nature

SCHISTOSOME GENOMICS

Whole-genome sequences of neglected pathogens *S. mansoni* and *S. japonicum*

THE APOLLO LEGACY

Can human spaceflight inspire again?

GLACIAL CLIMATES

Variations on a theme

PRIMATE ARCHAEOLOGY

An alternative past

Abstractions

FIRST AUTHOR

In the birth defect phocomelia, the long bones of the limbs are shorter than normal or, in the worst cases, do not develop at all, meaning that the hand or foot is attached directly to the shoulder or hip. Scientists studying the disorder — also seen in the 1950s and 1960s in the children of women who were prescribed the sedative thalidomide while pregnant — can mimic its characteristics by exposing the limb buds of developing chick embryos to irradiation.

Researchers had thought that the effects of both X-irradiation and thalidomide resulted in a defect in patterning — the correct laying-out of cells during development. In this model, a cell's identity is determined by the length of time it spends in a region near the tip of the developing limb bud. The thinking was that X-rays or thalidomide forced cells to stay too long in this region, called the progress zone, blocking their ability to form the bone cells of the limb regions closest to the shoulder and hip. But postdoc Jenna Galloway at Harvard Medical School in Boston, Massachusetts, and her colleagues have discovered that X-irradiation leads to problems with differentiation, the process by which progenitor cells assume specialized characteristics and functions (see page 400). She tells *Nature* more.

How does irradiation affect limb-bud cells?

During normal development the limb bud grows out and skeletal progenitor cells differentiate into bone cells. After we irradiated chick embryos, their limb buds lost some of the progenitor cells that become bone cells in the upper and middle parts of a limb, or proximal segments. These proximal segments still form, based on the presence of segment-specific, or patterning, markers. But when we looked for a differentiation marker for skeletal progenitors, we could not find it in the proximal segments. The marker was left only in the distal segments, the cells that would develop into hands or feet.

Were any results unexpected?

We were surprised that a major signal for limb growth and for telling cells to become distal cells was not disrupted by X-rays. This signal remained intact, which is why that part of the limb develops even after irradiation.

Did you encounter challenges?

When I started irradiating chick embryos, I put whole eggs into the X-ray machine. The embryos were fine, but did not get irradiated because of the shell. Then I tried cutting out a section of eggshell, leaving the embryo exposed, but it died. Finally, I cut a hole in the egg and placed shell pieces over part of the embryo to shield it, and that worked. ■

MAKING THE PAPER

Susan Alberts and Greg Wray

Baboon susceptibility to a parasite reveals parallels with humans.

Susan Alberts (left) and Greg Wray.

During the past four decades, US and Kenyan scientists have been monitoring a population of yellow baboons (*Papio cynocephalus*) in the Amboseli National Park in Kenya. The animals' detailed life histories are now being combined with genetic analyses to gain a glimpse into their evolutionary past.

As a first step, Susan Alberts and Greg Wray of Duke University in Durham, North Carolina, and their colleagues have identified a variation in a baboon gene that is associated with susceptibility to a malaria-like pathogen — the first report of a link between a gene and a complex trait in a wild population of non-human primates (see page 388). The fact that a similar association exists in humans suggests that both species endured comparable pressures from pathogens during their evolution.

"Like early humans, baboons are savannah foragers that adapted to a wide range of habitats and opportunities. They can be put in almost any environment and will thrive," says Alberts. "We think they share a parallel evolutionary history with humans."

Because of such parallels, the Amboseli Baboon Research Project was established in 1971 by Jeanne and Stuart Altmann of Princeton University, New Jersey, to study a population of about 300 baboons organized into 5 social groups, or extended families. "Literally every day someone is observing them," says Alberts, who co-directs the project with Jeanne Altmann.

In addition to careful monitoring of baboon behaviour, such as eating and sleeping habits, researchers have gathered physical measurements, as well as hormone and DNA samples from the animals' faeces. A few years ago, Alberts and Wray, a geneticist, started discussing the idea of applying genetic methods to gain further insight into these animals.

Both researchers credit first author Jenny Tung with getting that idea off the ground. She wanted to find connections between variations in certain genes, or genotypes, and particular traits (phenotypes) in the Amboseli baboons.

As a first phenotype, they chose to examine the susceptibility of baboons to infection with the blood parasite *Hepatocystis kochi*, which is related to the *Plasmodium* parasite that causes human malaria. Human susceptibility to pathogens is a trait that is often linked to single genes, and is thus amenable to genetic analyses. The main challenge for doing this type of study in wild animals is obtaining enough high-quality DNA and RNA samples for analysis.

Tung and Alberts spent two months during three consecutive summers living in tents at the Amboseli research site, collecting blood samples from baboons to extract DNA and RNA. Working with Kenyan researchers, they first had to put the animals to sleep by darting them, while being careful to minimize disruption to the group. "We would only dart animals who were alone, away from the rest of their social group, and no more than two animals each day and no more than six each week," says Alberts. "As a result the animals never lost habituation to us."

Back in the lab at Duke, the group analysed DNA samples from 190 baboons for the presence of *Hepatocystis* DNA. They discovered that infection rates varied substantially between individuals. They then looked at the DNA sequences of several candidate genes to identify changes that were associated with the observed variations in *Hepatocystis* susceptibility. In the end, they found that a change of one nucleotide in a gene called *FY* confers protection from the parasite. A different variation in the human *FY* gene is known to confer protection against malaria.

The fact that a genotype is associated with similar traits in baboons and humans shows that "given the same evolutionary pressures, the same type of thing can happen in parallel in different organisms", says Wray. "One lesson from this is the precision with which natural selection acts." Another important conclusion, notes Wray, "is that field studies such as this one can be done despite all the challenges".

Alberts and Wray next plan to tackle an even more ambitious project: finding genotypes associated with behavioural traits. In a 2003 paper in *Science* on baboon social integration, Alberts and her colleagues examined the observation that, although female baboons spend their whole lives in the same social group, some are well integrated — having lots of other baboons to groom with, for example — whereas others are more isolated. "We found that the degree of social integration predicts how well the offspring will do," says Alberts. "So there is an evolutionary pressure on females to be social."

The team will now look for the genetic determinants of social isolation. "It is a risky project. But this is the kind of population you need to do it in," says Wray. "And if we can't do it here, then it probably can't be done." ■

Visit Nautilus for regular news relevant to *Nature* authors ▸ http://blogs.nature.com/nautilus and see Peer-to-Peer for news for peer reviewers and about peer review ▸ http://blogs.nature.com/peer-to-peer.

nature

www.nature.com/nature

Vol 460 | Issue no. 7253 | 16 July 2009

Net gains

The Moon landing was not the only world-changing event in the summer of '69.

An international, cross-disciplinary survey by *Nature* on page 314 reveals just how powerfully the Apollo programme motivated young people to become scientists 40 years ago — a fact today's space scientists ignore at their peril (see pages 325 and 327).

Yet other events in the summer of 1969 would lead to a far deeper empowerment of scientists — and, indeed, many others. Even as *Apollo 11* was putting the first humans on the Moon, Ken Thompson at AT&T's Bell Labs was working to get *Space Travel*, a computer game he'd written for a mainframe computer, to run on a new, smaller machine. That effort led him to join with Dennis Ritchie and others to write a new computer operating system, which they named Unix. The rest is history: Unix triggered a still-ongoing boom in scientific computing, set the pattern for the open-source software movement and, along with its descendants, laid the foundations for the Internet.

Thompson wrote the first version of Unix in four weeks. It was initially for internal use only. But when the code was licensed for use outside of Bell Labs in the 1970s, it was quickly adopted by scientists worldwide. They embraced Unix for its power, simplicity and its ability to let machines interact with multiple users at once. No longer did they have to run stacks of punch cards through huge mainframes and wait hours or days to get their results on reams of fan-fold paper. Now they could just type and the machine would respond.

Timing was also key to the success of Unix. It appeared just as the mainframes were being challenged by a new generation of smaller, cheaper, interactive 'minicomputers' suited to individual departments and research groups. The new operating system could easily be adapted to run on any of these machines. Perhaps most importantly, researchers loved the fact that Unix was written in C, a new programming language that made it easy for them to write and share applications — and even to refine the operating system itself. They lost no time in doing so, in a flurry of innovation that presaged the open-source movement.

In the 1980s and 1990s, that openness fell victim to shortsighted commercial interests, when infighting among vendors of proprietary Unix machines produced incompatible versions of the operating system. This allowed Unix to be blindsided by an upstart called Microsoft, which quickly acquired a dominant position in the burgeoning microcomputer market with its own operating systems.

Nonetheless, Unix's future had been guaranteed in the late 1970s by the US Defense Advanced Research Projects Agency, which chose a version of the operating system developed at the University of California, Berkeley, as the basis for a number of its projects. And by funding Berkeley to include an implementation of the then-new TCP/IP protocols that underlay the Internet, the agency essentially made Unix Internet-ready. At the same time, the open-source movement became central to the Internet's development, with the Unix clone Linux, written in 1991 by Finnish student Linus Torvalds, coming to dominate.

Today, Linux or some other flavour of Unix runs most of the servers, routers and other elements of Internet infrastructure, as well as most mobile phones and GPS devices. And last week Google announced that it will use Linux as the base for its planned open-source 'Chrome' operating system.

The difference between the development of Unix and the top-down, colossal science and engineering project that was Apollo could hardly have been greater. Yet both are examples of the power of joint efforts. Unix, like Apollo, has earned its place in history. ■

Nowhere to hide

The G8 has laid down a marker by promising to restrict the rise of global temperatures.

At their annual summit last week, leaders of the Group of Eight (G8) industrialized nations pledged to try to keep the planet from warming by more than 2 °C above pre-industrial temperatures — an ambitious goal that has also been adopted by other countries including China, Brazil and South Africa (see page 313).

The absence of any commitment to reduce emissions before 2050 is less promising but — assuming that the G8 can be taken at its word — it shouldn't be a fatal defect. If warming is to stay within those two degrees, the global 'decarbonization' project must be tackled without delay anyway.

This will not be cheap or easy. When the two-degree ceiling is translated into a per-capita emissions limit, it is clear that the industrialized world, and particularly the United States, has already generated more than its fair share of greenhouse gases. By continuing to emit at these levels, these countries are taking up emissions allocations that developing nations need to use to grow. This creates a 'carbon debt' that must be repaid with technology and money, both of which will be necessary if poorer countries are to leapfrog the dirty development pioneered by the industrialized nations.

The only viable strategy is therefore a massive research-and-development drive for energy efficiency in the near term, coupled with a long-term move towards virtually carbon-free energy production by 2040 at the latest. Cap-and-trade regulations can play a part, but they are only the beginning. Smart urban planning, clean transport systems and lifestyle changes will also be important.

History teaches that great advances — such as agriculture, sanitation and mechanization — arrive in bursts. There is hope that, despite an agonizingly slow start, the energy revolution of the twenty-first century can still gain the required momentum. Avoiding long-term investment in high-emission technology will be more important than throwing around potentially unachievable emission-reduction numbers.

The G8 has set a point of reference for the UN climate-change conference in December. Global leaders must now put together a framework that allows — and requires — all nations to do their part. ■

RESEARCH HIGHLIGHTS

Smothered by a swarm

Naturwissenschaften doi:10.1007/s00114-009-0575-0 (2009)

Honeybees don't just kill hornets with heat — they gas them, too. The giant hornet (*Vespa mandarinia japonica*) is a ferocious predator of the Japanese honeybee (*Apis cerana japonica*). The bees fight back by smothering the invader inside a ball formed by their swarming bodies, killing the hornet in minutes. It was thought that the temperature inside the ball — which can rise to around 46 °C — killed the hornet while sparing the more heat-tolerant bees.

Michio Sugahara and Fumio Sakamoto of Kyoto Gakuen University, Japan, found that outside a bee ball, hornets survive such temperatures. Their heat tolerance falls, however, as carbon dioxide levels rise. The air in a bee ball contains about 3.7% CO_2, and in this atmosphere the temperature is lethal to the hornet.

PALAEOCLIMATOLOGY

Tropical ice

Geophys. Res. Lett. doi:10.1029/2009GL037643 (2009)

Ice cores can reveal historical information about climate and vegetation through trapped aerosols such as dust, soot and complex molecules.

Matthew Makou of Ohio State University in Columbus and a team of fellow ice enthusiasts looked at compounds derived from tropical plants that were blown to high altitudes, where they became trapped in ice. The researchers tried out relatively new methods — stir-bar sorptive extraction and thermal desorption — that allowed useful information to be extracted from such compounds without the need for many kilograms of ice.

The authors suggest that the methods will be useful in reconstructing past characteristics of tropical areas, such as forest-fire occurrence, vegetation cover and aridity.

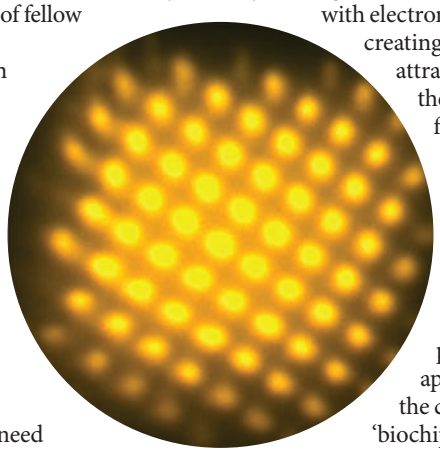
NANOTECHNOLOGY

Penned in protein

Angew. Chem. Int. Ed. doi:10.1002/anie.200900950 (2009)

Etching precise protein patterns on a substrate is important both for medical diagnostics and for fundamental studies. Until now, this has usually been done by first creating protein-attracting regions and then filling in the rest of the surface with protein-repellent molecules.

Michael Zharnikov of the University of Heidelberg in Germany and his coworkers have now developed a technique that 'writes' protein patterns on to a substrate more easily. They do this by depositing an ethylene glycol layer on a gold surface and then zapping it with electron-beam lithography, creating hydrophobic areas that attract proteins only where the electron beam was focused (pictured left; yellow dots represent adsorbed protein). If the adsorbed proteins carry specific binding sites, additional proteins can then be attached as a secondary patterning motif. The approach could lead to the creation of sophisticated 'biochips'.



CHEMICAL BIOLOGY

A glowing report

Nature Chem. Bio. doi:10.1038/nchembio.190 (2009)

Michael Tsang and colleagues at the University of Pittsburgh, Pennsylvania, wanted to find small-molecule modulators of the fibroblast growth factor (FGF) pathway, which is central to embryonic development. They screened chemicals in zebrafish that had been engineered to glow when FGF was active. One chemical known as BCI supercharged the glow, indicating it was boosting FGF activity.

Further study revealed that BCI blocks the activity of dual-specificity phosphatase 6 (Dusp6), an FGF feedback regulator, thereby stopping it from impairing FGF and resulting in the extra glow. Compared with untreated zebrafish embryos, BCI-treated embryos had more cardiac progenitor cells — which eventually give rise to the heart — but fewer of the cells that go on to line blood vessels and form blood cells. The results, say Tsang's team, demonstrate the utility of zebrafish in chemical screens *in vivo*.

EVOLUTION

Nice guys finish last

Behav. Ecol. Sociobiol. doi:10.1007/s00265-009-0814-6 (2009)

Selfish individuals can profit from the altruism of others in their group, and can even exploit a group's resources so much that the resources become exhausted — an event known as the tragedy of the commons.

Omar Eldakar, of the University of Arizona in Tucson, and his colleagues have now shown experimentally that aggressive mating in water striders (*Aquarius remigis*) can result in one such tragedy. Harassment of female water striders by males has previously been shown to drive away females, diminishing the mating success of all males in a group.

The team built pools and manipulated the number of aggressive and nonaggressive males in each. They found that hyperaggressive males had greater mating success than those which were not aggressive within mixed pools. But as the number of hyperaggressive insects increased, the mating success of both types decreased.

ORGANIC CHEMISTRY

Cockroach cruncher

Angew. Chem. Int. Ed. doi:10.1002/anie.200902192 (2009)

Powdered leaves from the Mexican cockroach plant, *Haplophyton cimicidum*, have been used to poison cockroaches and to repel fleas for hundreds of years. And attempts to synthesize one of the plant's active ingredients — the indole alkaloid (+)-haplophytine — have frustrated organic chemists ever since its complex molecular structure was solved in 1973.

Hidetoshi Tokuyama at Tohoku University in Sendai, Japan, and his colleagues now report the first total synthesis of (+)-haplophytine, creating 4.2 milligrams of the molecule. In a key step, the researchers performed an internal rearrangement to create the left-hand segment of the molecule. They then used a traditional Fischer indole synthesis to join that, through a central bond, to the previously synthesized right-hand segment.

IMMUNOLOGY

Themis in the thymus

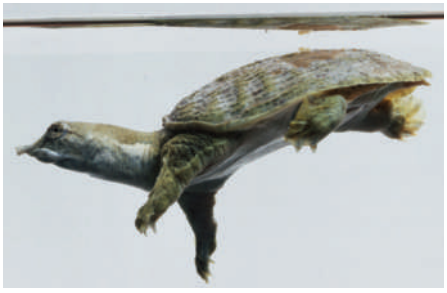
Nature Immunol. doi:10.1038/ni.1768; doi:10.1038/ni.1766; doi:10.1038/ni.1769 (2009)

The maturation of pathogen-fighting T cells relies on a mysterious protein called Themis, three independent groups have found.

Themis is expressed exclusively in T cells, but its protein sequence — although highly conserved in vertebrates — provided few clues to its function. All three teams generated mouse mutants that lack the full-length Themis protein, and found that the usual process by which properly functioning T cells are selected to continue maturation in the thymus was impaired. The result: fewer T cells survived this selection, and the mice contained smaller numbers of mature T cells.

Paul Love of the National Institutes of Health in Bethesda, Maryland, and his colleagues found that this defect could be reversed by stimulating signalling through a protein complex called T-cell antigen receptor (TCR), which is crucial for the activation and development of T cells.

Meanwhile, a second group led by Nicholas Gascoigne of the Scripps Research Institute in La Jolla, California, demonstrated that signalling through the TCR mediated by two other key molecules — calcium and a protein called Erk — is reduced when Themis is lacking. And a team headed by Richard Cornall of the University of Oxford, UK, and Ronald Schwartz, also of the National Institutes of Health, found that the expression of several genes, including those involved in cell survival, was reduced in these mutants — possibly contributing to eventual T-cell death.



DK LIMITED/CORBIS

DEVELOPMENTAL BIOLOGY

The turtle fold

Science **325,** 193–196 (2009)

The bodies of turtles and other chelonians are fundamentally different from those of all other animals. Their distinctive shells develop from the ribcage, which lies above, rather than below, the shoulder blades.

Shigeru Kuratani and his colleagues at the RIKEN Center for Developmental Biology in Kobe, Japan, compared the embryogenesis of the Chinese soft-shelled turtle (*Pelodiscus sinensis*, pictured below left) with that of other animals to understand how developmental changes may have resulted in the turtle's unusual form. The authors found that whereas the ribs of mice and chickens grow to encircle the body, those of turtles grow outwards and then stop, with the shoulder blades tucking under the edges of the ribs.

NEUROLOGY

New neurons show the way

Science **325,** 210–213 (2009)

Researchers have uncovered a function of new neurons that are generated in the brains of adult mice.

The role of adult neurogenesis in a brain region called the hippocampus has been difficult to pin down because the effects seem to be subtle and it is technically tricky to selectively disrupt neuronal development in adult animals. Fred Gage of the Salk Institute for Biological Studies in La Jolla, California, Timothy Bussey at the University of Cambridge, UK, and their colleagues designed challenging tasks that required mice to distinguish between very similar surroundings.

Using two techniques to knock down neurogenesis, the authors discovered that although the new neurons were not necessary for easy navigational tasks, they were important for more complex ones. Notably, for remembering very small differences in the spatial arrangement of their environment.

JOURNAL CLUB

Friedhelm von Blanckenburg
GFZ German Research Centre for Geosciences, Potsdam, Germany

A geochemist learns that mountain building does not accelerate rock weathering.

Mountain building has been deemed essential for stabilizing Earth's climate over the scale of millions of years. As tectonic forces push mountains into the sky, they provide fresh rock surface that is degraded by the physical effects of rain and temperature change, and by chemical weathering as carbon dioxide is dissolved in rain to form carbonic acid. Atmospheric CO_2 is thus consumed to convert rock into soil, which happens fastest where erosion rates are highest, exposing new rock to be weathered. Rivers then transport this carbon to the oceans where it is disposed of as carbonate sediment.

But many Earth scientists have questioned this story. Wouldn't periods of high tectonic activity, such as the rise of the Himalayas, provide enough rock to consume so much CO_2 that the planet would turn into an ice house?

Yet the consequences of mountain building are perhaps less profound than expected. The amount of weathering over time can be accurately measured on hill slopes using new geochemical methods that combine solute loss from soils with radioactive isotopes formed by cosmic rays to determine how long it takes for rocks to break down into soil. Two recent papers modelled the implications of this approach numerically. Surprisingly, the prediction is that weathering decreases rather than increases at erosion rates typical of high, active mountains (K. L. Ferrier and J. W. Kirchner *Earth Planet. Sci. Lett.* **272,** 591–599; 2008; E. J. Gabet and S. M. Mudd *Geology* **37,** 151–154; 2009). So hill-slope weathering in the Himalayas might do no more to withdraw CO_2 than any actively eroding, mid-altitude mountain range found worldwide.

Perhaps geochemists have been looking in the wrong place. Does the CO_2-consuming mineral decomposition thought to occur on high slopes actually happen on the floodplains below large, active mountains? We might need to take a closer look at these areas before we really understand the geological carbon cycle.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

NEWS

NIH nominee draws scrutiny

Francis Collins is likely to face funding challenges — and criticism of his Christian evangelism.

As Francis Collins prepares to take the helm of the US National Institutes of Health (NIH), opinions are divided about how the geneticist will steer the agency through its extraordinary funding boom.

Following President Barack Obama's long-anticipated nomination of Collins on 8 July, Harold Varmus, NIH director from 1993 to 1999, and now president and chief executive of the Memorial Sloan-Kettering Cancer Center in New York, described him as "a terrific scientist, inspirational leader, superb manager, and adept politician. I have great confidence in his ability to lead the NIH in complicated times."

But some believe that Collins has focused too much on genomics to the detriment of other approaches to improving health. Collins directed the NIH's National Human Genome Research Institute (NHGRI) for 15 years until 2008, led the public effort to sequence the human genome and was a co-discoverer of the gene responsible for cystic fibrosis (see *Nature* **460,** 164–169; 2009). "The genetic approach is extremely important, but it is a tool, it is not the be all and end all," says Fran Visco, president of the Washington-based National Breast Cancer Coalition, which has called on the NIH to back more research on the environmental causes of breast cancer.

There are also concerns about whether Collins's very public expressions of his evangelical Christian faith will affect his job. He laid out his beliefs in a 2006 book, *The Language of God: A Scientist Presents Evidence for Belief*, and last year founded the BioLogos Foundation, which aims to help Christians integrate their faith with contemporary science.

P. FRANZ/AP

Francis Collins has been a staunch advocate of genomic medicine.

Steven Pinker, a psychologist at Harvard University, says he has "serious misgivings" about the nomination. "Collins is an advocate of profoundly anti-science beliefs, and it is reasonable for the scientific community to ask him how these beliefs will affect his administration of the NIH and his efforts on behalf of the scientific enterprise."

Several scientists close to Collins say that his faith will not affect his NIH duties. "In all the years that I worked with Francis — and there were a lot of years — I never once saw evidence where his religious faith was in conflict with his scientific judgement," says developmental biologist Shirley Tilghman, president of Princeton University in New Jersey, who sat on Collins's advisory council when he directed the NHGRI.

"We would count him as an ally," adds Joshua Rosenau, a policy analyst at the National Center for Science Education, a non-profit organization in Oakland, California, that defends the teaching of evolution in schools. "It is helpful to have scientists like Francis Collins speaking out about how they personally reconcile science and religion."

The BioLogos Foundation has confirmed that Collins would step down from his role there before taking up the reins at the NIH. It's the right move, says Varmus. "Discussion about the foundation and his involvement with it could readily become a distraction from the business of running the NIH," he told *Nature*.

Malaria drug-makers ignore WHO ban

There is a growing risk that malaria parasites will develop resistance to artemisinin because almost half of both its manufacturers and malaria-affected countries are failing to comply with World Health Organization (WHO) demands to sell it only in combination with other drugs. Artemisinin and its derivatives are the leading treatments for the disease, being the only antimalarials that have not yet seen widespread resistance in malaria parasites.

The full scale of the problem is revealed in a soon-to-be-published WHO briefing seen by *Nature*, "Stop the marketing of oral artemisinin monotherapies", which calls for governments to empower national drug-regulatory authorities to clamp down on offending companies.

