding from NASA's Interdisciplinary Research in Earth Science programme (grant number 80NSSC17K0028 P00003) exploring the influence of rainfall in triggering volcanism.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Farquharson, J.I., Amelung, F. Reply to: Rainfall an unlikely factor in Kīlauea's 2018 rift eruption. *Nature* **602**, E11–E14 (2022).
<https://doi.org/10.1038/s41586-021-04164-0>

- Published: 02 February 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04164-0>

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

[Extreme rainfall triggered the 2018 rift eruption at Kīlauea Volcano](#)

- Jamie I. Farquharson
- Falk Amelung

Article 22 Apr 2020

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04164-0>

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

Amendments & Corrections

- [**Author Correction: Multiple hominin dispersals into Southwest Asia over the past 400,000 years**](#) [10 January 2022]
Author Correction •
- [**Author Correction: A proximity-dependent biotinylation map of a human cell**](#) [11 January 2022]
Author Correction •
- [**Author Correction: Histidine catabolism is a major determinant of methotrexate sensitivity**](#) [11 January 2022]
Author Correction •
- [**Publisher Correction: Giant modulation of optical nonlinearity by Floquet engineering**](#) [12 January 2022]
Publisher Correction •

- Author Correction
- Open Access
- [Published: 10 January 2022](#)

Author Correction: Multiple hominin dispersals into Southwest Asia over the past 400,000 years

- [Huw S. Groucutt](#) [ORCID: orcid.org/0000-0002-9111-1720](#)^{1,2,3},
- [Tom S. White](#) [ORCID: orcid.org/0000-0001-9056-7286](#)⁴,
- [Eleanor M. L. Scerri](#) [ORCID: orcid.org/0000-0002-7468-9977](#)^{5,6,3},
- [Eric Andrieux](#) [ORCID: orcid.org/0000-0003-1688-0543](#)^{7,8},
- [Richard Clark-Wilson](#)^{8,9},
- [Paul S. Breeze](#) [ORCID: orcid.org/0000-0003-2575-8021](#)¹⁰,
- [Simon J. Armitage](#) [ORCID: orcid.org/0000-0003-1480-9188](#)^{8,11},
- [Mathew Stewart](#) [ORCID: orcid.org/0000-0002-0413-8133](#)¹,
- [Nick Drake](#)^{10,2},
- [Julien Louys](#) [ORCID: orcid.org/0000-0001-7539-0689](#)^{12,13},
- [Gilbert J. Price](#) [ORCID: orcid.org/0000-0001-8406-4594](#)¹⁴,
- [Mathieu Duval](#) [ORCID: orcid.org/0000-0003-3181-7753](#)^{12,15},
- [Ash Parton](#)^{16,17},
- [Ian Candy](#)⁸,
- [W. Christopher Carleton](#) [ORCID: orcid.org/0000-0001-7463-8638](#)¹,
- [Ceri Shipton](#)^{18,19},
- [Richard P. Jennings](#) [ORCID: orcid.org/0000-0001-9996-7518](#)²⁰,
- [Muhammad Zahir](#) [ORCID: orcid.org/0000-0002-1765-6319](#)^{2,21},
- [James Blinkhorn](#)^{5,8},
- [Simon Blockley](#)⁸,
- [Abdulaziz Al-Omari](#)²²,
- [Abdullah M. Alsharekh](#) [ORCID: orcid.org/0000-0002-2666-3262](#)²³ &

- [Michael D. Petraglia](#) ORCID: orcid.org/0000-0003-2522-5727^{2,12,24,25}

Nature volume **602**, page E15 (2022)

- 1122 Accesses
- 1 Citations
- 4 Altmetric
- [Metrics details](#)

Subjects

- [Archaeology](#)

The [Original Article](#) was published on 01 September 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03863-y>
Published online 1 September 2021

In the version of this Article initially published, the department in affiliation 15 appeared incorrectly as “Geochronology and Geology, Centro Nacional de Investigación sobre la Evolución Humana, Paseo de Atapuerca, Burgos, Spain.” This has now been corrected to read “Geochronology and Geology, Centro Nacional de Investigación sobre la Evolución Humana (CENIEH), Paseo de Atapuerca, Burgos, Spain.”