Treatments that use only artemisinin need to be taken for seven days to kill all parasites, but patients often stop treatment after a few days when they begin to feel much better. This leaves the remaining parasites in contact with low levels of the drug — a recipe for resistance. The WHO recommended in January 2006 that artemisinin should always be given in combination with other drugs for at least three days, because a cocktail reduces the chances of resistance. The need to move away from monotherapies has become all the more urgent with recent reports of resistance arising in Cambodia.

Although artemisinin-based combination therapies (ACTs) have become the treatment of choice for malaria, with a three-day programme curing more than 95% of patients, monotherapies are cheaper to produce and sell.

Of the 69 manufacturers of artemisinin monotherapies that the WHO has identified, 21 have withdrawn monotherapies, and 14 say they intend to comply with the WHO's recommendations. But the remaining 34 have not yet disclosed their intentions. Many have not even replied to multiple WHO e-mail

Collins takes over an NIH awash with cash, yet bracing for leaner times after a one-time bolus of US$10.4 billion in economic stimulus spending runs out in September 2010. Spending the money both rationally and quickly is already proving a daunting challenge for the NIH — the world's largest funder of biomedical research — whose peer-review system is close to being overwhelmed by a flood of applications (see *Nature* **459,** 763; 2009). "Everybody and their brother is going to have an idea of how to do this," says Tilghman. "He's going to have to sift through all those ideas and plans and ultimately do what is best for the United States. And he is going to have to do it relatively quickly."

Since Obama's 2010 budget requested only a 1.4% increase for the $30.3-billion agency, the NIH could also face the kind of boom and bust it confronted in 2003, when its rapidly expanded ranks of scientists were left competing for slices of a suddenly very finite pie. "He's going to have to fight for budget," says Paul Nurse, the Nobel-prizewinning biologist who is president of the Rockefeller University in New York.

The NIH is a behemoth that now has 27 institutes and centres and about 18,000 employees — much larger than the more focused NHGRI. "A big challenge for him is going to be to shift into a new role from his old role," says Elias Zerhouni, who left the NIH director's job last October. "You can't just dictate what you think is right. You have to listen a lot more."

White House rules prevent Collins from speaking to journalists before his nomination is considered by the Senate Committee on Health, Education, Labor, and Pensions, chaired by Senator Edward Kennedy (Democrat, Massachusetts). Kennedy has called the choice of Collins "inspired", and said he would work to see that the nomination is approved "without delay". ■

Meredith Wadman

and fax requests for information, says Andrea Bosman, an official at the WHO's Global Malaria Programme (see http://tinyurl.com/m4gqmn).

Regulatory authorities in just 39 of the 76 countries using malaria drugs have so far complied or said they intend to comply with the WHO's recommendations (see http://tinyurl.com/m94wqk). National regulatory authorities could help by banning monotherapies, says Bosman, but many are weak, understaffed and lack expertise.

"It's terrible," says Bosman, adding that every country where malaria is endemic is affected by the problem. "Who says there is no profit to be made in malaria? When you see the number of companies operating in Africa, and the diversity of products, you'd just be amazed." ■

Declan Butler

PUNCHSTOCK

APPROVALS ON TRIAL
European rules for clinical studies need streamlining.
www.nature.com/news

Flu furore hits Argentina

BUENOS AIRES

Argentina's government has for more than two weeks ignored recommendations from a committee of influenza experts to declare a national health emergency. Such a declaration would prompt greater federal resources to combat the spread of the pandemic H1N1 2009 virus.

The committee, specifically set up to advise the Ministry of Health on swine flu, issued its call just days before Argentina's national election on 28 June, which it feared could accelerate the spread of the virus.

Following the committee's call for action, the health minister, Graciela Ocaña, asked for the election to be postponed, but her request was rejected. Ocaña subsequently resigned and left office on 29 June. Mirta Roses, director of the Pan American Health Organization — the Latin American and Caribbean arm of the World Health Organization (WHO) — has also criticized the decision to go ahead with the elections.

"The delay was recommended because the agglomeration of people in places with little ventilation favours the transmission of the disease," says Jorge San Juan, coordinator of the expert committee and head of the intensive-care unit at Muñiz Hospital in Buenos Aires. The government now faces two lawsuits brought by individual lawyers who claim that its decision not to delay the elections recklessly endangered its citizens, who are legally required to vote.

Epidemiologist Emilio Santabaya, former head of the Malbrán Institute in Buenos Aires — the country's main flu monitoring centre — adds that Argentina knew an epidemic was coming and did nothing, and that flu monitoring in the country is not adequate to track the spread of the virus. The situation contrasts with that of neighbouring country Chile, which has many more flu surveillance centres, according to Jorge Jimenez, who has worked both as Chile's health minister and as chairman of the WHO's executive committee. On 7 July, Chilean President Michelle Bachelet, who trained as an epidemiologist, signed a decree that gives health authorities in the country the power to cancel events that could spread swine flu.

Meanwhile, a group of doctors, hospital workers and scientists called Médicos Sin Banderas (Doctors without Flags) has also accused the national government of deliberately hiding the full extent of the swine-flu outbreak.

Ariel Umpierrez, a health economist who heads the group, says that its members have been sharing information about flu cases from their places of work across the country. Those data show that the government has lowered the official number of cases by excluding infected people who visit private clinics, he argues. Argentina's Ministry of Health has not responded to *Nature*'s queries on the matter.

E. MARCARIAN/REUTERS

Argentines have been travelling abroad to buy flu drugs.

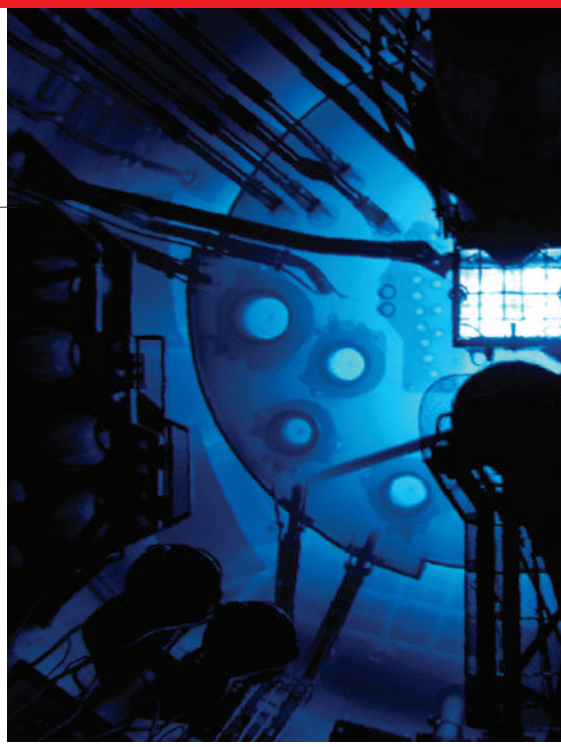
Although scientists inside the country say they are being ignored, those outside are worried by reports of panicked Argentines travelling to Chile and Uruguay to buy flu drugs. The drugs, which are being taken without medical advice, are probably being used in a way that promotes the development of resistance, says Eric Toner, a senior associate at the University of Pittsburgh's Center for Biosecurity in Pennsylvania. Because Uruguay is heavily dependent on trade with Argentina, it is highly unlikely to prevent Argentines from entering the country — but the governments of Brazil and Bolivia are considering closing their borders to prevent the spread of infection. ■

Anna Petherick

NATURE|Vol 460|16 July 2009

Medical isotope shortage reaches crisis level

Robust solutions sought urgently to shore up fragile supply chain.



T. TAPSELL, ANSTO

The worldwide shortage of medical isotopes is about to get much worse this week, as the High Flux Reactor in Petten, the Netherlands, closes for a month-long maintenance inspection.

It joins the National Research Universal reactor in Chalk River, Ontario, Canada, which has been closed since 15 May because of a heavy-water leak and is unlikely to restart before late 2009, according to Atomic Energy of Canada Limited, the government-sponsored body that runs the facility.

Together, the reactors produce two-thirds of the global supply of molybdenum-99, which decays to form technetium-99m, an isotope that is used in about 70,000 medical imaging procedures worldwide every day.

The shutdown has prompted calls for a major review of the way that medical isotopes are made and distributed. All five commercial nuclear reactors that use neutron-induced fission of highly enriched uranium-235 to make ^{99}Mo (see map) are more than 40 years old and cracks are beginning to emerge. The Petten reactor is scheduled to close again in early 2010 for up to six months so that deformed pipes in its cooling system can be repaired.

With both reactors unavailable for significant periods, "you really do have to say: 'right, let's clear the table and start all over'", says Thomas Ruth, senior research scientist at TRIUMF, Canada's national laboratory for particle and nuclear physics in Vancouver.

The problem arises because isotopes cannot be stockpiled — ^{99}Mo has a half life of 66 hours and ^{99m}Tc of just six hours. ^{99}Mo created in nuclear reactors must be shipped quickly to facilities that process the material into ^{99m}Tc generators. These generators are delivered to hospitals on a weekly basis, where they are 'milked' every morning to deliver ^{99m}Tc. Although the reactors in Belgium and South Africa will be working at full power over the next few weeks to produce more ^{99}Mo than usual, isotope suppliers say they will probably not be able to meet the shortfall. The shortage has been particularly acute in the United States and Canada, and "the situation will get much worse", says Michael Graham, director of nuclear medicine at the University of Iowa in Iowa City, and president of the international Society of Nuclear Medicine based in Reston, Virginia. "It is likely many studies will be cancelled or postponed."

Leaky pipeline

Week-by-week availability of ^{99}Mo is now coordinated by the Brussels-based Association of Imaging Producers and Equipment Suppliers (AIPES), which in the past year has taken the initiative to stimulate global action on the isotope crisis. Both the AIPES and the Organisation for Economic Co-operation and Development (OECD), based in Paris, began to investigate long-term solutions to the chronic shortage of ^{99}Mo at the start of 2009. But given the importance of ^{99m}Tc in health care, the ageing reactor infrastructure and highly fragmented supply network needs a root-and-branch overhaul.

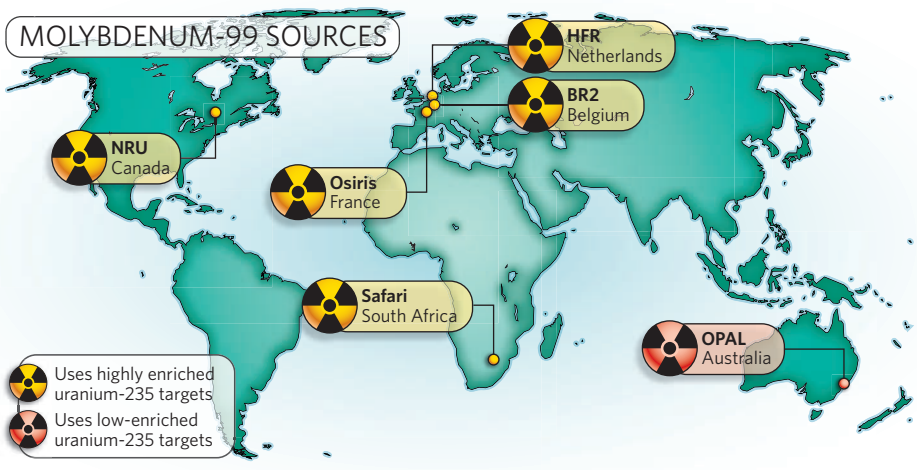
"We must upgrade and invest in updates to these ageing reactors," says George Segall, chief of the nuclear-medicine service at the Veterans' Affairs Palo Alto Health Care System in California. "Many of these reactors have already exceeded their useful lifetime. No one expected that they would still be needed now when they were built."

The Canadian government has so far resisted calls to switch on two Multipurpose Applied Physics Lattice Experiment (MAPLE) reactors, built at the Chalk River site to replace the 52-year-old National Research Universal reactor. The MAPLE reactors were to be the first nuclear reactors in the world dedicated to production of medical isotopes, but are lying dormant after Atomic Energy of Canada Limited halted development in May 2008, citing safety issues that it claimed it would be too expensive to resolve. If started up, these reactors alone could deliver more than the current global requirement for medical isotopes, according to Harold Smith, who was a manager on the MAPLE project.

Some medical-isotope suppliers are looking to academic research reactors to fill the gap. MDS Nordion, a medical-isotope provider based in Ottawa, Canada, that had invested heavily in the ill-fated MAPLE project, is negotiating with the Karpov Institute of Physical Chemistry in Moscow, which holds a 90% share of the market for ^{99m}Tc generators in Russia.

Researchers at Germany's Technical University of Munich are seeking financial backing from the government to support an upgrade to the university's FRM II neutron source to create ^{99}Mo. If highly enriched ^{235}U targets were used, the reactor could satisfy almost all of Europe's needs for ^{99m}Tc once it becomes fully operational, says Winfried Petry, scientific director of the FRM II.

An alternative option, says Ruth, could be to fire high-energy electrons at ^{100}Mo targets, forcing out a neutron to create ^{99}Mo. This



SOURCE: LANTHEUS MEDICAL IMAGING

Australia's OPAL reactor is the only one currently irradiating low-enriched uranium.

process could be carried out in small particle accelerators, similar to those already used to make medical isotopes such as fluorine-18 for positron emission tomography scans. Although this strategy would be much less efficient than ^{235}U fission, it could at least alleviate ^{99m}Tc shortages at a local level.

Another possible pathway is electron-induced ^{238}U fission. This, too, would bypass the need for nuclear reactors, although it would need to be done in dedicated high-powered accelerators (T. Ruth *Nature* **457,** 536–537; 2009).

For now, ^{235}U fission remains the most efficient way to generate ^{99}Mo in the quantities needed by the medical community. However, the reliance on highly enriched ^{235}U has made many governments reluctant to build more reactors because of fears that the material could be used to make a nuclear weapon.

Australia's OPAL reactor, located in New South Wales, is the only site in the world that is irradiating low-enriched ^{235}U targets, which pose less of a proliferation threat and also produce less hazardous waste. However, the University of Missouri in Columbia is seeking funding to upgrade its research reactor, which proponents say could start ^{99}Mo production from low-enriched uranium by the end of 2011. Meanwhile, the reactor-construction firm Babcock & Wilcox in Lynchburg, Virginia, in partnership with the medical-isotope supplier Covidien of Mansfield, Massachusetts, wants to build a series of small nuclear reactors with a novel liquid core that also relies on low-enriched uranium.

Ultimately, governments now need to secure the long-term supply of medical isotopes, says Graham. "There needs to be a major effort in several countries to either convert existing reactors or build new reactors to ensure a stable long-term supply of ^{99}Mo." ■

Paula Gould

HAVE YOUR SAY
Comment on any of our News stories, online.
www.nature.com/news

G8 climate target questioned

The path to a meaningful deal at the Copenhagen climate summit in December seems more treacherous in the wake of last week's meeting of the G8 nations.

ROAD TO COPENHAGEN

Leaders gathered in the quake-struck Italian town of L'Aquila promised to try to prevent global temperatures from rising more than 2 °C above pre-industrial temperatures. They also said that rich countries would cut their aggregated greenhouse-gas emissions by 80% by 2050, and, with other countries, would cut global emissions by 50% over the same period. But they made no commitment to reducing emissions any earlier and did not agree on a base year against which to measure national emissions cuts.

The choice of base year "matters a great deal", says Gwyn Prins, an economist at the London School of Economics. A recent year, say 2005, would make targets easier to achieve for countries such as Canada, Japan and the United States, whose economies and emissions have grown constantly over the past 20 years. It would be even more attractive for booming countries such as China and India. But the European Union would benefit from a 1990 baseline — as emissions there have fallen since then — as would Russia, which saw a dramatic industrial decline in the 1990s.

"So much for science — the rest is up to politicians and voters."

Others argue that the 2 °C limit is, in fact, a crucial commitment because it marks the first time G8 countries have agreed on what constitutes 'dangerous' climate change. Two degrees warming translates with "mathematical clarity" into the required "decarbonization" of the global economy, says Hans Joachim Schellnhuber, director of the Potsdam Institute for Climate Impact Research in Germany. A study earlier this year, for example, suggested that there is a 50% chance that global warming will exceed 2 °C by the end of the century if around 1,400 billion tonnes of carbon dioxide is emitted during the period 2000–50 (M. Meinshausen *et al. Nature* 458, 1158–1162; 2009). Emissions for 2000–06 alone amounted to around 234 billion tonnes.

"We have now a politically accepted and science-based threshold that allows us to calculate precisely how much greenhouse gas we can still afford to emit if we don't want to exceed a given probability of getting into dangerous territory," Schellnhuber says. "So much for science — the rest is up to politicians and voters."

Packed schedule

However, much remains to be done before the Copenhagen meeting in December, when nations will meet to negotiate a successor treaty to the 1997 Kyoto Protocol on climate change.

Last week, China and India reiterated their unwillingness to sign up to binding emissions cuts. If that stance does not alter, it is likely to jeopardize a bill to introduce a cap-and-trade programme to cut US emissions by 83% from 2005 levels by 2050.

The US House of Representatives last month passed its version of the bill. But Senator James Inhofe (Republican, Oklahoma), a senior member of the panel drafting the Senate's version of the bill, warned last week that "unless supporters of cap-and-trade legislation can develop a plan to persuade China and India to make meaningful emissions reductions on par with the United States, no such bill will pass the US Senate".

Serious questions also remain over whether the 80% reduction goal agreed by the G8 nations can be achieved using 'indirect' policy tools such as emissions trading and carbon 'offsetting' by land-use change and forestry projects. The European Union's emissions trading scheme, for example, has failed to deliver substantial emissions reductions since its launch in 2005.

Prins doubts that such policies, including cap and trade, will work (see G. Prins and S. Rayner *Nature* 449, 973–975; 2007). "We don't need meaningless targets, but real action on carbon intensity and energy efficiency," he says. "But this would require leaders admitting there's something fundamentally wrong. I'm afraid this won't happen unless the facts eventually slap us in the face." ■

Quirin Schiermeier

See Editorial, page 307, and www.nature.com/roadtocopenhagen for more climate coverage.

SPECIAL REPORT

Shooting for the Moon

The Apollo programme inspired thousands of people to pursue careers in science. Today, they still support human spacefaring — but baulk at the price. **Richard Monastersky** reports on the results of a *Nature* poll.

Humans have not walked on the Moon since the *Apollo 17* mission of 1972.

NASA

In the heady days after the *Apollo 11* lunar landing on 20 July 1969, countless children travelled to the Moon by cupping their hands over their mouths and announcing: "Houston, Tranquility Base, here. The Eagle has landed." With giant strides, they bounded across the landscape just like their heroes, Neil Armstrong and Buzz Aldrin. They practised lift-off by counting backwards to zero and blasting themselves skywards. And many of them landed, decades later, in scientific careers.

In fact, the Moon landings deserve credit for motivating a large fraction of today's scientists, according to a survey of almost 800 researchers who have published in *Nature* in the past three years[1].

Half of the researchers who responded to an online poll last month said that the Apollo missions had inspired them to pursue science — and not just in astronomy or planetary science (see 'Survey respondents'). "I became completely space crazy," said one life scientist. "I was certain I'd be an astronaut. My interest shifted to biology, but I still believe *Apollo 11* was a major influence on me."

The survey, intended to capture broad impressions rather than to precisely measure attitudes, also reveals surprisingly widespread support among scientists for human space exploration. Nearly 80% of scientists polled said that there are scientific research justifications for continuing human spaceflight.

More than 80% felt that the life sciences, physical sciences, engineering and human physiology all benefited to some degree from human spaceflight, and almost 90% said that it still inspires younger generations to study science (see 'Survey results'). Their impressions are not without justification. The US National Research Council concluded in 2003 that work in microgravity has had a major effect on several areas of physical sciences, including research into fluids, combustion and crystal growth[2]. The Neurolab experiments, which flew on NASA's space shuttle *Columbia* in 1998, are regarded as the high point of life-sciences research in space and resulted in 100 publications, according to Joan Vernikos, former director of life sciences at NASA.

Up, up and away

In principle, the scientists who participated in the poll seemed to support manned space programmes, with 62% of Americans and 83% of Europeans rating their own country's expenditures on human spaceflight as "about right" or "not enough". The stronger support in Europe perhaps reflects a vast difference in spending on either side of the Atlantic. Including money provided by the stimulus package, NASA will pour US$9.7 billion into activities related to human spaceflight in the current fiscal year, roughly half of the agency's $18.7-billion budget. The European Space Agency (ESA) allocates only €481 million ($670 million), about 13% of its €3.6-billion ($5.0 billion) budget.

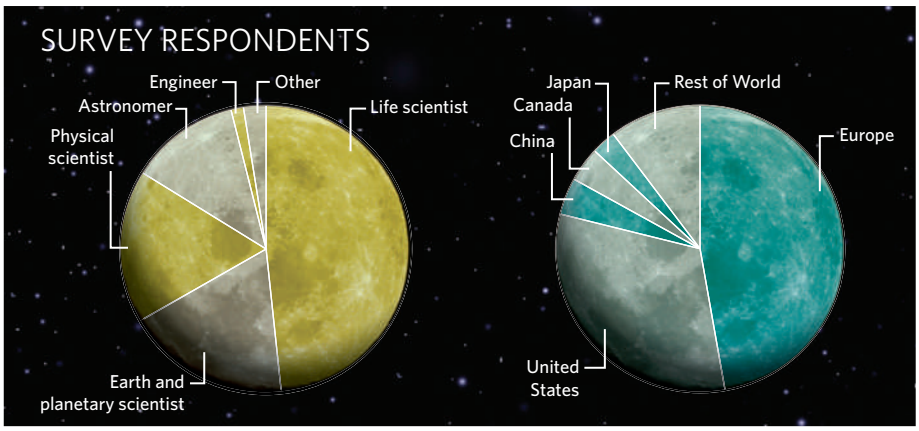
But support for sending people to space flagged when participants were asked to prioritize spending. Nearly 80% would welcome a shift in spending from human spaceflight to unmanned space missions, whereas 70% would prefer some of that cash to be spent on other areas of scientific research.

David Southwood, director of science and robotic missions for ESA, says that the affection for shifting money rests on a false premise, because money taken from human spaceflight would not go to unmanned projects. Indeed, written responses from several of those who completed the survey revealed the opposite sentiment: some supported human spaceflight thinking it would increase spending generally on space research.

Southwood says that the positive attitudes seen in the *Nature* poll reflect the broad impact of human spaceflight, beyond simply yielding data for science. "It gives humans a vicarious means of breaking the surly bonds of gravity," he says, borrowing a phrase from the pilot–poet John Gillespie Magee.

During the lead-up to the Apollo missions, scientists had similar outlooks to those seen in the current survey. A study conducted by the American Association for the Advancement of Science in 1964 found that scientists generally supported the goal of landing humans on the Moon, but they objected to the tight timetable and the extreme costs[3]. At the time, the NASA budget consumed 4.3% of the total US budget; it now accounts for

NASA



Moondreams

"I was a small boy living in provincial New Zealand when the Apollo missions were going on, and the astronauts and scientists behind that remarkable series of missions were the one group of people who genuinely inspired me." *Earth and planetary scientist, Oceania*

"There were no women involved that I could see, so I wanted to be the first female astronaut, or at least show them that women could be scientists too." *Life scientist, US*

"We can do everything we need to do using robotics much cheaper." *Earth and planetary scientist, Europe*

"Putting men on the Moon meant that with science, nearly anything is possible." *Life scientist, US*

"Without human spaceflight, we will only be able to scratch the surface of space exploration." *Engineer, US*

"I knew nerds were heroes at NASA. That was inspiring." *Physical scientist, US*

just 0.5%. Scientists in the public eye have generally been more sceptical about human spaceflight than those in *Nature*'s poll, says John Logsdon, an aerospace historian at the Smithsonian Institution in Washington DC. "There have been very few scientists publicly giving this kind of support for human spaceflight," he says.

With some scientists, the support for spaceflight was quite personal (see 'Moondreams'). About 5% have applied to be an astronaut and 14% said they would pay between $10,000 and $50,000 for a brief sub-orbital trip into space.