Further, there was a duplication error in the Late Pleistocene lithic data and principal components analysis reported in the Supplementary Information initially published online. The error had no significant effect on the original results. A revised analysis with detailed description of the error and links to corrected and uncorrected data, scripts and replication instructions are available in the Supplementary Information accompanying this amendment,

while the revised [Supplementary Information](#) appears alongside the main article. The original Article has been corrected online.

Supplementary Information is available in the online version of this Amendment.

Author information

Affiliations

1. Extreme Events Research Group, Max Planck Institutes for Chemical Ecology, the Science of Human History, and Biogeochemistry, Jena, Germany

Huw S. Groucutt, Mathew Stewart & W. Christopher Carleton

2. Department of Archaeology, Max Planck Institute for the Science of Human History, Jena, Germany

Huw S. Groucutt, Nick Drake, Muhammad Zahir & Michael D. Petraglia

3. Institute of Prehistoric Archaeology, University of Cologne, Cologne, Germany

Huw S. Groucutt & Eleanor M. L. Scerri

4. Department of Life Sciences, Natural History Museum, London, UK

Tom S. White

5. Pan-African Evolution Research Group, Max Planck Institute for the Science of Human History, Jena, Germany

Eleanor M. L. Scerri & James Blinkhorn

6. Department of Classics and Archaeology, University of Malta, Msida, Malta

Eleanor M. L. Scerri

7. Department of Archaeology, Durham University, Durham, UK

Eric Andrieux

8. Centre for Quaternary Research, Department of Geography, Royal Holloway University of London, Egham, UK

Eric Andrieux, Richard Clark-Wilson, Simon J. Armitage, Ian Candy, James Blinkhorn & Simon Blockley

9. Department of Geography and Environmental Science, University of Reading, Reading, UK

Richard Clark-Wilson

10. Department of Geography, King's College London, London, UK

Paul S. Breeze & Nick Drake

11. SFF Centre for Early Sapiens Behaviour (SapienCE), University of Bergen, Bergen, Norway

Simon J. Armitage

12. Australian Research Centre for Human Evolution, Griffith University, Brisbane, Queensland, Australia

Julien Louys, Mathieu Duval & Michael D. Petraglia

13. College of Asia and the Pacific, The Australian National University, Canberra, Australia Capital Territory, Australia

Julien Louys

14. School of Earth and Environmental Sciences, University of Queensland, Brisbane, Australia Capital Territory, Australia

Gilbert J. Price

15. Geochronology and Geology, Centro Nacional de Investigación sobre la Evolución Humana, Paseo de Atapuerca, Burgos, Spain

Mathieu Duval

16. Human Origins and Palaeoenvironments Research Group, School of Social Sciences, Oxford Brookes University, Oxford, UK

Ash Parton

17. Mansfield College, University of Oxford, Oxford, UK

Ash Parton

18. Institute of Archaeology, University College London, London, UK

Ceri Shipton

19. Centre of Excellence for Australian Biodiversity and Heritage, Australian National University, Canberra, Australia Capital Territory, Australia

Ceri Shipton

20. School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK

Richard P. Jennings

21. Department of Archaeology, Hazara University, Mansehra, Pakistan

Muhammad Zahir

22. Heritage Commission, Ministry of Culture, Riyadh, Saudi Arabia

Abdulaziz Al-Omari

23. Department of Archaeology, College of Tourism and Archaeology,
King Saud University, Riyadh, Saudi Arabia

Abdullah M. Alsharekh

24. Human Origins Program, National Museum of Natural History,
Smithsonian Institution, Washington, USA

Michael D. Petraglia

25. School of Social Science, University of Queensland, St Lucia,
Queensland, Australia

Michael D. Petraglia

Corresponding authors

Correspondence to [Huw S. Groucutt](#) or [Michael D. Petraglia](#).

Supplementary information

Supplementary Information

This file contains a revised analysis with detailed description of the error and links to corrected and uncorrected data, scripts and replication instructions.