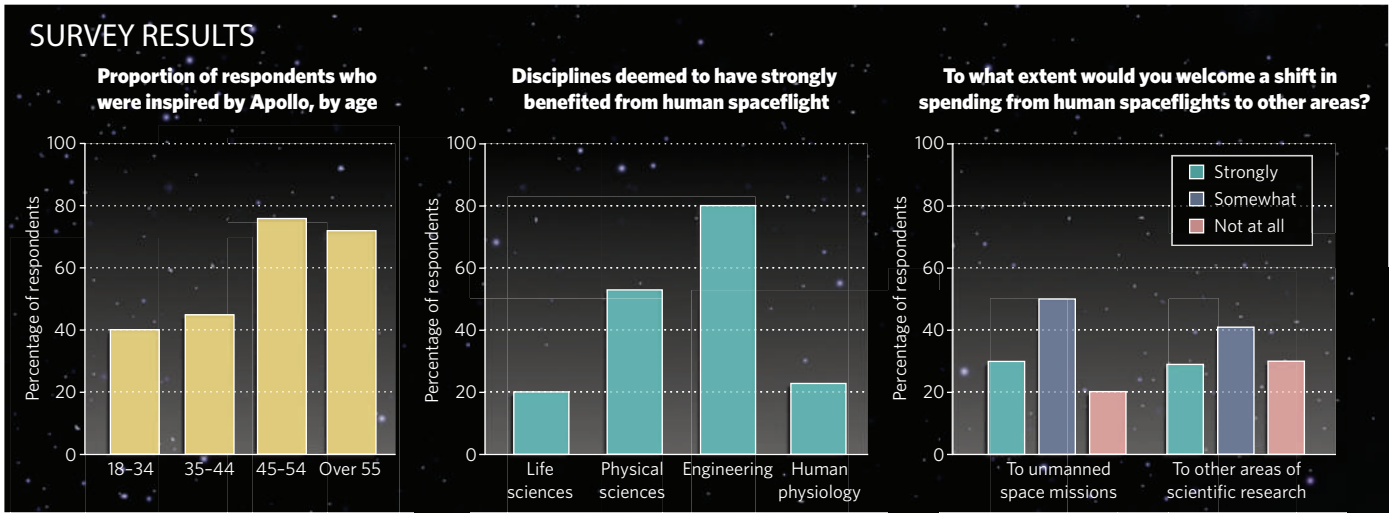
It remains to be seen whether NASA's space programme will continue to earn their enthusiasm. Next month, a panel convened by US President Barack Obama will issue its assessment of America's plans to return to the Moon by 2020 (ref. 4), and its approval of the programme is far from certain.

Yet more than 36 years after man last ventured beyond the confines of low-Earth orbit during the *Apollo 17* mission, *Nature*'s survey would suggest that many will support the next grand voyage to the Moon and beyond. ■

Survey work was aided by Sara Grimme.

1. www.nature.com/nature/newspdf/apollo_results.pdf
2. *Assessment of Directions in Microgravity and Physical Sciences Research at NASA* (National Academies Press, 2003).
3. Abelson, P. H. *Science* **145,** 539 (1964).
4. Hand, E. *Nature* **459,** 1038–1039 (2009).

See Editorial, page 307, and *Nature*'s Apollo special at www.nature.com/apollo.



Arms pact could boost US–Russian science

Nuclear scientists in Russia and the United States say that the sputtering relationship between their weapons laboratories could be rekindled by an agreement to cut the nations' nuclear-weapons arsenals.

US President Barack Obama and Russian President Dmitry Medvedev announced on 6 July that they intended to reduce their stockpile of warheads. The pair also discussed scientific collaboration as a way to strengthen ties between the two nations.

Even during the cold war, US and Soviet scientists worked together on specific issues, such as the development of fusion energy. Wider programmes began in earnest after the fall of the Soviet Union, when the US government stepped in with millions of dollars in funding for scientific collaborations.

A. ZHDANOV/UPI PHOTO/EYEVINE

Presidents Dmitry Medvedev and Barack Obama want to cut back their nations' nuclear-weapons stockpiles.

That cooperation remained strong until the new millennium, but "has dropped off sharply during the past eight years or so", says Siegfried Hecker, a metallurgist at Stanford University in California, and former director of the Los Alamos National Laboratory in New Mexico, one of the two main US nuclear-weapons labs.

In 2000, Vladimir Putin, Russia's newly elected president, began tightening up security and scaling back access to the nation's military laboratories. Meanwhile, politicians in the United States were growing increasingly sceptical of programmes that directed tax dollars into the Russian defence establishment, says David Holloway, a political scientist at Stanford University who has studied the Soviet weapons complex.

Programmes to improve the physical security and accounting of nuclear materials have survived the cooling political climate, but Hecker says that "most of the lab-to-lab scientific exchanges are gone".

Obama's promise to "reset" the US–Russian relationship has sparked hopes that these exchanges can be renewed. In March, Hecker met with other US and Russian scientists in Moscow to discuss how to renew strategic and scientific cooperation.

The collaboration is likely to be more even-handed than it was in the 1990s because conditions at Russian laboratories have improved so dramatically, according to Adriaan van der Meer, the executive director of the International Science and Technology Center in Moscow, which was set up in 1992 to foster collaboration between the former Soviet states and the West.

"We need to move to a true partnership," agrees Hecker. Among the various projects Hecker would like to see supported are efforts to develop 'safer' warheads that are more resistant to tampering and accidental detonation, improved detectors for combating nuclear smuggling and databases for tracing the origin of bomb material. There is also great potential for cooperation on civilian nuclear-power research, he says. "We have superb modelling and the Russians have excellent experimental facilities."

Holloway, who attended the March meeting in Moscow, says that scientists from both countries are clearly eager to embark on a new cooperative relationship. But he notes that much more diplomatic work must be done before researchers can start purchasing plane tickets.

Russia is currently unhappy with US intentions to place a missile defence system in Europe, for example, while the United States has halted a plan for civilian nuclear cooperation after Russian forces invaded the Republic of Georgia last year. Until the two sides begin to tackle these larger issues, says Holloway, scientific collaboration may be slow to blossom. ■

Geoff Brumfiel

Europe attacks tactics that delay generic drugs

The European Commission last week promised tougher scrutiny of drug industry deals that maximize profits on branded medicines by slowing consumers' access to cheap generic drugs.

It opened an investigation into Servier, a pharmaceutical company headquartered in Neuilly-sur-Seine, France, and several other drug manufacturers, over agreements that may have breached anti-trust laws.

Neelie Kroes, the European competition commissioner, also announced the final report of her department's 18-month inquiry into competition in the pharmaceutical sector (see http://tinyurl.com/EUreport). The report says that 'company practices' are delaying the market entry of generic drugs, but notes that shortcomings in Europe's regulatory framework might be a factor too.

An industry trade body, the European Federation of Pharmaceutical Industries and Associations, approved of the reference to regulatory obstacles. It says that the findings show a "welcome shift away from the emotive language" of an interim report released last November.

Obama names next head of US Geological Survey

A California-based geophysicist is set to become the first woman to direct the US Geological Survey (USGS) in its 130-year history.

Marcia McNutt, now president and chief executive of the Monterey Bay Aquarium Research Institute (MBARI) in Moss Landing, expects her Senate confirmation hearing by September, following her nomination last week by President Barack Obama. The USGS has an annual budget of $1 billion and 8,800 members of staff.

McNutt, an avid horsewoman who has run the MBARI since 1997, would also become the science adviser to the secretary of the interior.

Marcia McNutt is set to take the reins at the USGS.

Mars Science Laboratory devours budgets

NASA/JPL-CALTECH

A further budget overrun for NASA's Mars Science Laboratory rover (pictured), due to launch in 2011, could for the first time delay other missions in the agency's cash-strapped planetary-science division.

In a report due to be handed to Congress by the end of July, NASA will announce that the mission needs between $15 million and $115 million more than its estimated price tag — which, at $2.28 billion, is already 40% above an official $1.63-billion estimate made in 2006.

The agency has so far raided technology-development funds within the Mars programme, but if additional costs rise too much, it may have to delay two planned Moon missions. "The time for some tough decisions is here," said NASA science chief Ed Weiler, breaking the news to planetary scientists at an advisory-committee meeting on 9 July at NASA headquarters in Washington DC. Weiler also confirmed, as expected, that NASA would work in partnership with the European Space Agency on all future major Mars missions.

For a longer version of this story, see http://tinyurl.com/marscash

El Niño arrives in the Pacific Ocean

US climate scientists last week announced the arrival of El Niño, a cyclical rise in sea-surface temperatures in the central and eastern Pacific Ocean. The event is known to influence fisheries and global weather patterns.

El Niños, which are associated with a weakening in the easterly trade wind, occur every two to five years and typically last for about a year. The current one is expected to continue developing over the next several months and to last through the winter, according to the National Oceanic and Atmospheric Administration in Washington DC.

Previous El Niños have seen more rainfall over the central tropical Pacific, drought in Indonesia and powerful winter storms in California as well as flooding and mudslides in Central and South America. The phenomenon has also been linked to less hurricane activity in the Atlantic Ocean and additional winter precipitation in the arid southwestern United States.

German molecular biologist to head EMBO

Maria Leptin was last week appointed the new head of the European Molecular Biology Organization (EMBO). She is only the fifth director in the 45-year history of the organization, which has more than 1,300 members.

Leptin, who works on cell shapes and immunity at the University of Cologne, Germany, says that she hopes EMBO will improve connections with India, China and the United States, and that it can influence governments on policy matters such as stem-cell research.

She will move to EMBO's base at the European Molecular Biology Laboratory in Heidelberg, Germany, in January next year.

For an interview with Maria Leptin, see http://tinyurl.com/Leptininterview

US renewable energy gets cash boost

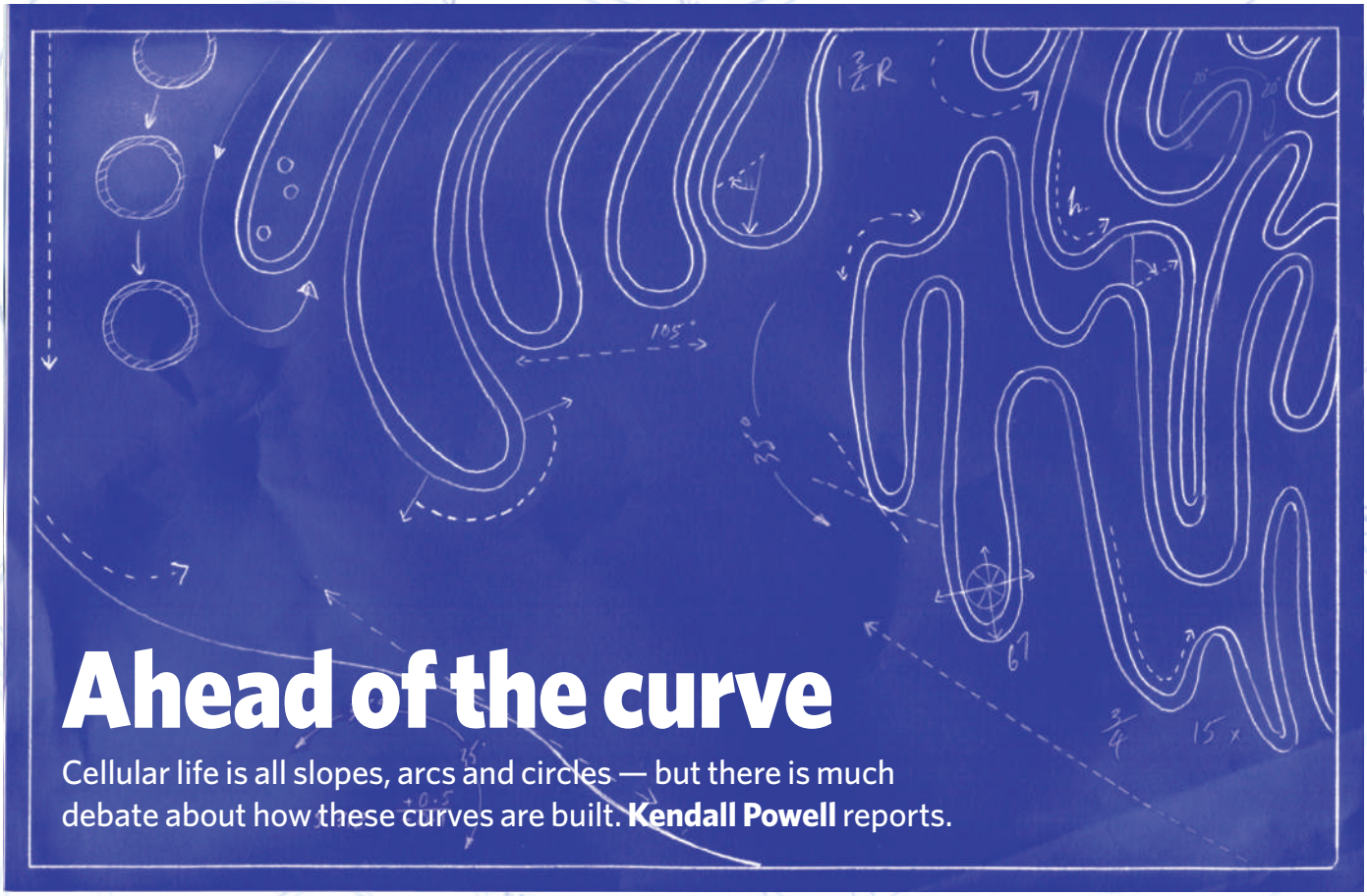
The Obama administration announced plans on 9 July to pump billions of dollars directly into the US renewable-energy market, providing a new avenue for funding that had dried up during the banking crisis.

The programme allows developers to receive cash payments, in lieu of existing tax subsidies, for wind, solar and other renewable-energy projects. Until now, such payments had tended to flow through banks, which provided upfront financing in exchange for tax credits. The new approach, sought by the renewable-energy industry and backed by Congress in its February stimulus bill, allows direct access to the subsidies.

The Department of Energy plans to begin accepting applications in August, and estimates that demand could exceed $3 billion, enabling some $10 billion to $14 billion in projects.

Correction

The News Feature 'One gene, twenty years' (*Nature* **460,** 164–169; 2009) incorrectly stated Robert Beall's current title. He is, in fact, president and chief executive of the Cystic Fibrosis Foundation.



Ahead of the curve

Cellular life is all slopes, arcs and circles — but there is much debate about how these curves are built. **Kendall Powell** reports.

To view the innards of a cell is to view architecture reminiscent of Antoni Gaudi: the gentle arc of the cell membrane, the contortions of internal tubing, the tight bubbles of vesicles. But for biologists, this architecture is an intellectual puzzle as well as a beautiful structure. Membranes generally prefer to be flat — so what, exactly, is generating all the curves?

Pietro De Camilli thought he had found one answer when, just over a decade ago, he saw a tangle of tubules in an electron micrograph. The pictures came from an experiment carried out by Kohji Takei, De Camilli's postdoc at Yale University in New Haven, Connecticut. Takei had mixed a protein called amphiphysin, which is thickly clustered at the tips of neurons, with large bubbles of artificial membranes to try to replay the process by which neurotransmitter chemicals are packaged up. It came as a surprise when the membranes snapped into masses of jumbled, twisted tubes[1]. "It looked like a plate of spaghetti, it was absolutely spectacular," says De Camilli. "It was the first example of a protein that, by itself, had dramatic membrane-deforming properties," he says.

The first, maybe — but not the last. Since this discovery, De Camilli and others have identified whole families of protein that have equally dramatic abilities to bend membranes, and a tightly knit community of researchers has built up to study them. Much debate has centred on whether the proteins create curves by wedging themselves into membranes, or by moulding them into shape using curved protein scaffolds. As it turns out, both may be at play — and more.

The issue is not just one of beautiful architecture: the ability to turn membranes into circles and tubes is central to almost every cellular process. Spherical vesicles are essential for carrying construction materials and communication signals around the cell. And the undulating network of membranes that makes up the mitochondria, endoplasmic reticulum (ER) and Golgi complex are vital to these organelles' roles in energy production, protein synthesis and protein processing, respectively. Some researchers are now finding that, once built, the cell uses its curves to position cellular processes. "The fact that the shape can also be information is very exciting," says Bruno Antonny, a biochemist at the CNRS Institute of Molecular and Cellular Pharmacology in Valbonne, France. "That you can sense curvature means that you can organize reactions in time and space."

> **"It looked like a plate of spaghetti, it was absolutely spectacular."**
> — Pietro De Camilli

Walling off spaces by wrapping membranes around them was also necessary for the evolution of cellular and organismal complexity, as it allowed cells to adopt specialized functions and trafficking to occur between compartments — the hallmarks of eukaryotic cells, says Harvey McMahon of the Medical Research Council Laboratory of Molecular Biology in Cambridge, UK. "Evolution from the primordial single-celled organisms to multicellular organisms was made possible by the appearance of small membrane-bound spaces in the cell," he says.

A matter of scale

The closest that most cells come to having straight lines are the surfaces of the encircling plasma membrane. On tightly packed and column-shaped cells, these are virtually flat; and even on curved cells, a particular spot on the circumference still seems level to the molecules there, much like the surface of Earth does to a person standing on it. Inside the cell, however, microscopy has revealed a landscape with much sharper contours. Yet for many years researchers gave little thought to how such curves were generated: cell membranes were thought of as fluid, pliable structures that could easily be pushed or pulled into shape.

That view changed in the early 1970s when physicist Wolfgang Helfrich, one of those behind the invention of the liquid-crystal display, proposed a model showing that membranes are relatively rigid structures that require substantial energy to contort[2], just as a liquid crystal does. Cell membranes are made up of lipid bilayers: two tightly packed rows of lipid molecules lined up with their hydrophilic lipid head

ILLUSTRATIONS BY CHRISTINE BERRIE

groups facing outwards, and water-repellent fatty-acid chains sandwiched on the inside. Work by Helfrich and other biophysicists has since shown that lipids prefer to be lined up in planes, and it takes a lot of energy to create the kind of disturbance that comes with bending a plane into a cylinder or sphere. The situation is similar to blowing a soap bubble from a circular wand — it takes continuous blowing to deform the flat soap film into a sphere.

Researchers spent many years looking for scaffolding proteins that could sustain a membrane curve, but none emerged that could, according to biophysicists' calculations, provide sufficient energy. De Camilli says that it wasn't until the 1999 amphiphysin experiment that biologists began to appreciate that proteins could be 'massaging' membranes into shape rather than forcing them. What made that discovery all the more intriguing to De Camilli was that amphiphysin is normally found floating in the cell cytosol, not integrated into membranes. He wondered whether there were more proteins that could distort membranes so strongly, and how they did it.

By 2002, De Camilli's group and others had identified a few more proteins that had membrane-flexing abilities[3–5] — and saw that they shared an intriguing spiral structure called an amphipathic helix. Looking down the barrel of the helix, one half is charged and the other half uncharged. Put the protein near a bilayer and it will immediately insert itself lengthwise with its charged half among the lipid head groups and the uncharged half nestled within the fatty-acid chains. This suggested that insertion into the outer half of the bilayer would create a wedge, and if enough wedges were inserted then the membrane would bend (see graphic, overleaf). It was an instantly appealing idea and one that De Camilli promoted — but it was about to be joined by an equally appealing rival.

In 2004, a group led by McMahon and his colleague Phil Evans solved the three-dimensional structure of the 'BAR' domain of amphiphysin, which includes the amphipathic helix[6]. The helix didn't crystallize well, but the rest of the BAR domain did — and their work showed that it was crescent shaped. "We could immediately see this banana shape, that, to me, is one of the most beautiful structures we've solved," McMahon says. Beauty aside, the banana suggested an alternative way for the BAR domain to bend membranes. The concave face of the banana is positively charged, which would make it stick to the negatively charged outer surface of membranes and mould them to their own curved shape.

The bananas did not go down well at meetings, and McMahon recalls more than a few scoffs and scowls. The idea that they were acting as scaffolding seemed too similar to previous scaffolding models, which the physicists had already proved wrong. Clearly BAR domains were important: evidence was piling up that they were a huge family and the various relatives could subject membranes to all sorts of bending. But were the banana-shaped molecules generating the curvature or simply stabilizing or sensing the curvature after it had been produced by insertion of the amphipathic helix?

"We could immediately see this banana shape, that, to me, is one of the most beautiful structures we've solved."
— Harvey McMahon

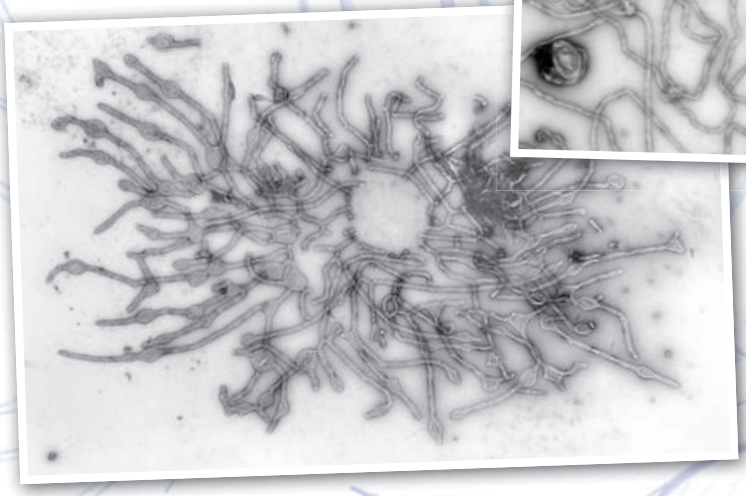
The question has inspired vigorous debate and some flip-flopping between ideas. Although De Camilli says he was originally sceptical of the scaffolding concept, his work now promotes it. Last year, his group, in collaboration with Yale colleague Vinzenz Unger, published stunning cryo-electron microscopy images of BAR domains without amphipathic helices. The domains stacked up end to end to form a spiral around membrane cylinders much like a barber's pole[7], suggesting that scaffolding alone can generate curvature — at least at the artificially high concentrations present in the test tube.

Rods and curves

McMahon now raises an eyebrow at De Camilli's latest model for how thousands of banana domains could generate curvature alone. Work from his lab has shown that this domain is much less efficient at forming membrane tubes in the lab dish if the amphipathic helix has been lopped off. "This was the first demonstration that the two functions might be joined in one protein," says McMahon.

Recently, McMahon has started to promote the wedging idea. He teamed up with biophysicist Michael Kozlov at Tel Aviv University in Israel, who calculated that insertion of an amphipathic helix alone is energetically sufficient to generate curves, whereas scaffolding alone is not. The study showed that the most efficient way to bend a membrane is to insert a short rod shape, similar to the amphipathic helix, into the outer layer of the membrane at a depth — and this is key — of around one-third of the membrane's thickness[8]. Their calculations predicted that the bending ability of BAR domains can be credited solely to the insertion of the amphipathic helix. "Now the feeling is that this banana-like shape stabilizes the curvature that is generated by some sort of insertion into the membrane," says Kozlov. Speaking of McMahon, De Camilli says, "Both of us got it partially right and partially wrong, so it has been a humbling experience."

Just as researchers were finding something to agree on about BAR domains, a whole new hypothesis for generating curves burst onto the scene. This one came from the lab of Tom Rapoport at Harvard Medical School in Boston, Massachusetts. In the late 1990s, his graduate student Lars Dreier was mixing nuclear-envelope membranes with DNA, attempting to imitate in a test tube the way that the envelope reforms after cell division. In the background of the microscope slide, Dreier noticed a network of membrane tubules forming that closely resembled the ER, the main site of protein synthesis in the cell. When



Oodles of tubules: amphiphysin triggers cell membranes to assemble into tubes.

K. FARSAD & P. DE CAMILLI CURR. OPIN. CELL BIOL. 15, 372–381 (2003)

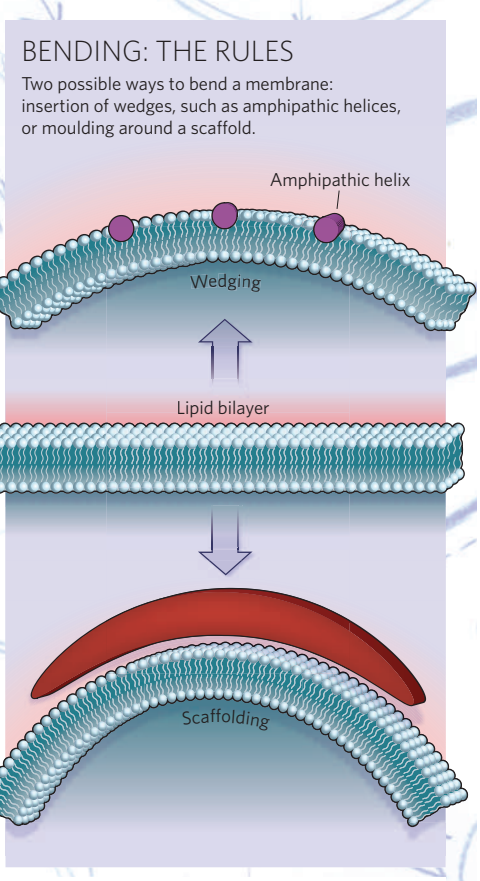
REF. 1

Rapoport saw it, he says, "I ran around the lab asking everyone, 'Did you see this? It's fantastic!' and getting everyone to look through the microscope". The experiment seemed to have triggered membranes into self-assembling an ER, which is normally continuous with the nuclear envelope membrane[9].