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

Groucutt, H.S., White, T.S., Scerri, E.M.L. *et al.* Author Correction: Multiple hominin dispersals into Southwest Asia over the past 400,000 years. *Nature* **602**, E15 (2022). <https://doi.org/10.1038/s41586-021-04289-2>

- Published: 10 January 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04289-2>

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

- [Multiple phases of human occupation in Southeast Arabia between 210,000 and 120,000 years ago](#)
 - K. Bretzke
 - F. Preusser
 - A. G. Parker

Scientific Reports (2022)

This article was downloaded by **calibre** from <https://www.nature.com/articles/s41586-021-04289-2>

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

- Author Correction
- [Published: 11 January 2022](#)

Author Correction: A proximity-dependent biotinylation map of a human cell

- [Christopher D. Go](#)^{1,2} na1,
- [James D. R. Knight](#)¹ na1,
- [Archita Rajasekharan](#)³,
- [Bhavisha Rathod](#)¹,
- [Geoffrey G. Hesketh](#) ORCID: orcid.org/0000-0002-5570-7615¹,
- [Kento T. Abe](#)^{1,2},
- [Ji-Young Youn](#)^{1,2} nAff7,
- [Payman Samavarchi-Tehrani](#)¹,
- [Hui Zhang](#)⁴,
- [Lucie Y. Zhu](#) ORCID: orcid.org/0000-0002-1048-5377⁴,
- [Evelyn Popiel](#)²,
- [Jean-Philippe Lambert](#) ORCID: orcid.org/0000-0002-0833-1221¹ nAff8 nAff9,
- [Étienne Coyaud](#)⁵ nAff10,
- [Sally W. T. Cheung](#) ORCID: orcid.org/0000-0003-0974-8720¹,
- [Dushyandi Rajendran](#)¹,
- [Cassandra J. Wong](#)¹,
- [Hana Antonicka](#) ORCID: orcid.org/0000-0002-2525-2867³,
- [Laurence Pelletier](#) ORCID: orcid.org/0000-0003-1171-4618^{1,2},
- [Alexander F. Palazzo](#) ORCID: orcid.org/0000-0002-9700-1995⁴,
- [Eric A. Shoubridge](#)³,
- [Brian Raught](#)^{5,6} &
- [Anne-Claude Gingras](#) ORCID: orcid.org/0000-0002-6090-4437^{1,2}

Nature volume **602**, page E16 (2022)

- 885 Accesses
- 2 Altmetric
- [Metrics details](#)

Subjects

- [Computational biology and bioinformatics](#)
- [Proteomics](#)
- [Systems biology](#)

The [Original Article](#) was published on 02 June 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-03592-2>
Published online 2 June 2021

In the version of this Article initially published, we omitted in the introductory paragraph a reference to a paper that reported the BioID profiles of 18 subcellular markers and suborganelle localization markers and proposed a localization score strategy. This paper is now cited as ref. [61](#) at the end of the last sentence of the first paragraph: “BioID has been successfully used to define the composition of many different protein complexes and the spatial organization of several membrane-bound and membraneless organelles (see, for example, refs. [5–8,61](#)).”

The authors believe that they should have included this reference and apologize for the omission.

References

1. 61.

Liu, X. et al. An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. *Nat. Commun.* **9**, 1188 (2018).

Author information

Author notes

1. Ji-Young Youn

Present address: Peter Gilgan Centre for Research and Learning,
Hospital for Sick Children, Toronto, Ontario, Canada

2. Jean-Philippe Lambert

Present address: Department of Molecular Medicine, Cancer Research
Centre, Big Data Research Centre, Université Laval, Quebec City,
Quebec, Canada

3. Jean-Philippe Lambert

Present address: CHU de Québec-Université Laval Research Center
(CHUL), Quebec City, Quebec, Canada

4. Étienne Coyaud

Present address: PRISM INSERM U1192, Université de Lille,
Villeneuve d'Ascq, France

5. These authors contributed equally: Christopher D. Go, James D. R.
Knight

Affiliations

1. Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Sinai
Health System, Toronto, Ontario, Canada

Christopher D. Go, James D. R. Knight, Bhavisha Rathod, Geoffrey G. Hesketh, Kento T. Abe, Ji-Young Youn, Payman Samavarchi-Tehrani, Jean-Philippe Lambert, Sally W. T. Cheung, Dushyandi Rajendran, Cassandra J. Wong, Laurence Pelletier & Anne-Claude Gingras

2. Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

Christopher D. Go, Kento T. Abe, Ji-Young Youn, Evelyn Popiel, Laurence Pelletier & Anne-Claude Gingras

3. Montreal Neurological Institute and Department of Human Genetics, McGill University, Montreal, Quebec, Canada

Archita Rajasekharan, Hana Antonicka & Eric A. Shoubridge

4. Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada

Hui Zhang, Lucie Y. Zhu & Alexander F. Palazzo

5. Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada

Étienne Coyaud & Brian Raught

6. Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada

Brian Raught

Corresponding author

Correspondence to [Anne-Claude Gingras](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Go, C.D., Knight, J.D.R., Rajasekharan, A. *et al.* Author Correction: A proximity-dependent biotinylation map of a human cell. *Nature* **602**, E16 (2022). <https://doi.org/10.1038/s41586-021-04308-2>

- Published: 11 January 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04308-2>

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/s41586-021-04308-2>

- Author Correction
- [Published: 11 January 2022](#)

Author Correction: Histidine catabolism is a major determinant of methotrexate sensitivity

- [Naama Kanarek](#)^{1,2,3,4},
- [Heather R. Keys](#) [ORCID: orcid.org/0000-0003-1371-2288](#)¹,
- [Jason R. Cantor](#)^{1,2,3,4},
- [Caroline A. Lewis](#) [ORCID: orcid.org/0000-0003-1787-5084](#)¹,
- [Sze Ham Chan](#)¹,
- [Tenzin Kunchok](#)¹,
- [Monther Abu-Remaileh](#)^{1,2,3,4},
- [Elizaveta Freinkman](#)¹,
- [Lawrence D. Schweitzer](#)⁴ &
- [David M. Sabatini](#) [ORCID: orcid.org/0000-0002-1446-7256](#)^{1,2,3,4}

[Nature](#) volume 602, pages E17–E18 (2022)

- 1125 Accesses
- 2 Altmetric
- [Metrics details](#)

Subjects

- [Acute lymphocytic leukaemia](#)
- [Cancer metabolism](#)
- [Cancer therapy](#)

The [Original Article](#) was published on 11 July 2018

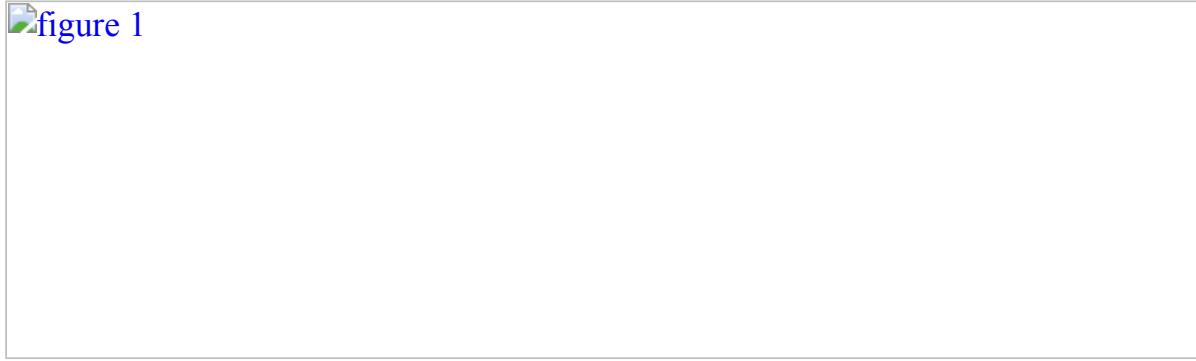
[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-018-0316-7> Published online 11 July 2018

In Fig. 1f of this Article, one of the three biological replicates of the control drug doxorubicin was erroneously duplicated from another experiment. The original Fig. 1f is shown next to the corrected panel without the erroneous replicate as Fig. 1 of this Amendment. The Source Data for the original Fig. 1, which had included erroneous data in the tab ‘survival MTX’ regarding this doxorubicin replicate, has also been corrected and is shown as Supplementary Information to this Amendment. In the Fig. 1 legend, the text “...compared to wild-type cells ($n = 3$, biological replicates,” should have read: ‘...compared to wild-type cells ($n = 3$, biological replicates, **except for doxorubicin in Ramos cells, where $n = 2$,**”.

Fig. 1: This is the original published Fig. 1f (left) and the corrected Fig. 1f (right)

 **figure 1**



of the original Article.

In Extended Data Fig. 11 of this Article, two of the histological sections in the ‘methotrexate’ group (top and bottom images in Extended Data Fig. 11a) and two sections in the ‘vehicle’ group (middle and bottom images in Extended Data Fig. 11b) were duplicated, and the sections in the ‘histidine supplementation’ (bottom image in Fig. 11b) and ‘methotrexate’ (top image in Extended Data Fig. 11b) groups were also duplicated. The original panels of Extended Data Fig. 11 are shown next to the corrected panels as Fig. 2 of this Amendment. In [Supplementary Fig. 3](#) of this Article, two of the histological sections in the ‘methotrexate’ group were duplicated. The original and corrected [Supplementary Fig. 3](#) is shown as Supplementary Fig. 1 to this Amendment, with the new panels marked by red outlines. There were errors in the Source Data of Fig. 2 that have been corrected and have been highlighted in red in Supplementary Information to this Amendment; the data in the original Fig. 2 were correct. In the Methods section, the reference for the sgRNA library that was used for the screen (Wang et al. (2014), ref. 7) was incorrect, and the updated library was instead from Wang et al. (2015)¹. The library can also be found here: <http://www.addgene.org/pooled-library/sabatini-crispr-human-high-activity-3->

[sublibraries](#). These errors do not affect the conclusions of the paper, and the original Article has not been corrected online.

Fig. 2: This is the original published Extended Data Fig. 11 and the corrected Extended Data Fig. 11 of the original Article.

 figure 2



The new panels are marked by red outlines.

David Sabatini is no longer affiliated with the Whitehead Institute or the Howard Hughes Medical Institute. At the request of the Whitehead Institute and to ensure execution of the duties of corresponding author, the corresponding author on this Author Correction is now Naama Kanarek (naama.kanarek@childrens.harvard.edu), replacing David Sabatini. All authors agree with the Author Correction and the change of the corresponding author.

Supplementary information is available in the online version of this Amendment.

References

1. 1.

Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).

Author information

Affiliations

1. Whitehead Institute for Biomedical Research and Massachusetts Institute of Technology, Department of Biology, Cambridge, MA, USA

Naama Kanarek, Heather R. Keys, Jason R. Cantor, Caroline A. Lewis, Sze Ham Chan, Tenzin Kunchok, Monther Abu-Remaileh, Elizaveta Freinkman & David M. Sabatini

2. Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

Naama Kanarek, Jason R. Cantor, Monther Abu-Remaileh & David M. Sabatini

3. Koch Institute for Integrative Cancer Research and Massachusetts Institute of Technology, Department of Biology, Cambridge, MA, USA

Naama Kanarek, Jason R. Cantor, Monther Abu-Remaileh & David M. Sabatini

4. Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA, USA

Naama Kanarek, Jason R. Cantor, Monther Abu-Remaileh, Lawrence D. Schweitzer & David M. Sabatini

Corresponding author

Correspondence to [Naama Kanarek](#).

Supplementary information

[Supplementary Figure 1](#)

The original and corrected Supplementary Fig. 3 is shown, with the new panels marked by red outlines.

[Supplementary Data](#)

This file contains the corrected Source Data for the original Fig. 1, which had included erroneous data in the tab ‘survival MTX’ regarding one of the doxorubicin replicates performed in the cell line Ramos.

Supplementary Data

This file contains the corrected Source Data for the original Fig. 2, with corrections highlighted in red.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Kanarek, N., Keys, H.R., Cantor, J.R. *et al.* Author Correction: Histidine catabolism is a major determinant of methotrexate sensitivity. *Nature* **602**, E17–E18 (2022).
<https://doi.org/10.1038/s41586-021-03487-2>

- Published: 11 January 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-03487-2>

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

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

- Publisher Correction
- [Published: 12 January 2022](#)

Publisher Correction: Giant modulation of optical nonlinearity by Floquet engineering