The excitement only intensified when Gia Voeltz, then one of Rapoport's postdocs, isolated a class of proteins responsible for curving the tubules in the ER[10]. Called reticulons, these proteins have neither a BAR domain nor an amphipathic helix and there is much discussion as to how they work. They do form a double hairpin-loop structure that inserts partway into the membrane, and Rapoport and Voeltz, along with De Camilli and Kozlov, think that these hairpins may act as a wedge. (Kozlov has calculated that the hairpins probably insert to the magical one-third depth.)

The barber pole effect

Because reticulons are sunk permanently into the membrane, Voeltz, who is now at the University of Colorado, Boulder, says that they and proteins like them might be better suited for the long-term maintenance of an organelle's tubular shape than a protein such as amphiphysin, which inserts temporarily as a budding vesicle forms. She originally thought that the reticulons might stack up to make a "nice little barber pole" scaffold around the ER tubes. But now, she's pretty sure it's not that simple — and when she pulls up her lab's latest cryo-electron microscopy images of the ER tubes, it's easy to see why. In these images, which show the ER's three-dimensional structure at nanometre-scale, the tubules look nothing like neat cylinders and more like gnarly tree roots with wide and narrow stretches. Voeltz now thinks that a cluster of several reticulon molecules may help to stabilize the tubes by forming a half-ring at the narrowest points of constriction. This fits with the idea that cells have evolved multiple wedges, with different shapes and bending abilities, to fit different purposes.



McMahon now says it was clear early on, "that we were going to have a whole repertoire of proteins — some driving curvature, some limiting or stabilizing curvature, and some sensing curvature". And Antonny's group has been focused on the sensing part of the repertoire. He is trying to understand how the bend of a membrane can act as a signal, such that its position, camber or direction can recruit additional proteins to a particular point on an organelle.

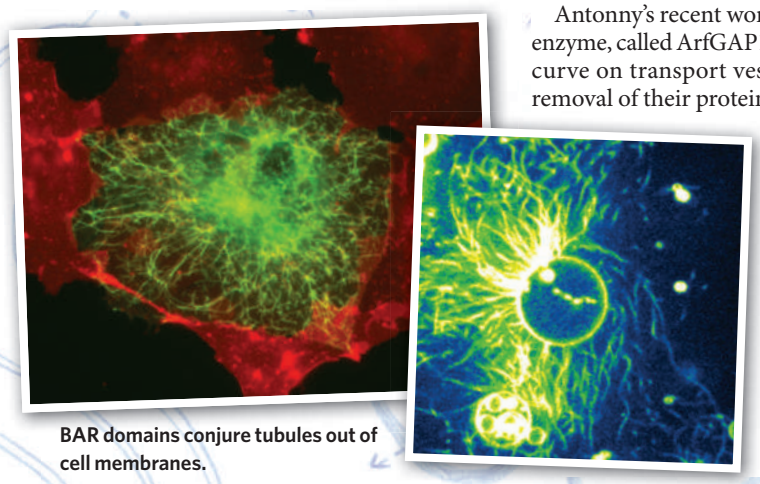
Antonny's recent work has focused on one enzyme, called ArfGAP1, that detects the tight curve on transport vesicles and directs the removal of their protein coat before they can fuse with their destination membrane. His team showed that ArfGAP1 has greater activity when bound to highly curved membrane spheres — with the same diameter as transport vesicles — than to ones with a broader diameter and a gentler curve[11]. The degree of curvature tells the enzyme that there's a vesicle that needs uncoating, Antonny says, so that it binds only to vesicles and not to other curved surfaces.

By chopping ArfGAP1 into smaller bits, the group found that its key curve-sensing stretch was an amphipathic helix. "But the chemistry of the helix was almost the opposite of what you would expect," he says. Unlike that in the BAR domain, the helix in ArfGAP1 has almost no positive charge on one side and lacks the strong attraction to the membrane[12]. Antonny compares the protein's behaviour to that of a nervous swimmer about to take the plunge. "It's as if the molecule is shy and the water is not warm enough when the membrane is flat. But when you bend the membrane, the molecule can sense the lipid packing defect and inserts." The molecule's shyness is what makes it a good sensor, he explains. It dives in only when the amount of curvature is right and gets out when that curvature changes.

Last year, Antonny's lab found a similar curvature-sensing helix that may help to maintain the shape of the Golgi complex[13], an organelle responsible for processing and trafficking proteins and other large molecules in the cell. The Golgi is a series of flattened membrane sacs stacked up like pancakes, with transport vesicles constantly budding off from its fringes. How it maintains its architecture in the midst of all this trafficking has been a puzzle. The new helix "is a start to explaining why the Golgi has this beautiful architecture", says Antonny.

Explaining beautiful — but functional — architecture is after all what the whole field is about. "They strike us, they surprise us, but we find a remarkably appealing harmony in such buildings," says De Camilli of Gaudi's constructions. And cells, he says, are the same. "I think the beauty of great architecture, like the beauty of cellular structures, resonates in us precisely because they build on natural physical principles." ■

Kendall Powell is a freelance science writer based in Broomfield, Colorado.



T. ITOH, K. ERDMANN, P. DE CAMILLI

BAR domains conjure tubules out of cell membranes.

M. WU, P. DE CAMILLI

1. Takei, K., Slepnev, V. I., Haucke, V. & De Camilli, P. *Nature Cell Biol.* **1,** 33–39 (1999).
2. Helfrich, W. *Z Naturforsch C* **28,** 693–703 (1973).
3. Farsad K. *et al. J. Cell Biol.* **155,** 193–200 (2001).
4. Ford, M. G. J. *et al. Nature* **419,** 361–366 (2002).
5. Antonny, B. *Curr. Opin. Cell Biol.* **18,** 386–394 (2006).
6. Peter, B. J. *et al. Science* **303,** 495–499 (2004).
7. Frost, A. *et al. Cell* **132,** 807–817 (2008).
8. Campelo, F., McMahon, H. T. & Kozlov, M. M. *Biophys. J.* **95,** 2325–2339 (2008).
9. Dreier, L. & Rapoport, T. A. *J. Cell Biol.* **148,** 883–898 (2000).
10. Voeltz, G. K., Prinz, W. A., Shibata, Y., Rist, J. M. & Rapoport, T. A. *Cell* **124,** 573–586 (2006).
11. Bigay, J., Gounon, P., Robineau, S. & Antonny, B. *Nature* **426,** 563–566 (2003).
12. Drin, G. *et al. Nature Struct. Mol. Biol.* **14,** 138–146 (2007).
13. Drin, G. *et al. Science* **320,** 670–673 (2008).

A lakeful of trouble

Africa's Lake Kivu contains vast quantities of gas, which makes it both dangerous and valuable. **Anjali Nayar** asks whether it is possible to tap the gas without causing a disaster.

C. ALS/BERLINGSKE/PANOS

In late 2001, Mount Nyiragongo in the Democratic Republic of the Congo (DRC) was growing restless. Plumes of smoke issued from the central crater, alarming volcanologists in the nearby city of Goma. Then, on 17 January 2002, lava fountained from a fracture on Nyiragongo's southern flank. The molten rock snaked down the sides of the volcano and razed the centre of Goma, engulfing houses and setting off a string of explosions at fuel stations and power plants. That evening, the lava streamed into nearby Lake Kivu, generating a plume of water vapour that clouded the area for days[1].

More than 100 people were killed and nearly 300,000 people fled their homes. The only obvious refuge for the displaced people was along the shores of the lake. But Kivu poses its own threats. Beneath its placid surface, the lake contains 300 cubic kilometres of carbon dioxide and 60 cubic kilometres of methane. A disruption to the lake — such as a bigger, closer eruption — could cause a gas burst, with potentially deadly consequences for the roughly 2 million people who live along Kivu's shores. The risks are hard to quantify, however. Although scientists have studied the lake for decades, basic details about Kivu and its gases are still relatively scarce, and there is now debate about how hazardous the situation is.

The issue is complicated by the lake's economic potential. The valuable methane dissolved in the water has started a feeding frenzy among energy companies working with the DRC and Rwanda, the other nation bordering the lake. In deals worth hundreds of millions of dollars, companies have started to siphon off the methane, in some cases working with the very scientists who have been assessing the lake's hazards. Proponents say that those degassing efforts will reduce the risk of gas eruption, but some researchers are worried that schemes to extract methane could make the situation more dangerous if they upset the lake's equilibrium.

"It could be one of the great remediation projects of all time: mitigating a lethal natural hazard and at the same time bringing power to people who desperately need it," says George Kling, a biogeochemist from the University of Michigan in Ann Arbor. "If it is done right."

> **"It could be one of the great remediation projects of all time."**
> — George Kling

An expanding problem

Lake Kivu lies in the Great Rift Valley, where tectonic forces are slowly ripping Africa apart. That movement brings up molten rock, which releases carbon dioxide that seeps into the bottom of Lake Kivu. Bacteria convert some of the carbon dioxide into methane, and other bacteria produce methane by breaking down organic matter in the deep waters (see graphic, overleaf).

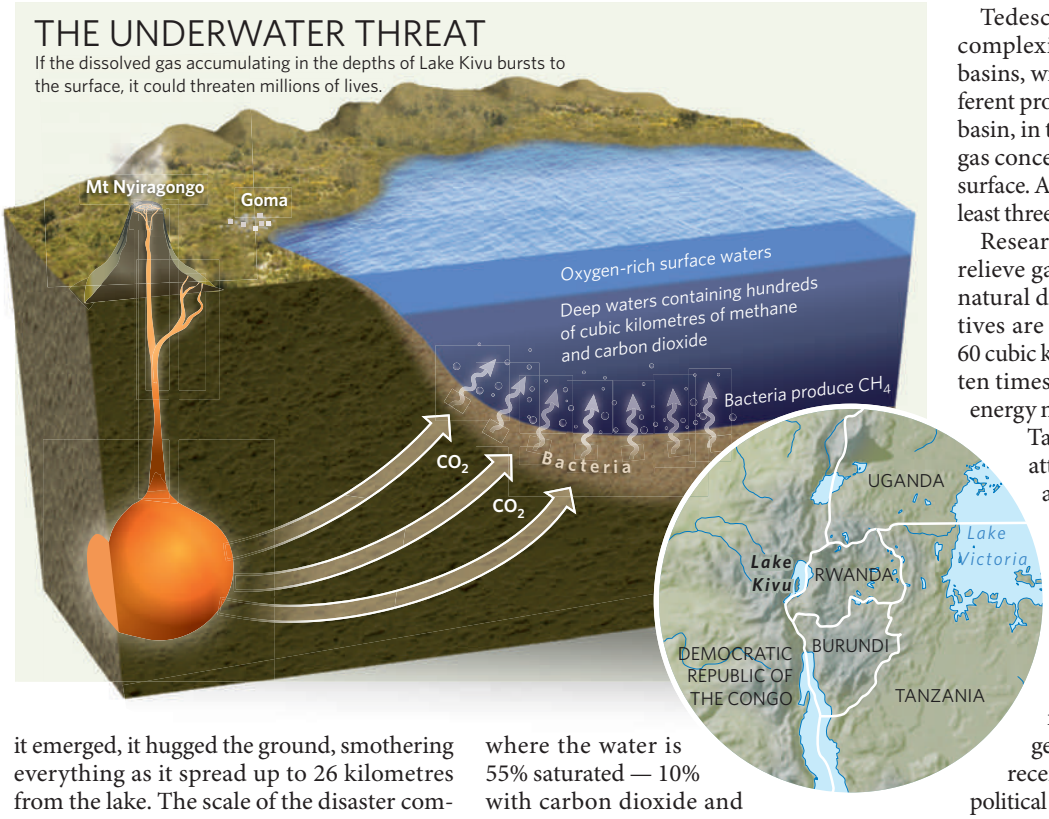
Kivu is permanently stratified, with layers of dense salt-rich water below fresh water at the surface. Deeper than about 50–80 metres, the lake is anoxic and the concentrations of dissolved carbon dioxide and methane increase with depth[2]. The differences in density prevent the layers of water from vertically mixing, and so trap the gases at the bottom of the lake.

Residents around the lake have known about the dissolved gases for many decades, but it wasn't generally thought to be a hazard. Then, in 1984, carbon dioxide erupted from Lake Monoun in Cameroon, killing 37 people. Two years later, another Cameroonian lake, Lake Nyos, spat up 0.3–1 cubic kilometres of carbon dioxide, asphyxiating more than 1,700 people.

Kling was part of a team that visited Lake Nyos in the weeks following the eruption. "The animals were all dead, thousands of cattle just lying about," he says. Kling had been to Nyos the year before, but had only sampled surface waters. "We knew nothing about the gas bomb in the bottom of the lake," he says.

It turned out that the deep waters of Lake Nyos were nearly saturated with carbon dioxide and, like in Lake Kivu, the gases were kept in solution by the pressure of the overlying water. Kling postulates that a landslide disturbed the lake's stratification, forcing gas-rich waters to move upwards[3]. That started a chain reaction. The reduction in pressure caused carbon dioxide to come out of solution and form bubbles, much like what happens when a bottle of champagne is uncorked. The rising bubbles dragged up the surrounding water, which also degassed, leading to a violent gas burst — a limnic eruption.

Carbon dioxide is denser than air so, when



it emerged, it hugged the ground, smothering everything as it spread up to 26 kilometres from the lake. The scale of the disaster compelled scientists to assess the risk at Kivu, the other lake known for its dissolved gases.

Latent threat

There are no historic records of limnic eruptions in Lake Kivu. But gaps in layers of plankton fossils at the bottom of the lake suggest that such paroxysms have struck several times in the past 5,000 years[4]. If Kivu were to undergo a limnic eruption soon, it would dwarf the Nyos disaster. Kivu is more than 3,000 times larger and contains more than 350 times as much gas as was released by Lake Nyos. Kivu's shores are also densely populated. "Kivu is basically the nasty big brother of Nyos," says Kling. But there is no scientific consensus on the current risks of that kind of a disaster.

Gas concentrations measured in 1974 and a limited study conducted in 2004 (ref. 5) show that there has been a 15–20% increase in methane and a 10% increase in carbon dioxide levels in the lake in the past 30 years, says Martin Schmid, a researcher at the Swiss Federal Institute of Aquatic Science and Technology in Kastanienbaum. If this trend continues, the lake will be saturated within the century and, like Lake Nyos, it could erupt with even the slightest disturbance. At the moment, however, the closest Kivu comes to gas overload is at a depth of 330 metres, where the water is 55% saturated — 10% with carbon dioxide and 45% with methane — says Schmid, so an eruption is less likely. (Methane contributes most of the gas pressure and the risk of eruption because it is less soluble than carbon dioxide.) According to modelling work, only an intense eruption in the gas-rich depths would be powerful enough to overturn the lake[6]. The 2002 eruption barely affected its stability because the magma did not reach those depths[7]. "The probability is low," says Schmid. "At least I am not afraid of swimming in Lake Kivu."

> **"I'm not afraid of swimming in Lake Kivu."**
> — Martin Schmid

But Dario Tedesco, a volcanologist from the Second University of Naples in Caserta, Italy, who is writing the DRC's Mount Nyiragongo eruption contingency plan for the United Nations, says an eruption at the bottom of Lake Kivu is a possibility. Tedesco has studied bathymetric surveys of the lake and found cone-like structures that are probably volcanic in origin, he says.

That evidence matches some other signs. During the 2002 Nyiragongo eruption, new fractures opened on the south side of the volcano, just a few kilometres from the lake. The lava from the fractures was also compositionally different from the volcano's crater lake[1], suggesting that there are separate reservoirs of magma in the region, some of which could extend under the lake. "Nyiragongo is going to erupt again," says Tedesco. "The only real question is where."

Tedesco's research at Kivu highlights its complexity. The lake contains at least five basins, with different characteristics and different probabilities of turnover[2]. The Kabuno basin, in the lake's northwest corner, has high gas concentrations only 12 metres below the surface. An eruption in Kabuno could release at least three times more gas than Lake Nyos did.

Researchers agree that it is important to relieve gas pressure at Lake Kivu to avoid a natural disaster — and the economic incentives are pushing that work forwards. The 60 cubic kilometres of methane equals roughly ten times the combined annual commercial energy needs of both the DRC and Rwanda. Tapping that reserve is particularly attractive to energy-starved Rwanda, and Lake Kivu has become the centrepiece of the country's plans to expand electricity production.

Extracting methane from Kivu is not a new idea. A brewery in Rwanda burned the lake's methane to heat its boilers for 40 years before it shifted to electricity. But commercial interest in using methane to generate electricity only burgeoned in recent years, in part because of growing political stability in Rwanda. Around 60 companies, most of them from foreign countries, have approached the government for access to the lake since 2005, says Albert Butare, Rwanda's minister responsible for energy, water and sanitation.

Rwanda has already handed out methane concessions totalling hundreds of megawatts to five consortia, including a US$325 million, 100-megawatt deal with ContourGlobal, an energy company based in New York. And in June, Rwanda and the DRC announced a new joint plan to develop an additional 200 megawatts from the lake.

The growing interest in Kivu has kept scientists busy discussing best practices for methane removal. Most of the proposed models use a floating platform to suspend a vertical pipe down into the gas-rich layers. A small pump initially pulls up some of the bottom water to lower pressures, until it becomes saturated with gas and starts forming bubbles. After this priming, the bubbling drives water up the pipe without additional pumping. The extraction works like a controlled limnic eruption.

The methane, being less soluble than the carbon dioxide, comes out of solution first. It is then piped onshore where it is burned to generate electricity. The problem is what to do with the carbon-dioxide-laden water.

From the standpoint of safety, it would be ideal to extract the carbon dioxide and reinsert the degassed water into the deep parts of the

Close call: lava from Mount Nyiragongo didn't go deep enough into lake Kivu in January 2002 to set off a limnic eruption, but next time it could.

G. MULALA/REUTERS

lake, where it wouldn't disturb the equilibrium, says Kling. But removing the carbon dioxide makes the water lighter, hence less stable at depth. "We don't want to generate any sinking or rising of water masses that will cause mixing," he says. Only water from the very bottom of the lake would still be dense enough to be reinserted in the deepest layers once it was degassed, but that would be prohibitively expensive. So Kling suggests keeping the carbon dioxide in the water. "Every solution is a compromise," he says.

Conflict of interests?

Klaus Tietze, a geophysicist and the director of Physik-Design-Technology, a consulting company based in Celle, Germany, argues that with carbon dioxide concentrations in the lake increasing at a rate of more than 3% a decade, "leaving the carbon dioxide in the lake is a very bad solution". He is pushing for removing both methane and carbon dioxide as quickly as possible and reinserting the water above the main gas accumulations in the lake, so that it doesn't dilute the methane resource.

Schmid discounts that plan, however, because the nutrients in the degassed deep water would overload the lake's upper layers.

With so much disagreement among scientists, the best way forwards for developers is unclear, especially because the extraction technology has been problematic. So far only one 4-megawatt platform, called KP, is sporadically generating electricity for Rwanda's national grid. Another, a 3.6-megawatt project funded by the Rwandan Investment Group, sank last year a week before it was scheduled to begin production. Some have attributed the loss to sabotage but others have blamed bad engineering.

Complicating the situation is the potential for conflicts of interest; some of the scientists who studied the lake and identified its hazard are now involved in the methane extraction projects.

For example, Michel Halbwachs, a recently retired physicist from University of Savoie in France, has spent two decades studying Nyos and Kivu and now spearheads the Rwandan Investment Group's project. He says that the engineering of the sunken platform was sound and the company is now rebuilding it. Halbwachs says that his previous work on the lake is not a conflict of interest, but rather makes him uniquely knowledgeable about how to extract gases safely.

Tietze agrees that only the few scientists who "know the whole lake" through decades of experience are qualified to run the methane projects. He is also trying to get into the extraction game and is now looking for investors.

> **"This is pure business. This has nothing to do with the hazard of the situation."**
> — Dario Tedesco

But other scientists worry that the safety of the lake could be compromised if the researchers who are assessing the gas hazard are also working on commercial methane extraction. The best interests of the lake's two million inhabitants could get lost along the way, says Tedesco, who calls the plans for tapping methane "pure business". "This has nothing to do with the hazard of the situation," he adds.

"It's a little bit like someone has dropped free money on the street and everyone's running around trying to gather it up," says Kling. He is part of an international team that is working with the World Bank and the involved nations to create rules for methane extraction in the lake. The team wants every project to be monitored through a local institute, which then reports to an international group of experts. "There has to be a separation between who is doing the work and who is checking the work," he says. In mid-June, the team came out with its first draft of legislation, which will now have to get parliamentary backing both in Rwanda and in the DRC.

Although, in theory, removing the methane should make the lake safer, it remains unclear whether the hazard will be reduced. The plans for degassing are preliminary, and no one knows how they will affect the lake's stability.

The chances of avoiding a disaster depend on many factors, some well beyond the control of the scientists or even the governments in that region. Armed militias in the DRC recently took over four of the area's seven seismic monitoring stations, hampering the ability of volcanologists to predict when Nyiragongo will erupt. And in the past few months, renewed fighting in rural areas has displaced hundreds of thousands of people.

Goma's population has almost tripled since the 2002 Nyiragongo eruption, to an estimated 1.2 million people. There are shanties crammed into every usable piece of land, many constructed from lava blocks from the volcano's last eruption. So as scientists and developers fight over Lake Kivu's methane resource, the displaced people remain pinned between a volcano, the militias and an explosive lake. ■

Anjali Nayar is an International Development Research Centre fellow at *Nature*.

1. Tedesco, D. et al. *J. Geophys. Res.* **112,** B09202 (2007).
2. Tassi, F. *et al. Geochem. Geophys. Geosyst.* **10,** Q02005 (2009).
3. Kling, G. W. *et al. Science* **236,** 169–175 (1987).
4. Haberyan, K. A. & Hecky, R. E. *Palaeogeogr.Palaeoclimatol. Palaeoecol.* **61,** 169–197 (1987).
5. Schmid, M., Halbwachs, M., Wehrli, B. & Wüest, A. *Geochem. Geophys. Geosyst.* **6,** Q07009 (2005).
6. Schmid, M. *et al. Acta Vulcanologica* **14/15,** 115–122 (2003).
7. Lorke, A., Tietze, K., Halbwachs, M. & Wuest, A. *Limnol. Oceanogr.* **49,** 778–783 (2004).

CORRESPONDENCE

Sharing different mouse strains is not always so simple

SIR — I was disappointed by the view expressed in your Editorial 'The sharing principle' (*Nature* **459,** 752; 2009) that the mouse community does not share its strains. This is untrue. Most labs are very collegial, spending a considerable amount of time and effort on distributing their mouse strains. Although there are a few labs that withhold distribution, any community may contain such individuals. The fact that a mouse strain is not found in a repository does not mean that it is not being shared.

I was also puzzled by the conclusion of May's CASIMIR workshop, noted in the Editorial, that "the sharing problem urgently needs resolution" with regard to international mouse gene-knockout projects. Such mutant alleles will mostly be archived as embryonic stem-cell lines. Readers should also realize that repositories cannot keep all their mouse strains live 'on the shelf': most strains are frozen. The cost for a user to have a strain thawed is thousands of dollars and it takes many months before the recovered mice become available. This is a big disadvantage for labs on tight budgets. With regard to funding agencies: in grant proposals to the US National Institutes of Health, for example, applicants are required to write a 'resource-sharing plan' that includes genetically modified mice.

It was suggested that sharing avoids duplication of effort. But it is essential that more than one group generates mutations in the same gene as a crosscheck. No two labs generate the same allele, and every geneticist knows that the expression of different alleles can lead to very distinct phenotypes.

Your claim that sharing mice "has never been easier" is questionable, considering all the paperwork, health certificates, veterinary screens, special serology screens, costs, time and logistics involved. This is quite different from uploading DNA sequences in the comfort of your office.

It would be great if funding agencies supplemented grants involving the generation of mouse strains to cover the costs of sending the strains to a repository. In these tough financial times, that seems unlikely.

Richard Behringer Department of Genetics, University of Texas, M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, USA
e-mail: rrb@mdanderson

For Nature*'s policy on sharing of biological materials, see http://tinyurl.com/33fg2r*

Invasion biology is a discipline that's too young to die

SIR — Mark Davis suggests in his book *Invasion Biology* that the discipline should be reabsorbed into general community ecology, as Emma Marris reports in her Review 'The end of the invasion?' (*Nature* **459,** 327–328; 2009). In fact, invasion biology has a much wider remit than the management of a few highly invasive species. The comparison of alien and native species has contributed to a better understanding of population dynamics, ecosystem function and species evolution. The field still has much to teach us about biodiversity and ecological processes in general, and its relevance to conservation policy is growing.