- [Jun-Yi Shan](#) [ORCID: orcid.org/0000-0001-7665-2169^{1,2}](#),
- [M. Ye³](#),
- [H. Chu^{1,2}](#),
- [Sungmin Lee⁴](#),
- [Je-Geun Park](#) [ORCID: orcid.org/0000-0002-3930-4226^{4,5,6}](#),
- [L. Balents³](#) &
- [D. Hsieh](#) [ORCID: orcid.org/0000-0002-0812-955X^{1,2}](#)

[Nature](#) volume **602**, page E19 (2022)

- 511 Accesses
- 1 Altmetric
- [Metrics details](#)

Subjects

- [Electronic properties and materials](#)
- [Nonlinear optics](#)
- [Ultrafast photonics](#)

The [Original Article](#) was published on 08 December 2021

[Download PDF](#)

Correction to: *Nature* <https://doi.org/10.1038/s41586-021-04051-8>
Published online 8 December 2021

In the version of this article initially published, the reference list in the main text of the PDF was incomplete, with refs. [28](#),[29](#),[30](#),[31](#),[32](#),[33](#),[34](#),[35](#),[36](#),[37](#) appearing online only. For completeness, the refs. are now listed below.

The corrections have been made to the online version of the article.

References

1. 28.

Grasso, V., Neri, F., Perillo, P., Silipigni, L. & Piacentini, M. Optical-absorption spectra of crystal-field transitions in MnPS₃ at low temperatures. *Phys. Rev. B* **44**, 11060–11066 (1991).

2. 29.

Fiebig, M., PavlovV. V. & Pisarev, R. V. Second-harmonic generation as a tool for studying electronic and magnetic structures of crystals: review. *J. Opt. Soc. Am. B* **22**, 96–118 (2005).

3. 30.

Boyd, R. W. *Nonlinear Optics* (Academic Press, 2003).

4. 31.

Muthukumar, V. N., Valentí, R. & Gros, C. Microscopic model of nonreciprocal optical effects in Cr₂O₃. *Phys. Rev. Lett.* **75**, 2766–2769 (1995).

5. 32.

Harter, J. W., Niu, L., Woss, A. J. & Hsieh, D. High-speed measurement of rotational anisotropy nonlinear optical harmonic generation using position-sensitive detection. *Opt. Lett.* **40**, 4671–4674 (2015).

6. 33.

Wildes, A. R., Rønnow, H. M., Roessli, B., Harris, M. J. & Godfrey, K. W. Static and dynamic critical properties of the quasi-two-dimensional antiferromagnet MnPS₃. *Phys. Rev. B* **74**, 094422 (2006).

7. 34.

Autler, S. H. & Townes, C. H. Stark effect in rapidly varying fields. *Phys. Rev.* **100**, 703–722 (1955).

8. 35.

Sie, E. J. et al. Valley-selective optical Stark effect in monolayer WS₂. *Nat. Mater.* **14**, 290–294 (2015).

9. 36.

Bloch, F. & Siegert, A. Magnetic resonance for nonrotating fields. *Phys. Rev.* **57**, 522–527 (1940).

10. 37.

Sentef, M. A., Li, J., Künzel, F. & Eckstein, M. Quantum to classical crossover of Floquet engineering in correlated quantum systems. *Phys. Rev. Res.* **2**, 033033 (2020).

Author information

Affiliations

1. Department of Physics, California Institute of Technology, Pasadena, CA, USA

Jun-Yi Shan, H. Chu & D. Hsieh

2. Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA, USA

Jun-Yi Shan, H. Chu & D. Hsieh

3. Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA, USA

M. Ye & L. Balents

4. Department of Physics and Astronomy, Seoul National University, Seoul, Republic of Korea

Sungmin Lee & Je-Geun Park

5. Center for Quantum Materials, Seoul National University, Seoul, Republic of Korea

Je-Geun Park

6. Institute of Applied Physics, Seoul National University, Seoul, Republic of Korea

Je-Geun Park

Corresponding author

Correspondence to [D. Hsieh](#).

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Shan, JY., Ye, M., Chu, H. *et al.* Publisher Correction: Giant modulation of optical nonlinearity by Floquet engineering. *Nature* **602**, E19 (2022). <https://doi.org/10.1038/s41586-021-04368-4>

- Published: 12 January 2022
- Issue Date: 03 February 2022
- DOI: <https://doi.org/10.1038/s41586-021-04368-4>

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/s41586-021-04368-4>