The acceleration of international trade and travel has led to progressive mixing of biota from across the world, introducing increasing numbers of species to new regions. Although our knowledge of the drivers and consequences of these invasions is still poor, the wisdom of distinguishing species in relation to their geographical origin is confirmed by discoveries of the often marked difference between alien and native species in behaviour, traits and impact (see, for example, P. E. Hulme *Global Ecol. Biogeogr.* **17,** 641–647; 2008). However, the position of invasion biology as a separate discipline will continue to be questioned until we have a better framework for understanding the impacts of invasive species, as well as for risk assessment and strategies to deal with them.

Plurality of discourse is, of course, valuable in any science, and Davis's book should help to instil a balanced perspective among scholars in a young field that is still defining its own unique contribution to ecological understanding.

Petr Pyšek Institute of Botany, Academy of Sciences of the Czech Republic, Zámek 1, 252 43 Průhonice, Czech Republic
e-mail: pysek@ibot.cas.cz
Philip E. Hulme Bio-Protection Research Centre, PO Box 84, Lincoln University 7647, Canterbury, New Zealand

For cancer, seek and destroy or live and let live?

SIR — In his Essay 'A change of strategy in the war on cancer' (*Nature* **459,** 508–509; 2009), Robert Gatenby proposes that controlling a tumour's growth may be more efficient in the long term than trying to eradicate it. However, it could be some time before we can establish whether such a switch in treatment strategy is clinically beneficial.

To keep cancer in check, Gatenby suggests giving progressively lower doses of cytotoxic drugs, rather than the maximum-tolerated doses used in conventional chemotherapy to try to destroy the tumour.

In fact, oncologists have for several years been using daily low doses of drugs such as etoposide or cyclophosphamide to control tumours that do not respond to chemotherapy at higher doses, an approach called metronomic therapy. Robert Kerbel and Judah Folkman and their colleagues discovered the potential of metronomic chemotherapy for restricting tumour blood-vessel growth some ten years ago (see, for example, G. Klement *et al. J. Clin. Invest.* **105,** R15–R24; 2000) and Isaiah Fidler and Lee Ellis have proposed that cancer is a chronic disease and should be treated as such (I. J. Fidler and L. M. Ellis *Nature Med.* **6,** 500–502; 2000).

But despite some promising results in adult and paediatric oncology, this approach remains marginal and limited to patients in relapse, or to those for whom conventional treatment has failed. Perhaps this is because a metronomic strategy challenges the predicted success of conventional regimens based on dose intensity.

Paradoxically, many oncologists have adopted metronomic scheduling with blockbuster drugs such as Gleevec, Tarceva and Sutent — targeted therapies administered daily at low doses — even though no one really knows whether a classical regimen of higher doses every three weeks might prove more efficient.

Gatenby's idea of 'adaptative therapy', derived largely from mathematical modelling, is a *tour de force*. He may be paving the way for advocates of traditional chemotherapy, targeted therapies and metronomic strategies to work together in the patient's best interest. We hope that stakeholders will be prepared to test Gatenby's approach and to revisit the idea of treating cancer as a chronic disease.

Nicolas André Department of Paediatric Oncology, Hôpital pour Enfants de la Timone, Bd Jean Moulin, 13385 Marseille Cedex 5, France
e-mail: nicolas.andre@ap-hm.fr
Eddy Pasquier Children's Cancer Institute Australia for Medical Research, Randwick, New South Wales 2031, Australia

Readers are welcome to comment on this topic at http://tinyurl.com/n4jn45

ESSAY

Why we need space travel

Giovanni Bignami reflects on the people who persuaded him that we must send humans beyond Earth's orbit to inspire public and political support for science.

The seasoned army officer slapped me on the back. "So it's scientists like you who made this possible," he thundered. "And we always called you sissies."

We were in the 'officers' club', a bare room with a lousy television set, in a small barracks in the Italian Alps. It was the early hours of 21 July 1969. This captain of the Alpini — an elite Italian mountain infantry corps — and I were watching the fuzzy, black and white images of the *Apollo 11* Moon landing thanks to rickety reception from an ice-axe doubling as an antenna on the roof.

I was a 25-year-old conscript. I'd been made a junior officer on the strength of my physics degree, despite my political leanings being dangerously in tune with the 1968 cultural revolution.

A few months later I was back at the University of Milan, ready to serve under a much more challenging commander: Giuseppe Occhialini, a baron of the academic community, and a demigod for me. He had just put himself in charge of Italian science in space after 30 years in fundamental particle physics.

Occhialini gave me three curt instructions: I was to go and work on a space astronomy project somewhere in the fogs of northern Europe. I was to note that all the work would take place in English (of which I could remember little). I was above all to remember a crucial truth — that science in space should be real science, not the semi-militaristic mucking around with astronauts we'd recently seen broadcast around the world.

The last order was one of the few mistakes the great man made. But it took me decades to work that out.

Meanwhile, the Vietnam War was surely and swiftly terminating the Apollo Programme. Few realize today that NASA's funding of Apollo, its biggest expenditure ever, overlapped with the United States undertaking its greatest military effort since the Second World War. After putting men on the Moon, Apollo had a futile, détente-fuelled fling with the Russian Soyuz human-spaceflight programme. Between 1972 and 1975, NASA unwisely got rid of its few remaining Saturn V launch rockets developed by Werhner Von Braun, an Americanized German physicist. At the time, they were the most powerful machines ever built and, alas, we have to re-invent them today. Six long years passed, until the shuttle era began in 1981, before the American conquerors of the Moon had access to space again.

D. PARKINS

NASA's backpedalling only confirmed my conviction that Occhialini was right: the agency seemed to have wasted more than a decade and an astronomical amount of money. And the situation, apparently, was only getting worse. The shuttle had nothing to do with science. It was designed to deliver materials and astronauts into space needed to build the Space Station — a Reaganite invention designed to crush the 'evil empire' of the Soviet Union technologically, but disingenuously marketed as a means to do science.

Throughout the 1980s and early 1990s, it was à la mode in Europe to join US scientists in shouting that the planned station was robbing us of precious space science. I joined in wholeheartedly — but in the late 1980s began to have doubts. My misgivings were triggered by the advice of a learned friend, whom I won't name in deference to his modesty.

This man was my mentor when I started working as an adviser for the European Space Agency in 1984. He instilled in me the idea that governments — of all stripes — will never part with the huge sums space demands for just 'doing science'. He helped me grasp that the funding of space science, such as the building of satellites for astronomy, happens because projects such as the station and its astronauts are easy for politicians to understand, and hugely attractive to industry. So if the whole lumbering station-plus-astronauts machine was kept rolling, then space science could ride smoothly on its back, more or less unnoticed. And it did so very well for two decades.

As the millennium turned, I saw that the already ageing International Space Station was not enough. You can't sell the future, especially in a time of crisis, on something that is decades old. You need a new mission to spark enthusiasm — such as sending people to Mars.

Which brings us to today. The Alpini captain was right: it is just ordinary people doing science that drive progress. Occhialini was wrong to pit science against astronauts. Forty years after the first Moon landing, it is clear that only a great new programme of space exploration, robotic and manned, will carry us forwards. It is the only way to keep improving on the beautiful technologies we've swiftly turned into services, from telecommunications to Global Positioning System navigation satellites.

Without astronauts and their cheering crowds to rally us along, space applications and space science will quickly wither into oblivion. Forty years from now there will be no space activities at all if we do not send people beyond Earth's orbit, where no one has been since 1972. ■

"Without astronauts and their cheering crowds, space science will quickly wither into oblivion."

Giovanni Bignami is at the Institute for Advanced Study of Pavia (IUSS), Lungo Ticino Sforza 56, 27100 Pavia, Italy, and Accademia dei Lincei, Rome, Italy. He was chairman of the ESA Space Science Advisory Committee from 2003 to 2006.
e-mail: giovanni.bignami@gmail.com

See also www.nature.com/Apollo.

BOOKS & ARTS

The slow slide towards a new battlefield?

The cold war saw governments develop international policies to regulate outer space for military and civil uses. Loopholes in those policies must now be closed, writes **Roald Sagdeev**.

Satellites form an integral part of global infrastructure through their use in telecommunications, remote sensing and navigation. They have also strengthened international security. Indeed, as James Clay Moltz explores in his book *The Politics of Space Security*, space regulation has been shaped by defence interests ever since the Soviet launch of Sputnik 1 in 1957.

Sputnik did not carry any military stigma; but, as the first artificial object in space, it opened up a potential new battlefield. The US administration understood the dangers immediately. In 1958, US President Dwight D. Eisenhower wrote to Soviet Premier Nikita Khrushchev to advocate strategic restraint and the right of any country to free access to space as part of the common heritage of humankind. Yet such advocacy did not exclude the military from space altogether; after all, satellite reconnaissance from orbit could reduce the risk of a surprise attack.

It is remarkable how the leaders of that dramatic epoch chose negotiation and treaties over a display of brute force in space. Such a spirit was expressed in the National Aeronautics and Space Act legislated by US Congress in 1958: "activities in space should be devoted to peaceful purposes for the benefit of all mankind".

In a decisive step in 1959, the United Nations established the Committee on the Peaceful Uses of Outer Space to designate the legal framework for space-related activities. In a treaty signed in 1963, the United States, the United Kingdom and the Soviet Union agreed to ban nuclear-weapons tests in the atmosphere, underwater and in space. The UN's comprehensive Outer Space Treaty, which prohibited the space deployment of weapons of mass destruction, was adopted in 1967. Yet this treaty left open the loophole of placing conventional armaments in orbit, allowing both superpowers to develop technologies for anti-satellite weapons.

The Politics of Space Security: Strategic Restraint and the Pursuit of National Interests
by James Clay Moltz
Stanford University Press: 2008. 384 pp. $70 (hbk), $35 (pbk)

Moltz dissects the debates and battles that have taken place between those who have formulated space policy. He describes, for instance, the episode when US President John F. Kennedy got the 1963 Partial Test Ban Treaty approved by Congress despite strong opposition — including that of physicist Edward Teller. In the seemingly better times after the cold war, President Bill Clinton failed to get the US Senate to ratify the Comprehensive Test Ban Treaty, which would ban all nuclear tests.

The peak of success in regulating military activity in space was reached in 1972 at the signing of the Antiballistic Missile Treaty by the United States and the Soviet Union. It was patiently championed by the US administrations of both Lyndon B. Johnson and Richard Nixon until the Soviet Union was persuaded by arguments of strategic stability. The treaty prohibited the development and deployment of missile defences — such as the 'Star Wars shield' — in orbit. The regime of deterrence was strengthened and space was left militarized, yes, but free of any weapons. The treaty even survived Ronald Reagan's later obsession with the Strategic Defense Initiative for protecting the United States from ballistic nuclear attack.

Steady technological progress made navigation from space possible, Moltz explains. Signals sent by Global Positioning System (GPS) satellites, operated by the US Air Force, have been freely available since 1995. That fulfilled the promise Reagan made in 1983 after a Korean Air flight was shot down by the Soviet Air Force in an incident allegedly caused by navigation error. For the military, GPS provides greater navigational accuracy for targeting weapons; hence sci-fi writer Arthur C. Clarke's description of the first Iraqi campaign as "the first space war".

With George W. Bush's 2001 arrival in

EUROPEAN SPACE AGENCY/SPL

Will President Barack Obama fulfil his campaign promise to tackle space debris, illustrated here?

the White House, the mix of technological exuberance and triumphalism, born of the United States' superiority in space, started to challenge the old paradigm of space security based on an international legal regime. The first signal of change came in January 2001 from a US commission on space security, chaired by Donald Rumsfeld before he became defence secretary. Warning that the United States could one day face a "space Pearl Harbor" — a devastating surprise attack on its strategically important space assets — the commission members wanted to ensure "that the President will have the option to deploy weapons in space to deter threats to, and if necessary, defend against attacks on US interests".

Rumsfeld presided over several Pentagon activities that prepared arguments for the full-spectrum use of force in space. Thus, the Air Force Space Command described in its 2004 Strategic Master Plan the possibility of using anti-satellite weapons, or even weapons that could be deployed in space to target objects on Earth. Most controversial was the US withdrawal in 2001 from the international Antiballistic Missile Treaty, which had by then been in force for three decades.

The existing set of space laws has loopholes. It does not restrict conventional weapons in space that could destroy or damage orbiting satellites. Satellites are extremely vulnerable even to primitive attacks, such as being hit at high speed by a projectile, and anti-satellite weapons exploit this fragility. Such weapons were not added to military arsenals during the cold war. However, both sides have kept open the possibility of developing these weapons to gain superiority, while each fearing that the other will overtake technologically.

To close these loopholes, experts are calling for the introduction of 'rules of the road' for orbital traffic, the prohibition of all weapons in space, the testing and deployment of anti-satellite weapons on the ground and support for international arrangements to mitigate the growing risk from space debris to assets in orbit.

China and Russia, for example, have tried during the past few years to introduce an initiative to ban space weapons in the UN Conference on Disarmament. The United States, with almost a lone voice, has blocked this discussion. China's test of an anti-satellite weapon in 2007 increased the pressure to act, and demonstrated that not everyone is ready to accept the space dominance of one country.

The collision in February this year of an active US telecommunications satellite from the Iridium constellation with a long-dead Russian one adds an alarm call. Introducing regulations that require satellites to be de-orbited at the end of their active life would need international cooperation and exchange of information on orbital traffic and debris. As with military threats, the United States has most to lose in an unregulated environment because it has the greatest presence in space.

A regulatory regime for orbiting objects can be introduced in various ways. An international law to ban space weapons, adopted by the UN, would offer a radical solution. In the interim, a regime of restraint could be agreed among those countries that are active in space. A precedent is the Missile Technology Control Regime — a voluntary association of 34 countries that has been promoted by the United States and others since the end of the cold war. Governments could also pledge not to place weapons in orbit.

Moltz ends on a pessimistic note: the long and difficult political battle to come. Proponents of space-based weaponry point to risks from North Korea, a nation seemingly embracing both nuclear weapons and seeking orbital capability. Iran's nuclear and rocket development is cited as another potential threat. However, the debate is changing thanks to the rapid rise of civilian space applications. Satellite networks for telecommunications, remote sensing and navigation are effectively global utilities. Space exploration today is also largely carried out by international collaboration, whether the mission is to Mars or the Moon. The presence of the International Space Station, which brought together the former adversaries of the cold war, itself safeguards against space becoming a battleground.

The rise of Barack Obama to the highest political orbit brings new hope. Promises he made during his presidential campaign include pursuing international negotiations to minimize space debris and banning the weaponization of space. He has declared that "the United States must show leadership by engaging other nations in discussions of how best to stop the slow slide towards a new battlefield". Such leadership is urgently needed. ■

Roald Sagdeev is a Distinguished University Professor of physics and director of the East West Science Center at the University of Maryland, College Park, Maryland 20742, USA. He was director of the Russian Space Research Institute (1973–88) and science adviser to Mikhail Gorbachev (1985–88).
e-mail: rsagdeev@gmail.com

The return path to the Moon

The Seventh Landing: Going Back to the Moon, This Time to Stay
by Michael Carroll
Springer: 2009. 115 pp. £19.99, $29.95

In *The Seventh Landing*, Michael Carroll adeptly lays out NASA's strategy for returning to the Moon, establishing a permanent outpost there, and, if all goes well, pushing on to Mars. It is a complex and ambitious undertaking, requiring a new fleet of launch vehicles, an updated crew module, a more versatile lunar lander, sophisticated all-terrain vehicles, souped-up spacesuits, lunar habitats and enough acronyms to make your head explode. Fortunately, using straightforward reportage, coupled with his own stunning space art, Carroll brings compelling order to what could have been a chaotic romp.

NASA's goal is to get moon boots on lunar soil by the year 2020. But to do so, milestones must be passed on schedule. The maiden flight of the booster rocket, Ares I, which will loft the crew module, *Orion*, into orbit, is scheduled for early 2013. *Orion*, which will accommodate a crew of up to six, may fly its first manned mission to the International Space Station by 2015. The more powerful Ares V booster will carry the lunar lander, named *Altair*, into Earth orbit. After docking with the crew module, the booster's upper stage will haul both *Orion* and *Altair* to the Moon. Ares V's maiden flight is scheduled for 2018. Deadlines, hardware and mission details are in flux but, Carroll recognizes, "the overarching goals, strategies and inspiration for the seventh landing will not change".

Those sentiments, however, may not go unsullied. Even now, under a directive by US President Barack Obama, a committee of ten aerospace executives, astronauts, engineers, scientists and a retired general are taking a down-to-earth look at the United States' space exploration strategies (*Nature* **459,** 1038–1039; 2009). Particularly sobering are the budget constraints on completing both the *Orion* module and the Ares V booster. Reservations from scientists about the value of a return to the Moon have also reached the committee's ears. Their report, which chairman Norman Augustine asserts will be unblinking in its findings, will be released in August.

After looking back at early robotic lunar missions such as the Luna and Ranger series and then the *Apollo* programme, Carroll shifts the focus from our lunar history to our potential lunar future and the justification for that

future. Whereas the *Mercury*, *Gemini* and *Apollo* programmes were politically motivated, the reasons to return are more practical. Jim Garvin, chief scientist at NASA's Goddard Space Flight Center in Greenbelt, Maryland, gives the most succinct reason: "Going to the Moon enables us to build a platform for technology that is transformational to our society, and that's not just Teflon."

As Carroll's interviews with experts show, those of us who remain on Earth will probably reap the benefits from innovations that emerge from the new Moon programme. These include advances in communications, materials science, computer technology, robotics, medical procedures and even building materials.

There are also purely scientific reasons for returning to the Moon. It is, as Carroll writes, "the Rosetta stone of the terrestrial planets", and there is still much to be learned about its geology and origins. If ice is locked up in the permanently shadowed regolith at the north or south poles, as many scientists think, it probably got there via countless impacting comets. Such material would be pristine, deposited when the Solar System was in its infancy. A study of these volatiles would prove invaluable to understanding the origins of the Solar System.

Inherent in the strategy is the long-range vision of sending humans to work and live on Mars. If we can learn to abide in a lunar environment, we should be able to deal with the harsh conditions on the Red Planet. As Carroll describes, however, many in the aerospace industry ask: why waste time and money on the Moon when those resources could be better used for a mission to Mars? This debate will not be settled for some time.

Carroll's vision of a shuttle craft coming in to land at a Mars settlement.

©MICHAEL CARROLL

What is not up for debate is the difficulty of establishing a long-term base on Mars. Just getting humans there will be a challenge. Travel times will be up to eight months, and by the time a crew lands the launch window to return home will practically be gone. Even if a crew stays for only a few weeks, they would have to wait another year and a half for the planets to reach a favourable alignment for the journey back.

Planetary scientist Chris McKay of NASA's Ames Research Center in Moffett Field, California, thinks our approach should resemble the efforts made to live and work in Antarctica. "We've been [in Antarctica] for 50 years and built a station that's designed to last at least another 30. That's the kind of mentality I would like infused into the Moon programme … and Mars base."

Carroll writes passionately that missions to the Moon and Mars are like instinctual siren calls to humans. "It is something larger than ourselves, something for the generations to come." To deny these instincts, he argues, is to deny our history and to deny our humanity.

Having won the race to the Moon with *Apollo 11* in 1969, US public interest in pushing on to Mars quickly waned. *The Seventh Landing* demonstrates that it has never died, however. Carroll's enthusiasm is infectious and will inspire readers who look back on the the *Apollo* landings as a time when mankind really did make giant leaps towards a promising future. Maybe we will finish what we started 40 years ago and set a course for Mars, either directly, or by way of the Moon. After all, that was the idea in the first place. ■

Jeff Kanipe is the author of *Cosmic Connection: How Astronomical Events Impact Life on Earth*. He lives in the Washington DC region.
e-mail: jeff.kanipe@comcast.net

For more on *Apollo*, see www.nature.com/Apollo.

APOLLO BOOKS

The anniversary of the first footstep on the Moon is being celebrated in an array of new books. Drawing on NASA's archive of oral histories, and declassified Central Intelligence Agency material on the space race, historian Craig Nelson tells in *Rocket Men* (Viking) the full story of the *Apollo* programme from the crews' training to the first moonwalkers' return as media stars.

In his straight-talking memoir *Magnificent Desolation* (Harmony), co-written with author Ken Abraham, Buzz Aldrin describes the great journey. More than 600 million people watched Aldrin become the second man to set foot on the Moon. But on his return to Earth he spiralled into depression and alcoholism.

Other *Apollo* astronauts describe their lunar experiences in Andrew Chaikin's *Voices From the Moon* (Viking Studio) and select their favourite photographs from the mission in *Apollo: Through the Eyes of the Astronauts* (Abrams), edited by Robert Jacobs and others.

NASA's official history, *Apollo Expeditions to the Moon*, edited by Edgar Cortright, is being reissued by Dover. It will include first-hand accounts from those involved in all areas of the mission and a DVD with historic space footage.

A shelf of testimonies from the men who walked the walk.

CROWN PUBLISHING GROUP; STUDIO

In Retrospect: Calvino's *Cosmicomics*

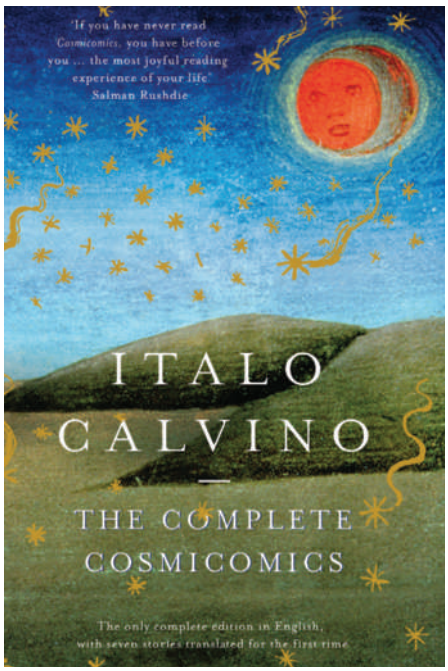
The soaring imagination of the Italian author abounds in a new compilation of his cosmic fables. Mostly written in the age of the space race, they are heavily informed by science, finds **Alan Lightman**.

In one of the stories in Italo Calvino's *The Complete Cosmicomics*, the inimitable character named Qfwfq is gambling with Dean (k)yK in the nothingness before time and space began. "I don't want to boast," announces Qfwfq, "but from the start I was willing to bet there was going to be a universe, and I hit the nail on the head … We were always gambling, the Dean and I, because there was really nothing else to do, and also because the only proof I existed was that I bet with him, and the only proof he existed was that he bet with me."

Wildly inventive, philosophical and playful at once, addicted to sustained metaphors, Calvino is almost unique in his fictions. One might compare him to José Saramago or Jorge Luis Borges, but few others. In Calvino's dazzling *Invisible Cities*, the explorer Marco Polo reports to Genghis Khan about each of the marvellous cities in the Khan's empire, the descriptions exquisite in their crystalline prose. In *The Baron in the Trees*, a man lives among the leaves and branches of trees, viewing the world from far above the ground.

The Complete Cosmicomics combines several earlier works: *Cosmicomics* (1968) and *Time and the Hunter* (1969), both translated by William Weaver; four stories from *Numbers in the Dark* (1995), translated by Tim Parks; and seven tales from *World Memory and Other Cosmicomic Stories* (1968), newly translated by Martin McLaughlin.

Most of these stories were written between 1963 and 1968, in the heat of the space race between the United States and the Soviet Union. They are heavily informed by science, especially astronomy and cosmology but also evolution and cell biology. In one of the yarns, for example, the protagonist is an ageing dinosaur who has somehow survived the extinction of his species and hides out, like a war criminal, in a tribe of descendant reptilians called 'New Ones'. The interloper is, by turns, embarrassed, secretive, defensive and finally indignant about his origins, falling in love and then being rejected by one of the new females. In another tale, two men and a woman fall through empty space for untold centuries, the narrator attempting to steer his trajectory closer to the woman and imagining a rivalry with the other man, a moustachioed lieutenant. In another, the main character is a mollusc: "I try to persuade myself that the two holes I had were a mouth and an anus … Every now and then, I was seized by fantasies, that's true, for example, the notion of scratching my armpit, or crossing my legs, or once even growing a moustache."



Colloquialisms nestle happily with scientific terminology in Calvino's playful approach.

Many of the stories concern cosmic beings who offer witty commentary on the formation of the Universe and have the wondrous capacity to casually wander across time and space. They exist in the cushiony folds of cosmic nebulae, or in empty space, or falling in the gravitational pull of distant galaxies. Calvino makes no attempt to be logically consistent or scientifically plausible, and herein arises much of the humour and charm of the book. Molluscs imagine armpits, cold exists before the Universe existed, people climb ladders and jump on to the Moon.

In 'At Daybreak' a squabbling family stumbles across "pinnacles, spires, and battlements" of condensing gas in a new-formed nebula. In 'As Long as the Sun Lasts', Qfwfq's grandparents trade insults about which stars to set up camp near: "And why not the other one? The bigger they are, the more I trust them." "Are you mad, don't you know what that is? Don't you know about the blue ones? They burn so fast, you don't even notice, and barely a couple of thousand millennia go by and you've already got to start packing." Colloquialisms nestle happily with scientific terminology.

We never know who these cosmic supermen are, where they came from, or how they got to be where they are. In *Cosmicomics*, there are no houses and chairs, no roads or cars, no banks or telephones, no Berlin or London. There is no human civilization at all. Yet we do meet intelligent beings, even families and lovers, attempting to make sense of life in its most elemental form. It seems that Calvino is attempting to fathom how much of a reality can be created without familiar geography of time and space. He plays with science the way a found-object artist throws together bits of silvered glass, odd metal brackets, bits of coloured paper.

It is impossible not to compare Calvino to another of his Italian countrymen, Primo Levi. Both wrote imaginative fiction inspired by science. Calvino's *Complete Cosmicomics* is more fanciful and playful than Levi's *Periodic Table*. Levi's characters in the latter are far richer. The most interesting of Calvino's fables, in my opinion, are the ones with human drama. The least interesting are those in which the author goes into long pedagogical riffs on science or indulges in pages of philosophical reflection. When Calvino brings together human drama with his imaginative scenarios, and philosophy with his beautiful language, humour and wit, he is unparallelled.

In the story titled 'World Memory', an organization is entrusted with the job of storing all of the world's knowledge and culture in the face of an impending cataclysm. "The duty of the director is to make sure that nothing is left out, because what is left out is as if it had never been." Fortunately for future generations of world civilization, Calvino has been, and he will certainly not be left out. ■

The Complete Cosmicomics
by Italo Calvino
Translated by William Weaver, Tim Parks and Martin McLaughlin
Penguin Classics: 2009. 304 pp. £20

Alan Lightman is a physicist, writer and adjunct professor of humanities at the Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. His forthcoming book, the narrative poem *Song of Two Worlds*, will be published in October 2009.

PENGUIN

NEWS & VIEWS

AGEING

A midlife longevity drug?

Matt Kaeberlein and Brian K. Kennedy

The small molecule rapamycin, already approved for clinical use for various human disorders, has been found to significantly increase lifespan in mice. Is this a step towards an anti-ageing drug for people?

Anti-ageing drugs — compounds that slow the hands of time and allow humans to live far beyond their natural span — have long been fertile ground for science-fiction writers. More recently, however, the possibility that such compounds might exist, and might perhaps even be within reach, has gained scientific credibility. In this issue (page 392), Harrison *et al.*[1] provide evidence that pharmacological intervention in the ageing process is feasible in mammals*. They report that dietary supplementation with rapamycin — a compound known to be linked to lifespan in invertebrates — significantly increases the lifespan of mice.

The US National Institute on Aging's Interventions Testing Program (ITP) was designed to test compounds of interest for effects on ageing in mice[1]. Anyone from the scientific community can nominate a compound for consideration by the ITP, and selected compounds are tested in parallel longevity studies at laboratories at three sites, providing built-in triplicate replication and high statistical power. Several compounds have already been tested. Of these, rapamycin is the first to robustly increase lifespan across all three centres and in both male and female mice.

As is often the case in science, this study benefited from a fortuitous accident. Early on, the ITP researchers realized that simply adding rapamycin to feed failed to maintain high levels of the drug, so a specially formulated feed was developed in which rapamycin is encapsulated for timed release in the intestine. It took more than a year to develop the special feed, which meant that mice in the first cohort to receive rapamycin were 600 days old when supplementation was initiated. As Harrison *et al.*[1] note, this is "roughly the equivalent of a 60-year-old person". Amazingly, both the median and maximum lifespan of these middle-aged mice were significantly increased by rapamycin supplementation. For instance, rapamycin increased maximum lifespan (defined by the 90th survival percentile) from 1,094 days to 1,245 days for female mice and from 1,078 days to 1,179 days for male mice. This translates into a striking increase in life expectancy at the time rapamycin supplementation was initiated of about 38% for females and 28% for males. In an ongoing study, mice are being fed rapamycin beginning at 270 days, with a significant increase in survival also being apparent in this cohort.

*This article and the paper concerned[1] were published online on 8 July 2009.

H. KANUS/SPL

Figure 1 | Rapa Nui and rapamycin. Rapamycin, the compound shown by Harrison *et al.*[1] to prolong lifespan in mice, was first identified in soil samples taken on Easter Island, the Polynesian island famed for its rock-carved figures. Rapamycin compounds are used clinically as immunosuppressants for organ transplants, as a treatment for advanced kidney cancer, and to prevent narrowing of coronary arteries after angioplasty. Easter Island is also known as Rapa Nui, from where rapamycin gets its name.

Rapamycin was first identified as a natural product of the bacterium *Streptomyces hygroscopicus* in soil samples from Easter Island — famous for its impressive rock-carved human statues (Fig. 1). The compound was selected for inclusion in the ITP on the basis of its known property as an inhibitor of the kinase enzyme, target of rapamycin (TOR). TOR signalling has previously been linked to the ageing process in invertebrates[2–6], but until now it had remained an open question as to whether TOR signalling also has a central role in mammalian ageing. The findings of Harrison *et al.*[1] make TOR the first protein that has been shown to modulate lifespan in each of the four model organisms most commonly used to study ageing: yeast, worms, flies and mice[7].

How does TOR activity influence ageing? Among other functions, TOR promotes translation of messenger RNA into protein by the ribosome, and inhibits a pathway that degrades cellular products in lysosomal vesicles (autophagy) — both of which have been implicated in ageing in invertebrate species[7]. Regulation of mRNA translation by TOR, in particular, has emerged as a lifespan-determining pathway that is highly conserved between yeast and the nematode worm *Caenorhabditis elegans*. In both species, mutations in targets of TOR, such as ribosomal S6 kinase (an enzyme involved in protein translation), several translation-initiation factors and multiple ribosomal proteins, increase lifespan[8]. In addition, TOR influences cell growth, cell-cycle progression, mitochondrial metabolism and insulin-like signalling. Untangling the relative contributions of each of these processes to the lifespan extension in mice conferred by rapamycin is likely to stimulate much interest during the next few years.

TOR signalling has also received attention for its role as a possible mediator of dietary restriction. This is defined as a reduction in nutrient availability without malnutrition, and has long been known to increase lifespan in species ranging from yeast to rodents[7]. TOR activity is reduced by dietary restriction, and genetic studies in invertebrate models have linked the inhibition of TOR to increased longevity by dietary restriction[7]. For example, a recent study in yeast showed that TOR inhibition increases

the amounts of a nutritionally responsive transcriptional activator Gcn4, and demonstrated that this is required for full lifespan extension from dietary restriction[9]. Similarly, autophagy must be induced for lifespan to be extended by dietary restriction in *C. elegans*[10].

On the basis of these studies, it is tempting to speculate that rapamycin may be functioning as a dietary-restriction mimetic — a small molecule that provides the benefits of dietary restriction without requiring a reduction in food intake. Like dietary restriction, TOR inhibition not only increases lifespan, but also confers protection in invertebrate and rodent models against age-associated disorders, including cardiovascular dysfunction, diet-induced obesity and cancer[7]. Cancer inhibition in particular is a hallmark of dietary restriction in rodents, and rapamycin analogues are already used clinically as a treatment for certain forms of cancer.

Despite these links, Harrison *et al.*[1] do not strongly favour the idea that rapamycin is mimicking dietary restriction in mice. This is based on their data that rapamycin extends lifespan without reducing body weight, and when treatment is initiated during middle age (late-life onset of dietary restriction has shown inconsistent effects on longevity in previous studies). It is worth pointing out, however, that a true dietary-restriction mimetic may not reduce body weight if it mimics the signalling events (and downstream responses) associated with dietary restriction without changing food consumption. Also, dietary restriction has not yet been extensively characterized in mice of the genetically diverse background used by Harrison *et al.*, so it is difficult to predict whether dietary restriction in these animals would have effects similar to rapamycin. Thus, although it is premature to say for certain that rapamycin is functioning as a dietary-restriction mimetic in mice, the known role of TOR in the nutrient response, and the genetic relationship between TOR signalling and dietary restriction in invertebrates, make this a reasonable possibility.

Is this the first step towards an anti-ageing drug for people? Certainly, healthy individuals should not consider taking rapamycin to slow ageing — the potential immunosuppressive effects of this compound alone are sufficient to caution against this. On the basis of animal models, however, it is interesting to consider that rapamycin — or more sophisticated strategies to inhibit TOR signalling — might prove useful in combating many age-associated disorders. Also, as relevant downstream targets of TOR are better characterized, it may be possible to develop pharmacological strategies that provide the health and longevity benefits without unwanted side effects. So, although extending human lifespan with a pill remains the purview of science-fiction writers for now, the results of Harrison *et al.*[1] provide a reason for optimism that, even during middle age, there's still time to change the road you're on. ■

Matt Kaeberlein and Brian K. Kennedy are in the Departments of Pathology and Biochemistry, University of Washington, Seattle, Washington 98195, USA.
e-mail: kaeber@u.washington.edu

1. Harrison, D. E. *et al. Nature* **460,** 392–395 (2009).
2. Jia, K., Chen, D. & Riddle, D. L. *Development* **131,** 3897–3906 (2004).
3. Kaeberlein, M. *et al. Science* **310,** 1193–1196 (2005).
4. Kapahi, P. *et al. Curr. Biol.* **14,** 885–890 (2004).
5. Powers, R. W. 3rd, Kaeberlein, M., Caldwell, S. D., Kennedy, B. K. & Fields, S. *Genes Dev.* **20,** 174–184 (2006).
6. Vellai, T. *et al. Nature* **426,** 620 (2003).
7. Stanfel, M. N., Shamieh, L. S., Kaeberlein, M. & Kennedy, B. K. *Biochim. Biophys. Acta* advance online publication doi:10.1016/j.bbagen.2009.06.007 (2009).
8. Smith, E. D. *et al. Genome Res.* **18,** 564–570 (2008).
9. Steffen, K. K. *et al. Cell* **133,** 292–302 (2008).
10. Hansen, M. *et al. PLoS Genet.* **4,** e24 (2008).

Competing financial interests: declared (see online article for details).

ATMOSPHERIC PHYSICS

Cosmic rays, clouds and climate

Ken Carslaw

Galactic cosmic rays could influence Earth's cloudiness by creating aerosol particles that prompt cloud formation. That possible effect looks to be smaller than thought, but the story won't end there.

Striking correlations have been observed between Earth's cloud cover and the flux of galactic cosmic rays entering our atmosphere. The decrease in galactic cosmic ray (GCR) flux by about 15% over much of the twentieth century has led to the hypothesis that GCRs could influence climate through their effect on cloudiness. This controversial possibility is revisited in a paper in *Geophysical Research Letters* by Pierce and Adams[1].

There are several plausible mechanisms that could link GCR flux and cloud properties[2]. A leading candidate is the 'ion–aerosol clear-air mechanism', in which atmospheric ions created by GCRs act as nuclei for the formation of atmospheric particles. The nucleation of new nanometre-sized aerosol particles is observed frequently, and in many parts of the atmosphere, and is thought to be a major source of cloud-condensation nuclei (CCN) — particles large enough for cloud droplets to form around them. The link between GCRs and climate is therefore plausible because any change in GCR-ionization rate might be expected to drive changes in cloud-droplet concentrations, and hence the amount of solar radiation that clouds reflect back to space.

Atmospheric ions can indeed seed new particles[3], but two outstanding questions have hampered progress. What fraction of nuclei is created this way? And what fraction of these particles grows large enough to influence CCN? To be relevant to recent climate change, it would be necessary to show that the decrease in GCR flux during the twentieth century could lead to significant changes in CCN and clouds.

In their paper[1], Pierce and Adams estimate the magnitude of the ion–aerosol clear-air mechanism. They used a global atmospheric model with a detailed treatment of aerosol physics to estimate some limiting values of CCN formation from changes in GCR flux. Their conclusion is clear: CCN concentrations just aren't very sensitive to the changes in GCRs that have occurred during the twentieth century. The authors predict that CCN concentrations will change by less than 0.1% between solar maxima and minima as GCRs change by 15% — about the same as the change seen during the last century. They estimate that this change in CCN translates into a change of 0.005 watts per square metre in solar radiation reflected from clouds, insignificant compared with the greenhouse-gas warming of 2 watts per square metre or more over roughly the same period.

Pierce and Adams's model is quite sophisticated in the way it treats the global lifecycle of aerosols, from formation at nanometre sizes to their eventual growth over days to weeks to CCN sizes. But rather than trying to model the complex ion–aerosol processes in detail (physics that is still incompletely understood), they make an upper-limit assumption that all nucleation is due to ions, thereby circumventing one obstacle to making such a global assessment.

Is this negative result the last word on the ion–aerosol clear-air mechanism? Climate modellers are always quick to point out that predictions can be model-dependent. Certainly CCN may be more sensitive to the ion-induced nucleation rate in a different model or under conditions not explored by Pierce and Adams. But other global-model studies[4,5] of nucleation suggest that CCN are fairly insensitive to the nucleation rate for a simple reason: during the time taken for nuclei to grow to CCN sizes, coagulation depletes particle concentrations — just as raindrops are always fewer in number than cloud drops. Unless there is some as-yet-undiscovered process that accelerates the growth of a few charged nuclei all the way up to CCN sizes, this low sensitivity is likely to be a robust conclusion.

Despite this result[1], it is likely that a cosmic-ray–cloud–climate connection will continue to be explored, for two reasons. First, scientists continue to be intrigued by correlations between cosmic rays, Earth's electrical state and climate variables (clouds, precipitation, drought and so on) on timescales from hours to millennia[6,7]. Because the climate displays a


C. DAVE/SPL

Light on stratocumulus: GCR–cloud physics will remain a hot topic.

multitude of cycles on almost all timescales, detection of a correlation among climate variables usually meets with initial and healthy scepticism. But variations in cloud properties observed on timescales that are unique to GCRs[8] will always prompt a hunt for a plausible mechanism.

The second reason that GCR–cloud physics will remain a hot topic is that we have yet to explore all the possible mechanisms. Attention may now shift to the 'ion–aerosol near-cloud' mechanism[2]. GCR ionization modulates the fair-weather conduction current (about 2 picoamps per square metre) flowing between the ionosphere and Earth, thereby altering the charge that has been observed to accumulate around cloud layers. Just like static electricity, this charge can influence how cloud drops stick to aerosol particles. If the particles are effective nuclei for ice formation, then GCRs may influence cloud glaciation and precipitation. And the charge on some aerosol particles in the near-cloud environment could possibly become large enough to influence the formation of cloud drops directly[9]. But our understanding of the relevant physics is incomplete, and it will be some time before global-impact investigations along the lines of Pierce and Adams's study can be made.

Some would argue that the link between cosmic rays and climate is just too tenuous to be worth pursuing. Others would point out that, by ignoring the fact that the atmosphere is actually a dilute plasma (that is, is weakly ionized), we are missing some potentially important cloud physics — and clouds are a very large lever by which to influence climate. Despite the controversy, it is clear that the study of cosmic rays in our climate system has come of age. Sophisticated models of ion–aerosol processes now exist. They are supported by observations and laboratory studies, which will include the upcoming CLOUD experiment at the CERN laboratory near Geneva, Switzerland, in which a proton beam will generate highly controllable ionization events in an aerosol–cloud chamber[10]. ■

Ken Carslaw is in the School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK.
e-mail: k.s.carslaw@leeds.ac.uk

1. Pierce, J. R. & Adams, P. J. *Geophys. Res. Lett.* doi:10.1029/2009GL037946 (2009).
2. Carslaw, K. S., Harrison, R. G. & Kirkby, J. *Science* **298,** 1732–1737 (2002).
3. Curtius, J., Lovejoy, E. R. & Froyd, K. D. *Space Sci. Rev.* **125,** 159–167 (2006).
4. Spracklen, D. V. *et al. Geophys. Res. Lett.* doi:10.1029/2007GL033038 (2008).
5. Pierce, J. R. & Adams, P. J. *Atmos. Chem. Phys.* **7,** 1367–1379 (2007).
6. Harrison, R. G. & Ambaum, M. H. P. *Env. Res. Lett.* **4,** 014003 (2009).
7. Kirkby, J. *Surv. Geophys.* **28,** 333–375 (2007).
8. Harrison, R. G. *Proc. R. Soc. A* **464,** 2575–2590 (2008).
9. Harrison, R. G. & Ambaum, M. H. P. *Proc. R. Soc. A* **464,** 2561–2573 (2008).
10. http://www.cloud-itn.uni-frankfurt.de/

GALAXY FORMATION

Anatomy of elliptical galaxies

Luca Ciotti

The family of elliptical galaxies is remarkable for the structural regularity of its members. Inspecting irregularities in this regularity could help in understanding how these galaxies form.

One of the most-debated subjects in modern astrophysics is how elliptical galaxies, which are among the oldest known objects in the Universe, formed. Among the various likely formation mechanisms, merging is the most popular. According to this theory, different galaxies are the aftermath of merger events between progenitors of different morphologies and of varying encounter geometries. But observations indicate that there is room for other mechanisms. Despite great endeavour in trying to match the regularities observed in the structures of elliptical galaxies with theoretical models, there is still no consensus view of how they formed. Writing in *The Astrophysical Journal Supplement Series*, Kormendy and colleagues[1] report a meticulous study of all known elliptical galaxies in the Virgo cluster (one of the clusters of galaxies nearest to Earth) that investigates how departures from the observed regularities can be diagnostic of the processes that triggered the formation of these galaxies (Fig. 1, overleaf).

The most striking property of elliptical galaxies is that their brightness profiles — that is, the way in which the combined luminosity of their stars varies with distance from the centre — depend in a regular way on their total luminosity (Sérsic's law). Other properties of elliptical galaxies that correlate with their total brightness include size, mean star velocity and metal content. Another trait shared by these stellar systems is a supermassive black hole, with a mass of the order of one-thousandth of the galaxy's stellar mass, at their centre[2].

In their study of the Virgo cluster of galaxies, Kormendy *et al.*[1] report galaxy luminosity profiles over large radial ranges and argue that elliptical galaxies are not arranged as a continuous sequence of objects with properties that scale well with their total luminosity. Instead, elliptical galaxies seem to branch out into two families according to a threshold value for the total luminosity. This dichotomy manifests itself in two kinds of departure from the Sérsic law at small radii. Luminous ellipticals have 'cuspy' cores — that is, their luminosity profiles are characterized by 'missing light' at small radii, because their brightness at such radii drops below the Sérsic-fitted, larger-radii profile. By contrast, less-luminous ellipticals are all 'coreless' — their central luminosity profiles seem to have 'extra light' at small radii (but see Graham *et al.*[3] for a different interpretation of the central-light profiles).

Kormendy and colleagues' results add weight to other observations that have hinted at a dichotomy in the properties of elliptical galaxies. Luminous-core galaxies are known to be slowly rotating; to be relatively anisotropic (properties such as stellar velocities depend on direction); to have triaxial shapes (they have different diameters in all three directions); to have quite 'steep' Sérsic profiles; and to have stars that are mostly very old and that formed on comparatively short timescales. Conversely, low-luminosity coreless ellipticals rotate rapidly; are more isotropic; have mostly oblate-spheroidal shapes; have quite


NASA/ESA/HUBBLE HERITAGE TEAM (STSCI/AURA)

Figure 1 | The deceptive looks of elliptical galaxies. Elliptically shaped galaxies, such as the larger system ESO 325-G004 (top left) seen in this Hubble Space Telescope image of the Abell cluster of galaxies S0740, appear to be quite simple objects. Kormendy and colleagues[1] show that these apparently uncomplicated galaxies disclose unexpected intricacies that could shed light on the mechanisms of their formation.

'shallow' Sérsic profiles; and have stellar populations that are younger and that formed on longer timescales. It is also well known that core ellipticals contain hot X-ray-emitting gas and tend to be radio-loud (they have strong radio emission), whereas the coreless galaxies do not[4].

The global regularities of the family of elliptical galaxies have already been used to set important constraints on their formation mechanisms[5,6]. Kormendy and colleagues' analysis now demonstrates that studies examining departures from those regularities could provide even more compelling clues. One previously recognized possibility is that the cores of luminous, large ellipticals could be naturally scoured by binary black holes[7], one likely product of the merging of two smaller elliptical galaxies, each of which harbours a central black hole. But binary black holes should also be found in coreless elliptical galaxies if merging is the main formation mechanism of the whole family of elliptical galaxies. So what thwarted core scouring by binary black holes in coreless ellipticals?

Kormendy *et al.*[1] suggest that if enough cold gas (provided by the merging companion) is carried to the centre of the forming coreless galaxy and turned into stars that provide the extra light, it could swamp core scouring by binary black holes. The picture that emerges from the new study[1], therefore, is one in which luminous-core ellipticals are the products of mergers between progenitors that had large numbers of old stars but little or no cold gas ('dry' mergers), whereas less-luminous coreless galaxies are the aftermath of mergers between (cold-)gas-rich galaxies ('wet' mergers). The wet mergers would trigger starburst activity; any energy released by the final black hole (the product of merging between the two progenitor black holes) because of gas accretion would not be sufficient to heat the gas and quench star formation.

Although the hypothesis discussed by Kormendy *et al.* is appealing, it is not the final word on elliptical-galaxy formation. Major evolutionary mechanisms are thought to be at work within elliptical galaxies. For example, in the several billion years of their evolution, the dying stars of an elliptical galaxy release an amount of gas equal to about 20–30% of the initial stellar mass of the galaxy. However, the observed present-day black-hole-to-galaxy mass ratio empirically dictates that, at most, 1% of this gas can be accreted by the central black hole. Therefore, essentially all of this gas is available to be recycled by the galaxy in some way other than black-hole accretion. If only a minor fraction of this gas flowed to the centre of the galaxy and was transformed into stars, it could suffice to produce the central extra light in low-luminosity ellipticals[8], whereas galaxy merging could be solely responsible for the creation of cores in high-luminosity ellipticals.

Increasingly refined computational simulations[9,10] are being devised to untangle the relative contributions of the different physical processes involved in galaxy formation. For now, Kormendy and colleagues' results demonstrate that research on elliptical galaxies, which were once considered rather dull objects, is revealing surprising secrets about galaxy formation. ■

Luca Ciotti is in the Department of Astronomy, University of Bologna, 40127 Bologna, Italy.
e-mail: luca.ciotti@unibo.it

1. Kormendy, J., Fisher, D. B., Cornell, M. E. & Bender, R. *Astrophys. J. Suppl. Ser.* **182,** 216–309 (2009).
2. Magorrian, J. *et al. Astron. J.* **115,** 2285–2305 (1998).
3. Graham, A. W., Erwin, P., Trujillo, I. & Asensio Ramos, A. *Astron. J.* **125,** 2951–2963 (2003).
4. Pellegrini, S. *Mon. Not. R. Astron. Soc.* **364,** 169–178 (2005).
5. Toomre, A. in *The Evolution of Galaxies and Stellar Populations* (eds Tinsley, B. M. & Larson, R. B.) 401–426 (New Haven Yale Univ. Observatory, 1977).
6. Ostriker, J. P. *Comments Astrophys.* **8,** 177–180 (1980).
7. Merritt, D. *Rep. Prog. Phys.* **69,** 2513–2579 (2006).
8. Ciotti, L. & Ostriker, J. P. *Astrophys. J.* **665,** 1038–1056 (2007).
9. Hopkins, P. F., Lauer, T. R., Cox, T. J., Hernquist, L. & Kormendy, J. *Astrophys. J. Suppl. Ser.* **181,** 486–532 (2009).
10. Naab, T., Johansson, P. H., Ostriker, J. P. & Efstathiou, G. *Astrophys. J.* **658,** 710–720 (2007).

ECOLOGY

Towards a theory of biodiversity

Jayanth R. Banavar and Amos Maritan

Models of ecological communities that incorporate mutation and spatial dispersal can yield results that go some way to explaining observations. A further step is to add sexual reproduction to the mix.

Humans are wreaking havoc on the planet[1,2], and the result is a continuing mass extinction of species. Yet there is little understanding of the various ecological effects of this loss of biodiversity. An initial requirement is to work out how an ecological community is assembled in the first place, and on page 384 of this issue de Aguiar *et al.*[3] offer an advance in tackling this formidable task by extending an explanatory approach to biodiversity known as neutral theory.

First, some context. Life is a non-equilibrium phenomenon involving a 'dual-diffusion' process. Organisms diffuse and distribute themselves in space, and compete for resources, such as light, water and nutrients. At the same time, on different timescales, there is diffusion in 'genome space' that leads to new species and serves to maintain diversity. The evolution of an ecological community corresponds to the dynamics of the distribution of occupied regions in real space and genome space, and the development of interactions between the two. Owing to the constraints of the available resources, growth in the population of one species must necessarily lead to a decrease in the population of another and, in some cases, to extinction. Niche effects[4,5], which arise from the competition for resources in a spatially heterogeneous environment, provide a simple

rationalization for species coexistence, rarity and extinction. They can, to some extent, be used to predict the observed distribution of species and their traits along nutrient gradients within the environment.

The environment is itself shaped by the individuals that make up the community, leading to a dynamic feedback mechanism. Thus, an ecological community is complex, with myriad interactions among and between species, as well as interactions of species with a spatially inhomogeneous and temporally changing environment. It is a many-body system exhibiting emergent characteristics[6]. To compound matters, many of the interactions are unknown or, at best, imperfectly known.

A fundamental understanding of biodiversity requires a theoretical framework based on a restricted set of verifiable hypotheses. An example of such an approach is neutral theory — presented in 2001 by Hubbell[7] — which predicts patterns of organisms that agree well with observational data. Hubbell's simplifying idea was to treat all species at the same trophic level (position in a food chain) as equivalent to each other. The rarity of one species and the high abundance of another do not result from their distinctive traits, but rather from stochastic drift in populations of essentially identical species. Species diversity comes about not from differences among species, as highlighted in niche theory, but rather from the interplay between extinction caused by random drift in species populations and the introduction of new species by either immigration from a surrounding region or speciation.

In the new paper, de Aguiar *et al.*[3] present computer simulations of a specific neutral model of dual diffusion along with the process of sexual reproduction. Previous approaches took into account diffusion in real space and postulated the appearance of new species through mutation without explicit consideration of the evolution of the genetic codes of the breeding organisms. In the simulations, de Aguiar *et al.* observe the spontaneous emergence of species, a group of individuals with genetic codes occupying a sub-region of the whole genome space, living in a sub-region of the space available to the community. This dual segregation arises from the double constraints on the difference in the genetic codes as well as the spatial distance between prospective mates, and does not require the presence of a geographical barrier. The average distance between a randomly chosen pair of points in a community of fixed size is higher in lower spatial dimensionality; consequently, segregation is easier in lower dimensions and leads to greater species diversity.

Remarkably, the results of the simulations, with many adjustable parameters, are in good accord with a variety of observed patterns. Examples of such patterns are the constant rate of speciation observed in the fossil record; the higher diversity of freshwater ray-finned fishes than of their marine counterparts; the species–area relationships of birds, flowering plants and tropical-forest trees; and the relative species abundance of birds and forest trees.

A framework for understanding biodiversity requires identification of the essential ingredients and an understanding of their effects. Along with other colleagues, we have used simplified models with no explicit interactions between species or with the environment to derive analytical expressions for relative species abundance[8], the spatial distribution of individuals[9], and the dynamics of an ecological community[10], and these fit the empirical data moderately well with relatively few adjustable parameters. A lesson from these calculations is that just because a theory fits the data, it does not necessarily imply that the assumptions underlying the theory are correct. In fact, one can sometimes fit data such as relative species abundance equally well through distinct hypotheses showing that additional data, not always measured, are needed to discriminate between competing biological possibilities. The proper definition of interactions, the means to infer them from data, and their incorporation into theory leading to testable predictions remain a huge challenge.

Among other lines of investigation, what is needed next is work along two fronts: the development of analytically tractable models that include the innovations introduced by de Aguiar *et al.*, and careful analysis of field data to assess whether the underlying assumptions of the simulations are valid. The journey towards understanding biodiversity will be a long one, but it is encouraging that we are at least taking steps down that road. ■

Jayanth R. Banavar is in the Department of Physics, The Pennsylvania State University, Pennsylvania 16802, USA. Amos Maritan is in the Department of Physics, University of Padova, Padova 35100, Italy.
e-mails: banavar@psu.edu; maritan@pd.infn.it

1. Walther, G.-R. *et al. Nature* **416,** 389–395 (2002).
2. Raven, P. H., Berg, L. R. & Hassenzahl, D. M. *Environment* 6th edn (Wiley, 2008).
3. de Aguiar, M. A. M., Baranger, M., Baptestini, E. M., Kaufman, L. & Bar-Yam, Y. *Nature* **460,** 384–387 (2009).
4. Sugihara, G., Bersier, L.-F., Southwood, T. R. E., Pimm, S. L. & May, R. M. *Proc. Natl Acad. Sci. USA* **100,** 5246–5251 (2003).
5. Tilman, D. *Proc. Natl Acad. Sci. USA* **101,** 10854–10861 (2004).
6. Anderson, P. W. *Science* **177,** 393–396 (1972).
7. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).
8. Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. *Nature* **450,** 45–49 (2007).
9. Zillio, T., Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. *Phys. Rev. Lett.* **95,** 098101 (2005).
10. Azaele, S., Pigolotti, S., Banavar, J. R. & Maritan, A. *Nature* **444,** 926–928 (2006).

CLIMATE CHANGE

Beyond the CO_2 connection

Rainer Zahn

At times in the past, mobile ocean fronts in the subtropics have exercised an influence on the magnitude of climate change by decoupling temperature from levels of carbon dioxide in the atmosphere.

During the past 800,000 years — the late Pleistocene and subsequent Holocene — Earth's climate has swung between cool and warm and back again many times. Variations in incoming solar radiation, modulated by Earth's orbital parameters[1], along with recurring ups and downs in atmospheric carbon dioxide concentrations[2], are among the prominent features associated with these climate swings. At first glance, then, it looks like climate is intimately coupled with solar radiation and CO_2. But on closer inspection, differences become apparent between the amplitudes of climate changes, and CO_2 variations in particular, that raise the question of how tight that coupling is.

As they describe on page 380 of this issue[3], Bard and Rickaby have addressed this question. They analyse palaeoclimatic records from a marine sediment core in the southwestern Indian Ocean, and show that the ocean front separating the warm subtropical ocean from the cold subantarctic zone may influence global-scale temperature control. Bard and Rickaby use a suite of data records to monitor latitudinal migrations of the subtropical front as climate cycled between ice ages and warm ages during the past 800,000 years. What makes their study particularly persuasive is that the sediment core from which they draw their data comes from the entrance to the ocean gateway that connects the Indian Ocean with the Atlantic Ocean at the southern tip of Africa.

At this gateway, warm and salty waters leak from the Indian Ocean into the South Atlantic, forming part of the global ocean thermohaline circulation and compensating for the export of deep water from the Atlantic to the rest of the world ocean. The salt-water transport to the South Atlantic causes a south-to-north density gradient in the Atlantic as a whole that has the power to influence the Atlantic's meridional overturning circulation and, ultimately, the Gulf Stream, with consequences for heat transport and climate across the North Atlantic region[4–6]. In theory, the latitudinal position of the subtropical front defines how wide the gate is left open for the salt-water transport from the Indian Ocean to the Atlantic.

Ocean fronts are boundaries between water masses of different temperature and salt content,

NATURE|Vol 460|16 July 2009

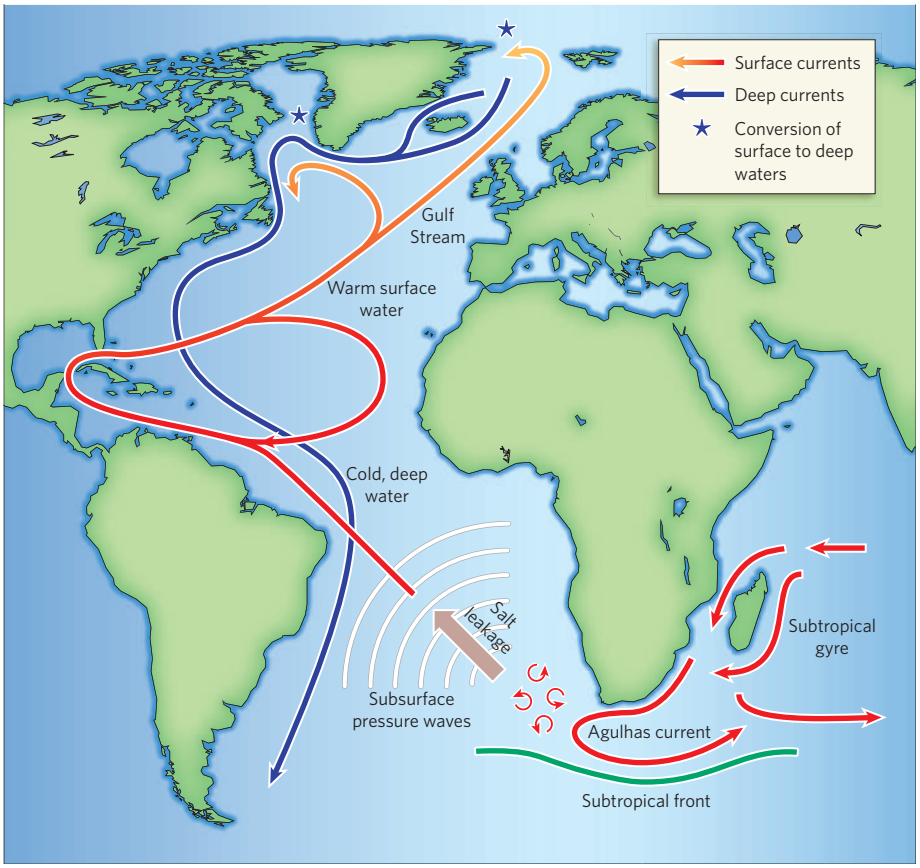


Figure 1 | The Agulhas current and the subtropical front. The Agulhas is the largest ocean jet current in the Southern Hemisphere, and carries warm, salty waters from the South Indian Ocean subtropical gyre along the eastern edge of southern Africa. Most of the current feeds back into the subtropical gyre, but some of its waters flow through the ocean gateway off the southern tip of Africa and into the Atlantic. This 'leakage' alters the density structure of the South Atlantic, with northward propagation of subsurface pressure waves potentially affecting the meridional overturning circulation in the North Atlantic[4–6], which has global effects. Migration of the mobile subtropical front, as described in the work of Bard and Rickaby[3], is diagnostic of shifted wind patterns that can alter the Agulhas Current and the southern African ocean gateway and so, potentially, the choreography of ocean–climate interactions.

and so of different density. That they affect ocean circulation and climate is not new. In particular, it has long been recognized[7] that, during the late Pleistocene, the North Atlantic polar front exerted control over the Atlantic overturning circulation and thus over climate in the Northern Hemisphere. Gains and losses in seawater density near the centres of deep-water formation were tightly linked with movements of the polar front. Shifting the front to a northerly position during the warm periods between ice ages allowed warm subtropical waters to penetrate far into cold subpolar latitudes, invigorating evaporation that increased the salt concentration and hence the density of sea water, thereby stimulating deep convection. A southerly position of the polar front during ice ages tended to reduce the thermal contrast between surface water and the overlying atmosphere, reducing evaporation and saltiness and causing a less-vigorous deep convection. Although it seems simplistic, the validity of this concept was demonstrated in the early days of palaeoceanography[8], and since then the focus has been fixed on the high latitudes.

Bard and Rickaby[3] shift attention to the subtropics by postulating that the subtropical front has likewise acted as an agent of change. Two 'superglacial' intervals serve as test cases — marine isotope stages 10 and 12 — during which Earth experienced unusually cold climates despite levels of atmospheric CO_2 that were no different from those during the preceding and subsequent ice ages. The suite of temperature-sensitive geochemical and faunal records that Bard and Rickaby present constitute a compelling case that, during both superglacials, the subtropical front was shifted by as much as 7 degrees latitude to its northernmost position of the past 800,000 years.

In this position, the subtropical front potentially would have prevented water transport to the Indian–Atlantic gateway by intercepting the Agulhas current — which runs south along the eastern edge of southern Africa — closing the gate on the inter-ocean transport of water (Fig. 1). In consequence, the Atlantic overturning circulation could have been forced into an unprecedentedly slow mode. This then reduced the poleward oceanic transport of heat, so the thinking goes[4–6], to the extent that ice sheets in the Northern Hemisphere grew beyond their normal limits and climate cycled into a severe cool state.

So much for the concept. But what might have caused the subtropical front to undergo such an exceptional northward migration? The position of the front is a consequence of the processes that alter the temperature distribution at the southern reaches of the large oceanic subtropical gyres. Today, a central element is the interplay between the easterly trade winds in the subtropics and the westerly winds that, in turn, interact with the ocean to move water (and heat) around its surface. The dynamics that would have caused the wind patterns to change, such that the subtropical front may have shifted that far north during marine isotope stages 10 and 12, are not yet clear.

It also remains to be seen if a northward-migrating subtropical front would, by itself, have been strong enough to block a current as mighty as the Agulhas and so affect water transport to the Atlantic. A different way of altering the scale of the leakage is for the Agulhas current to shift gears. The strengths of currents typically vary in parallel with shifting wind patterns and could act to invigorate or weaken the Agulhas leakage to the Atlantic. The details of this mechanism have not yet been fully explored[9,10]. But the strength of the Agulhas has been steadily increasing since the 1980s, driven by changed gradients in atmospheric pressure and in wind fields[11].

Bard and Rickaby[3] make a compelling case that, at times in the past, severely reduced water transport between the Indian and Atlantic oceans may have caused climate to cool beyond typical ice-age conditions. But will the reverse also hold? Will an increased leakage (forced by recent shifts in atmospheric pressure and winds[11]) compensate for the loss of salt in the North Atlantic[12,13] (caused by increased precipitation and changes in the freshwater flux from the Arctic)? Would this be enough to stabilize the Atlantic overturning circulation? Let's keep an eye on what the leakage does next. ■

Rainer Zahn is at the Institució Catalana de Recerca i Estudis Avançats (ICREA), Institut de Ciència i Tecnologia Ambientals and Departmento de Geologia, Universitat Autònoma de Barcelona, Bellaterra E-08193, Spain.
e-mail: rainer.zahn@uab.cat

1. Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194,** 1121–1132 (1976).
2. Siegenthaler, U. *et al. Science* **310,** 1313–1317 (2005).
3. Bard, E. & Rickaby, R. E. M. *Nature* **460,** 380–383 (2009).
4. Biastoch, A., Böning, C. W. & Lutjeharms, J. R. E. *Nature* **456,** 489–492 (2008).
5. Knorr, G. & Lohmann, G. *Nature* **424,** 532–536 (2003).
6. Weijer, W., de Ruijter, W. P. M., Sterl, A. & Drijfhout, S. S. *Glob. Planet. Change* **34,** 293–311 (2002).
7. Duplessy, J. C. *et al. Oceanologica Acta* **14,** 311–324 (1991).
8. Mix, A. C. & Fairbanks, R. G. *Earth Planet. Sci. Lett.* **73,** 231–243 (1985).
9. Franzese, A. M., Hemming, S. R. & Goldstein, S. L. *Paleoceanography* doi:10.1029/2008PA001706 (2009).
10. van Sebille, E., Biastoch, A., van Leeuwen, P. J. & de Ruijter, W. P. M. *Geophys. Res. Lett.* doi:10.1029/2008GL036614 (2009).
11. Rouault, M., Penven, P. & Pohl, B. *Geophys. Res. Lett.* doi:10.1029/2009GL037987 (2009).
12. Dickson, B. *et al. Nature* **416,** 832–837 (2002).
13. Dickson, R. R., Curry, R. & Yashayaev, I. *Phil. Trans. R. Soc. Lond. A* **361,** 1917–1934 (2003).

PHOTONICS

Light control at will

Sajeev John

Microchips that make use of light instead of electrons could outperform their electronic counterparts if light flow can be controlled at will. Photonic crystals are instrumental in achieving such a manoeuvre.

Photonic crystals are artificial materials made of periodically arranged insulating structures that are specially designed to aid the trapping and confinement of light within a very small area. These crystals offer the possibility of confining light on sub-wavelength scales in air, and provide a foundation for guiding light through circuit paths in a three-dimensional (3D) optical (as opposed to electronic) microchip without energy loss caused by light scattering. The crystals also enable unprecedented strong coupling effects between light and semiconductor nanocrystals, known as quantum dots. These effects provide a route to fast information processing and computing by optical means. However, such tantalizing opportunities require that photonic crystals be accurately synthesized and that light be efficiently manipulated both inside and at the surface of the crystal without loss. In this issue (page 367), Ishizaki and Noda[1] take an important step towards achieving these goals by demonstrating the control of light at the surface of a suitably constructed, gallium-arsenide-based, 3D photonic crystal.

The special, regularly repeating arrangement of insulating (dielectric) structures in photonic crystals gives them a special property: light waves with frequencies that fall within a range known as the photonic band gap (PBG) are prevented from propagating because of destructive wave-interference effects. This provides a clean slate on which to write optical circuit paths. Ishizaki and Noda's woodpile-type structure consists of layers of dielectric rods stacked together (see Fig. 1a on page 367). Although it exhibits a bulk 3D PBG — that is, frequencies within the PBG are blocked from flowing inside the material in all possible directions — there are frequencies (modes) within the otherwise forbidden gap in which light can leak from the crystal interior to the exterior surface, travel freely along the surface of the crystal and eventually escape into free space. In other words, the edges of the slate are not completely clean.

Modification

To control the leakage of light into unwanted surface modes, the authors[1] reapply the PBG concept to the surface itself. They modify the surface layer of the material by constructing a two-dimensional (2D) periodic structure to create a 2D PBG for the surface modes. The spectral overlap between the bulk 3D PBG and the 2D surface PBG then provides the desired frequency window in which no electromagnetic waves can propagate, either within the woodpile architecture or along its surface. With this completely clean slate, the last remaining escape pathways for light are removed and light within the 3D microchip can be controlled at will.

The experimental demonstration[1] of a simultaneous 3D PBG and 2D surface PBG has several implications. Ishizaki and Noda have used this to demonstrate a strongly localized state on the surface of their woodpile structure that persists for around 9,000 optical cycles before it releases its trapped light to free space. This light trap was engineered by introducing a defect — in the form of a local deviation from horizontal periodicity — on the surface of a thin 3D photonic crystal; that is, one with only two repeating units of vertical periodicity. If thicker 3D crystals were used, the lifetime would be expected to increase dramatically.

However, this is only one small indicator of the implications of the authors' successful integration of defects of different dimensionality within 3D PBG architectures. For example, circuits for light in an optical microchip are based on one-dimensional defects known as linear waveguides[2]. In optical communications and information processing, a crucial issue is the 'insertion loss' as light enters the in-chip waveguide from an external optical fibre. The efficiency of the coupling between the chip and the fibre is limited by light scattering into unwanted surface modes. Unlike their conventional equivalents, photonic-crystal fibres[3] — those carved out of otherwise-perfect 2D PBG crystals — would enable the efficient transfer of optical information from fibre to chip, provided that unwanted surface modes on the 3D chip are eliminated[4]. Ishizaki and Noda's results provide the required clean slate.

The 3D optical microchips of the future will require not only the bulk 3D PBG crystal to provide overall confinement of light within the chip, but also the integration of lower-dimensional defects and band-gap materials. This would completely control the flow of light and tailor its nonlinear interaction with the optical resonators embedded within the photonic crystal. One possible solution is to use PBG heterostructures, which involve sandwiching a layer of photonic crystal with a 2D PBG between two identical 3D PBG materials[5]. In their study[1], Ishizaki and Noda have, in effect, synthesized the open-faced version of the sandwich, a step towards constructing and demonstrating light flow within the fully 3D circuitry of a 3D optical microchip[5].

Optical computing in a photonic microchip requires that tiny pulses of laser light pass through the waveguide circuitry and that these pulses, in turn, control and switch the flow of other optical pulse streams. In free space, these pulse streams would simply pass through each other, transferring no information from one to the other. By coupling these optical pulses to 'artificial atoms' (quantum dots) embedded near or within a photonic-crystal waveguide, it should be possible to control light with light[6]. When the quantum dots are excited or de-excited by one pulse stream, they can coherently amplify or attenuate pulses in the second stream of optical information, enabling information processing.

Confinement

In conventional materials, high-power laser beams are needed to make this nonlinear process occur. However, with less than a milliwatt of laser power flowing through a 3D PBG waveguide, it is possible to deliver electric-field amplitudes in the range of 10,000 –100,000 volts per centimetre as a result of the confinement of light modes on sub-wavelength scales by the 3D PBG material. This tight confinement and concentration of light prompts an otherwise unattainable response from quantum dots at modest levels of laser power. Whereas the 3D PBG is often associated with the inhibition of spontaneous emission of light from a quantum dot, suitably embedded planar and linear waveguides can be used to engineer spectral windows within the PBG in which the rate of spontaneous emission relative to that in free space is increased by several thousand.

This enhancement of the coupling of light to quantum dots, or more generally of the strength of light–matter interactions, will help novel quantum-electrodynamics phenomena — such as 'vacuum Rabi splitting' in single quantum dots[7] and large 'Mollow splitting' in a collection of quantum dots[6] — to emerge. The synthesis of high-quality 3D PBG materials and the control of their surface optical modes as demonstrated by Ishizaki and Noda is progress towards these and other far-reaching goals in both basic and applied science. ■

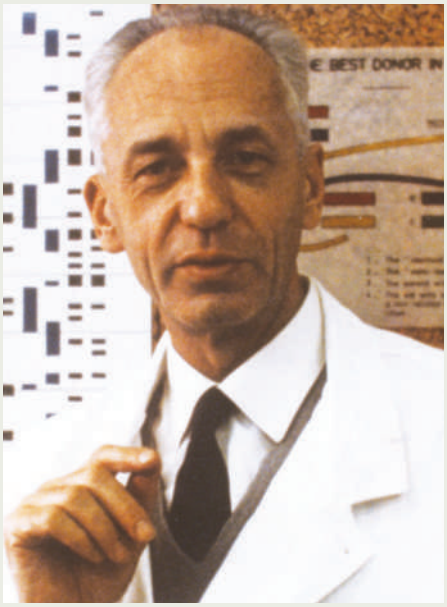
Sajeev John is in the Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada.
e-mail: john@physics.utoronto.ca

1. Ishizaki, K. & Noda, S. *Nature* **460,** 367–370 (2009).
2. Joannopoulos, J. D., Villeneuve, P. R. & Fan, S. *Nature* **386,** 143–149 (1997).
3. Knight, J. C., Broeng, J., Birks, T. A. & Russell, P. St J. *Science* **282,** 1476–1478 (1998).
4. Bauer, J. & John, S. *Phys. Rev. A* **77,** 013819 (2008).
5. Chutinan, A. & John, S. *Optics Express* **14,** 1266–1279 (2006).
6. Vujic, D. & John, S. *Phys. Rev. A* **76,** 063814 (2007).
7. Khitrova, G., Gibbs, H. M., Kira, M., Koch, S. W. & Scherer, A. *Nature Phys.* **2,** 81–90 (2006).

OBITUARY

Jean Dausset (1916–2009)

'Father' of the human leukocyte antigen system.


CEPH

Jean Dausset, who died in Majorca, Spain, on 6 June 2009, aged 92, will always be remembered for the discovery of the human leukocyte antigen (HLA) tissue system, one of the most significant scientific breakthroughs of the twentieth century. The HLA system — a collection of genes that encode polymorphic molecules expressed on the surface of cells — determines the biological identity of an individual's cells and tissues. Because recognition of HLA molecules by the immune system allows it to distinguish self cells from foreign cells, this discovery laid the foundations for successful organ transplantation. Dausset not only pioneered the scientific and medical research into HLA, but also embraced and promoted the sociological and philosophical impacts of this finding.

Dausset was born in Toulouse, France, into a provincial bourgeois family, and spent his youth in Biarritz before the family moved to Paris. He completed his medical education at the Faculty of Medicine at the University of Paris. During the Second World War, he joined the French army transfusion service and participated in the Tunisian and Normandy campaigns. It was in the Army Medical Corps, where he oversaw large numbers of blood transfusions, that his drive to understand the causes of frequent adverse reactions against blood cells was born. As part of the Marshall Plan, he attended Harvard University for a year, in 1948, where he worked as a clinical haematologist at the Peter Bent Brigham Hospital. On his return to Paris, he specialized in immunohaematology.

In 1952, Dausset made the seminal observation that serum from patients who had received multiple blood transfusions caused marked agglutination of white blood cells (leukocytes). Dausset attributed this leukoagglutination to an antibody in the serum directed against the white blood cells. In 1958, he described the first human leukocyte antigen, which he named MAC (now known as HLA-A2), in a single-author publication. He went on to define the genetic region that encodes the complex system of cell-surface antigens, which he termed HU-1, later renamed human leukocyte antigen.

In the 1960s, Dausset teamed up with the surgeon Felix Rapaport in Paris. Together they performed hundreds of skin-graft experiments on volunteers, correlating graft survival with the extent of HLA incompatibility and thereby demonstrating the role of HLAs in human transplantation. He considered these volunteers the 'real heroes' of the discovery of the HLA system and always acknowledged their contribution on receipt of honours.

Today, many years after his discovery, tissue typing (determining the combination of HLAs on cells) is an essential tool for selecting donors for organ and bone-marrow transplantation. Dausset's pioneering work on the HLA system was recognized in 1980 when he was awarded the Nobel Prize in Physiology or Medicine, together with Baruj Benacerraf and George Snell, for "their discoveries concerning genetically determined structures on the cell surface that regulate immunological reactions".

Although Dausset's own contributions were not in the field of immunology, he always insisted that HLAs were important in the control of immune responses as established by Hugh McDevitt, a colleague he particularly admired. They both anticipated that the HLA system contributed to susceptibility or resistance to disease, including autoimmune disease. With extraordinary prescience, Dausset extended this principle to espouse the concept of a new form of medicine that used an individual's genetic profile to predict disease — he named this 'predictive medicine' and foresaw its potential in personalized disease prevention.

Dausset recognized that the HLA system was a tool that could be used to analyse the genetics and migration patterns of populations. His interest in the anthropological implications of his discovery took him on a voyage to Easter Island, where he collected blood samples from this isolated population for genetic analysis.

He also realized that the HLA model could be extended to the entire human genetic system and, in 1984, together with Daniel Cohen, created the Human Polymorphism Study Centre (CEPH) — a resource that facilitates genetic studies of human populations. The genotyping of large numbers of reference families by the CEPH, and the sharing of this information with scientific peers, contributed to the determination of the first genetic linkage map of the human genome, and led the way to the discovery of genes that predispose to numerous diseases.

In addition to his medical research, Dausset worked as a French government adviser from 1955 to 1958, and was instrumental in shaping French law integrating science education and medicine, leading to the creation of the CHU (University Hospitals). In 1969, he founded France Transplant, an organization that coordinated organ transplantation across France. With Jean Bernard, his colleague at the Hopital St Louis, Paris, he established the French Bone Marrow Donors Registry of unrelated donors for bone-marrow transplantation. When these organizations were taken over by national administration, he voiced concern that the humanistic nature of these endeavours would be diluted by bureaucracy. Dausset was eventually elected president of the Universal Movement for Scientific Responsibility (MURS), and actively campaigned against patenting the human genome sequence.

Dausset never had a mentor. The discovery of the HLA system stemmed from his own vision and was initially driven solely by his ferocious determination. By contrast, for more than 50 years as a scientist, he embodied a spirit of international collaboration, gathering around him numerous students, colleagues and friends, whom he encouraged and supported. This extended worldwide community, which had a strong South American and Hispanic flavour, firmly established him as the 'father of HLA'. He enjoyed this parental status and designated the immunologist Rose Payne the 'mother' of HLA.

A passionate collector of modern plastic arts, for a time he owned an art gallery in Saint Germain des Prés, Paris. He poetically summarized his philosophy of life in the title of his autobiography *Clin d'Oeil à la Vie* (*A Wink at Life*). His wife Rosita, daughter Irène and son Henri were a constant support. Dausset died peacefully in Majorca, surrounded by his family and immersed in the Hispanic culture he was so fond of. His uniqueness will endure.

Dominique J. Charron
Dominique J. Charron is at the Laboratoire Jean Dausset, Hôpital Saint-Louis, and the Institut Universitaire d'Hématologie, 75010 Paris, France.
e-mail: dominique.charron@sls.aphp.fr

REVIEWS

Primate archaeology

Michael Haslam[1], Adriana Hernandez-Aguilar[1], Victoria Ling[1], Susana Carvalho[1], Ignacio de la Torre[2], April DeStefano[3], Andrew Du[3], Bruce Hardy[4], Jack Harris[3], Linda Marchant[5], Tetsuro Matsuzawa[6], William McGrew[1], Julio Mercader[7], Rafael Mora[8], Michael Petraglia[1], Hélène Roche[9], Elisabetta Visalberghi[10] & Rebecca Warren[4]

All modern humans use tools to overcome limitations of our anatomy and to make difficult tasks easier. However, if tool use is such an advantage, we may ask why it is not evolved to the same degree in other species. To answer this question, we need to bring a long-term perspective to the material record of other members of our own order, the Primates.

Several animal species use tools and selectively manipulate objects[1]. Primate tool use has received particular attention, in part because of the close evolutionary heritage that other primates share with technology-dependent humans. However, the effects of material culture on primate long-term adaptiveness have yet to be systematically explored. Here we review the interface of primatology and archaeology, following a recent intensification of research into the relevance of other primates for understanding hominin technology and behaviour.

Extensive and flexible tool use was once considered a defining human characteristic[2]. Detailed observation since the mid-twentieth century has, however, revealed a wide variety of habitual tool use among wild chimpanzees (*Pan troglodytes*)[3] in Africa, adding to records of less complex tool use by other species. At the same time, stone artefacts made by hominins—the human lineage since the split with chimpanzees some 5–7 Myr ago[4]—have been found dating back 2.6 Myr (ref. 5). Because this earliest Oldowan technology already shows much planning depth, spatial coordination and manual dexterity[6] in its creators, it is probable that earlier, currently unrecognized, tool manufacture occurred. Appeals to phylogenetic proximity therefore posit the last common ancestor of chimpanzees and humans to have been a tool user[7,8], and for many archaeologists chimpanzees have become the dominant referent for modelling early hominin behaviour. However, recent recognition that wild South American bearded capuchin monkeys (*Cebus libidinosus*) also habitually use tools[9], whereas wild bonobos (*Pan paniscus*, the chimpanzee's closest relative) rarely do[10], forces us to rethink the accepted roles of continuity and convergence in primate tool use. We may ask, for example, how many extinct primate groups independently 'invented' tool use during the past several million years, and what circumstances permitted or prevented such discoveries. Questions also arise as to the influence of long-term tool use on non-human-primate anatomy, and the reasons why hominins alone have taken tool use to such an extreme.

A long-standing separation of anthropocentric archaeology from primate ethology has obscured the holistic perspective required to address these questions. Here, our solution is to introduce a new interdisciplinary field—primate archaeology—that examines the past and present material record of all members of the order Primates. This field provides a comprehensive comparative and long-term evolutionary framework for understanding the biological, environmental and social contexts of primate behaviour, through analyses of tool making, tool use and the spatially patterned accumulation of refuse. Examination of the contexts for non-human-primate artefact and landscape use, alongside early hominin equivalents, provides a new understanding of the origins and evolution of human behaviour. Comparisons with the patterned use of material objects by living and extinct taxa outside the Primates will provide the first step towards a universal framework for unravelling the behavioural implications of both human and non-human components of the archaeological record.

The intersection of primatology and archaeology

Since the first systematic recording of wild-chimpanzee tool use, primatologists have stressed its direct relevance to early hominin studies[11]. Beyond phylogeny, the main reasons for this applicability are the convenient physicality and durability of material culture, which permit measurement of behaviour and inference of intentions and abilities even if the user is absent or dead. From an archaeological perspective, definitions of 'tool use'[12] are therefore less important than recognition of the adaptive benefits and interpretive potential of manipulated durable objects.

Free-living primate populations use a wide variety of plant materials for extractive foraging, social interaction and self-maintenance[9,13–16] (Fig. 1). Wild chimpanzees, orangutans (*Pongo* spp.) and capuchin monkeys use leaf, wood, twig, grass and bark tools, often modifying the items to better suit the targeted task[17–19]. Wild western gorillas (*Gorilla gorilla*) have also recently been reported to use branch tools for postural support[20]. These observations of primate plant-tool use expand the conceptual framework for our interpretations of early hominin behaviour, as hominin wooden artefacts are only known from within the past 800,000 years[21], and poor preservation of organic tools in the archaeological record means a significant amount of information about the origins of human technology is now lost. Use-wear traces and microscopic residues on stone artefacts may provide indirect evidence of plant-tool manufacture in the more remote past[22].

The enforced focus on inorganic archaeological materials has led to the development of highly informative methods of measuring and interpreting stone artefacts[23], often centring on raw-material selection and transport, and on the manufacturing process and resultant forms. The key characteristic distinguishing stone from most organic materials is the potential for controlled formation of acute margins during fracture (that is, detachment of a flake from a larger core), with the degree of

[1]Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, Cambridge CB2 1QH, UK. [2]Institute of Archaeology, University College London, London WC1H 0PY, UK. [3]Department of Anthropology, Rutgers University, New Brunswick, New Jersey 08901, USA. [4]Department of Anthropology, Kenyon College, Gambier, Ohio 43022, USA. [5]Department of Anthropology, Miami University, Oxford, Ohio 45056, USA. [6]Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan. [7]Department of Archaeology, University of Calgary, Calgary, Alberta T2N 1N4, Canada. [8]Division de Prehistoria, Universidad Autonoma de Barcelona, Barcelona 08193, Spain. [9]Maison de l'Archéologie et de l'Ethnologie, CNRS, Paris 92023, France. [10]Istituto di Scienze e Tecnologie della Cognizione, Rome 00197, Italy.

Figure 1 | Chimpanzee plant use. **a**, Plants as tools: fishing for termites (Gombe Stream National Park, Tanzania). Chimpanzee extractive foraging such as this varies culturally between groups, independently of biological or environmental influences. **b**, Plants as construction: sleeping in a nest (Gombe Stream National Park). Chimpanzees make a 'bed' every night, and recent work has shown that reoccupation of the same localities may result in recognizable, patterned debris accumulation. **c**, Plants as activity sites: cracking oil-palm nuts using a tree-branch anvil and stone hammer (Liberia). Photo, A. C. Hannah.

control reflecting motor skill and cognitive capacities[6]. The resulting sharp edges can be used for tasks such as butchering animals, processing plants or trimming wood. Oldowan assemblages dated to ~2.6–1.6 Myr ago typically consist of minimally worked cores and flakes with associated debris, although careful reconstructions have shown that on occasion more than 70 flakes were struck from a single cobble[6]. Oldowan sites also contain battered (not systematically flaked) stone tools (Fig. 2a), and although recognized early on as hammers and anvils[24], only recently have these been analysed in detail and compared with chimpanzee pounding tools[25]. Recovery of cut-marked and damaged fossil bones at some early archaeological sites[26] suggests that cutting edges were exploited primarily for greater access to meat or marrow. However, as a wide variety of plants were available to Oldowan hominins[27], the relative emphasis on vertebrate faunal exploitation over that of vegetation remains hypothetical. Early hominin invertebrate exploitation is little discussed, but is ripe for further enquiry owing to its prevalence in the primate world[28,29].

Chimpanzees and capuchins use stone hammer-and-anvil combinations to crack hard-shelled nuts[18,30] (Fig. 2b, c), as well as anvils for opening hard-shelled fruits[31,32]. Island-dwelling long-tailed macaques (*Macaca fascicularis*) use stones to crack molluscs and crabs[33], and bearded capuchins use stones to dig for tubers and to process

Figure 2 | Primate stone-tool use. **a**, Three ~1.7–1.6-Myr-old Oldowan pounding tools from Olduvai Gorge, Tanzania. Provenance (left to right): FLK North level 1–2; FLK North level 5; FLK North sandy conglomerate. Scale bars, 1 cm. **b**, Chimpanzee cracking nuts with a stone hammer and anvil (Bossou, Guinea). The full social complexity of this activity cannot be reconstructed from the archaeological record alone. **c**, Adult male capuchin cracking nuts using a stone hammer and wood anvil (Boa Vista, Brazil). Note erect body position and relatively large (1.44-kg) hammer.

plants such as cactus[19,34]. These tools can become fractured, abraded or pitted as a result of repeated impact damage, leaving potentially diagnostic use-wear patterns (Fig. 3). Although non-human primates have not been seen systematically creating sharp stone flakes, fortuitous breakage of an anvil margin while processing plant foods has been hypothesized as a first step towards deliberate stone knapping[32]. The origins of intentionally flaked stone technology are therefore an important meeting point between primatology and archaeology, and archaeologists regularly invoke chimpanzee plant-processing behaviour in discussions of either pounding tools or the accidental creation of flakes from anvils[35,36]. Continued input from both fields is required to resolve incongruities between the hypotheses that vegetal processing led to the critical discovery of stone flaking, but that meat and marrow processing were the first uses of the discovery.

The oldest archaeological sites have no hominin fossils in direct association with stone artefacts[5,6]. *Homo habilis*, currently the earliest recognized member of genus *Homo*, lived ~2.4–1.4 Myr ago[37] and was hypothesized to be the creator of the initial Oldowan tools[38]. However, the time gap between the start of systematic stone flaking and the earliest *Homo* fossils suggests that earlier hominin taxa could be the innovators[39]. The relative brain sizes and manipulative abilities of the hominin species potentially ancestral to early *Homo* (that is, *Australopithecus africanus*, *Australopithecus afarensis* and *Australopithecus garhi*) appear to be roughly equivalent to those of the extant great apes[8,40–42]. This has prompted studies of non-human-primate capabilities. When they are motivated, captive orangutans[43], bonobos[44] and capuchin monkeys[45] are capable of rudimentary stone reduction leading in some cases to sharp-edged, flaked cutting tools. The bonobo results are distinguishable from Oldowan artefacts, although the main bonobo subject displayed an unprompted inclination throughout the study to break rocks by throwing them onto a hard substrate, rather than using the typical Oldowan method of striking one hand-held stone against another[46]. This suggests an alternative avenue through which australopithecines and earlier hominins may have obtained sharp stone edges, and raises the possibility that extinct panins (the chimpanzee lineage since the split with humans) may have used the same technique.

Aspects of modern primate anatomy, such as a capacity for bipedal posture and the structure of the forelimb, wrist and hands[47], provide valuable comparisons in assessing the capacities of hominins living before 2.6 Myr ago. Primate dentition is sufficient for many scraping, slicing, crushing and trimming tasks[40,48], and this has been offered as an explanation for the apparent absence of stone flaking among wild non-human primates. Documenting the selective feedback from both technology and environment in shaping anatomical traits therefore forms an important research focus of primate archaeology. That both 45-kg chimpanzees and 3-kg capuchins use stone hammers averaging around 1 kg or more in mass[30,49] demonstrates that we cannot posit an a-priori relationship between the size of a primate and its potential tools. Similarly, *Cebus* has a significantly higher encephalization quotient (the ratio of brain size to expected brain size as calculated from body mass or metabolism) than do most other primates, including *Pan* and the extinct hominins before *Homo*[50]. Ideally, comparative anatomy allows us to track the divergent paths taken by the various primate lineages since their split. For panins, however, we are hindered by an almost total lack of fossils[51]. The hominin fossil record is better represented, but the likelihood of past adaptive radiations makes it unclear which of the known extinct species are direct ancestors of the human line.

It is also unclear which of the extinct hominins were tool users. *Australopithecus garhi* has been suggested as being responsible for cut-marked bones dated to 2.5 Myr ago in Ethiopia[26], and *Australopithecus robustus* has been posited as being a potential bone-tool user in South Africa[28]. Chimpanzees show cultural variation in their tool use; for example, not all populations use stone tools to crack nuts[14], and wild bonobos and gorillas have not been observed to use extractive tools of any kind[10]. Thus, we should not assume that groups of extinct hominins were homogenous in their forms of material culture. Moreover, the parsimonious explanation that widespread *Homo* and chimpanzee material culture indicates a tool-using common ancestor cannot reasonably be extended to the common ancestor of apes and *Cebus* or *Macaca*. The New and Old World monkeys split from the line leading to the apes around 35 and 25 Myr ago, respectively[52], and wild tool use in these lineages is rare. Convergence in tool use among primates is therefore plausible, probably owing to similar adaptive pressures, foraging requirements and physiological constraints. Significantly, acceptance of convergence in monkeys opens the door to the possibility of repeated gain and loss of tool use among extinct primates in multiple lineages over millions of years. From this perspective, it is unreasonable to expect that the only tool-using, non-hominin primates are extant species.

For a comprehensive comparative study of the development of technology, we must establish the antiquity and form of tools used by non-human-primate ancestors (and potentially also tool-using species outside the primates), along with their ecological contexts. This aim has been advanced through excavation of chimpanzee nut-cracking sites in Taï National Park, Côte d'Ivoire[53,54], where finds of accumulated cultural refuse dating back at least 4,300 years helped establish the legitimacy of primate-archaeological research. Recovered tools were interpreted as fragments of pounding stones used for processing nuts, supported by discriminatory analysis of adhering starch residues. Ongoing studies at Bossou and Diecke in Guinea focus on chimpanzee tool selection, spatial patterning and functional characteristics for a number of surface-context, present-day nut-cracking localities[55] (Fig. 4a). Similarly, research on wild bearded capuchins has examined such factors as the wear traces left by repeated use of anvils for pounding hard-shelled nuts[30], tool transport and selectivity[56] and the kinematics and energetics of nut cracking[57]. Capuchins consistently select hammer stones that are appropriate for the task by size and weight, displaying planning abilities and the rapid employment of visual, acoustic and haptic clues to correctly identify suitable tools[58]. Preferential accumulation of such materials at sites with appropriate anvils—the latter identifiable from wear traces—produces a capuchin 'activity area' (Fig. 4b), the remains of which may last for millennia.

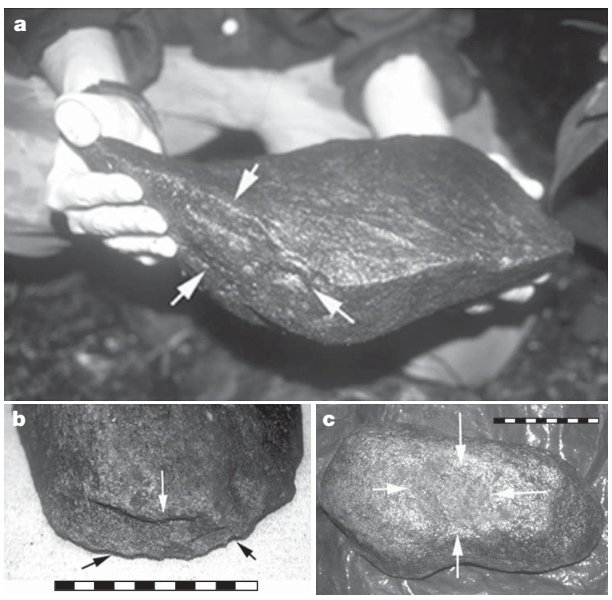


Figure 3 | Wear patterns on chimpanzee pounding stone tools. Repeated use of favoured hammers and anvils builds up distinctive damage patterns, permitting archaeological identification of these tools. **a**, **b**, Flaking, crushing, pitting and corner removals. **c**, Cavity formation. Scale bar divisions, 1 cm. Photos, C. Boesch and J.M. (Taï National Park, Côte d'Ivoire).

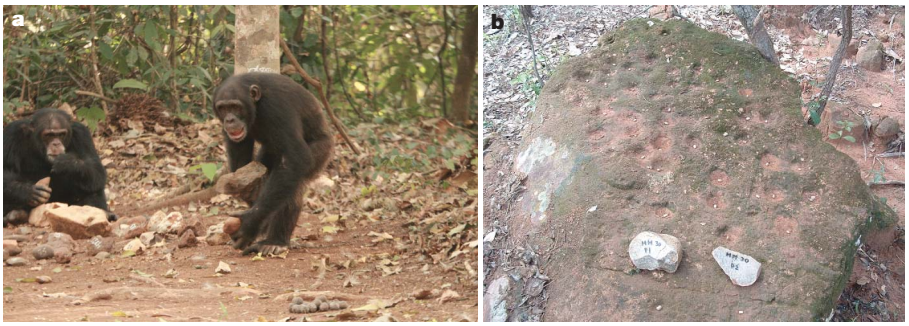


Figure 4 | Primate site creation. a, Selective transport of a stone hammer and anvil by an adolescent male chimpanzee (Bossou, Guinea). Over time, this behaviour accumulates artefacts in preferred tool-use sites. **b**, Stone anvil pitted by capuchin nut-cracking activity (Boa Vista, Brazil). Two stone hammers were found on the anvil, and nut debris was removed from the anvil before recording. Image from ref. 30.

Chimpanzee and capuchin spatial concentration of pounding tools such as anvils and hammer stones, and repeated, frequently seasonal, re-use of locations such as nut-tree sites, can form recognizable non-human archaeological assemblages[53]. The viability of identifying pounding tools from pre-Oldowan times onwards is strengthened by modern actualistic studies—of nut cracking, meat processing, marrow extraction, digging-stick manufacture and stone-tool manufacture—that leave distinct use-damage patterns on the tools and act as proxies for identifying past pounding activities. Recording use-wear traces on experimental anvils and hammer stones, and on those used by free-ranging primates, also gives a clearer understanding of the dynamic behavioural and decision-making processes involved. Such cross-taxa comparative work is still in its infancy; however, ongoing research applying this methodology to Plio-Pleistocene anvils from Koobi Fora, Kenya, and Olduvai Gorge[25] suggests that the various pounding activities may be differentiated by their use traces. Field experiments exposing wild chimpanzees to a variety of raw materials for extractive foraging complement this work[59], as does the study of percussion-marked fossil bones[60].

Pioneering studies applying archaeological recording methods to the spatial distribution of chimpanzee nests[61,62] suggested that chimpanzees accumulate organic artefacts through repeated occupation of a site. Current work confirms this hypothesis[63], providing insight into early hominin site formation processes. A recent report of wooden digging tools used by savanna-woodland chimpanzees also demonstrates the importance of an archaeological approach to primate plant-tool use[15]. The presence of digging tools, holes, masticated underground storage organs, faeces and knuckle prints indicated that chimpanzees had been digging for underground storage organs, despite the absence of direct observations of this activity. Wild capuchins also use sticks to probe[19]; moreover, they exhibit repeated use of caves to sleep in at night, to thermoregulate during the hottest part of the day, and to seek refuge in during heavy rain (Etho*Cebus* project, unpublished observations). All primates leave remnants that can enter the archaeological record (for example faeces as coprolites), and behavioural patterns that emphasize repeated use of particular tree, cliff or cave sites[64,65] may result in spatially patterned and therefore detectable accumulations in both open and forested environments. These findings have clear parallels with the archaeological interpretation of hominin activity areas; however, recognition of the highly uneven preservation of this ephemeral evidence in the archaeological record is challenging.

The role of primate archaeology

Acknowledging that only some primate taxa exhibit technology, despite their inherited biological similarities, primate archaeology has the following interconnected aims. The first goal is to use concurrent archaeological and primatological methods for recording living non-human-primate tool use to learn about the evolutionary trajectories of primate behaviour from both anthropocentric and 'primatocentric' perspectives. The second is to employ comparative studies of similar technologies across human and primate taxa (for example pounding, probing and digging tools) to investigate the origins of tool use among primates and its social, physiological, biomechanical and environmental contexts. The third is to examine the spatially patterned behavioural characteristics of primates that do not use tools, to determine whether these patterns are recognizable in the material record and how they might inform us about the likely activities of extinct primates (including hominins that did not use tools). The fourth aim is to identify how convergence has shaped primate tool use, and the implications that this has for the technologies of extinct primate taxa.

In implementing this agenda, the most complex results will probably derive from stone-tool-using species such as humans, chimpanzees, capuchins and macaques, as these generate a durable material record. However, until research methods in this field are much more rigorous and well established, it would be premature to make decisions on the potential contribution of any one taxon. Much of our knowledge is very recent: a mere decade ago the list of known wild-primate stone-tool users included only humans and chimpanzees, and half a century ago solely humans.

Despite a number of research projects that examine pre-*sapiens* contexts, most archaeological investigations worldwide remain focused on the activities of *Homo sapiens*, largely within the past few tens of millennia. This agrees well with definitions of archaeology as an anthropocentric science. However, we contend that archaeology should be viewed as a set of methods for the recovery and interpretation of behavioural evidence, irrespective of the species creating the record. This view is already common in the study of species that may be our direct ancestors (for example *H. habilis*), as well as those that are not (for example *Homo neanderthalensis* and *Homo floresiensis*). The difference is that the comparative study of dynamic living-primate behaviour offers greater potential for hypothesis testing and data generation than any study of static objects left by extinct hominins. Furthermore, although terms such as 'chimpanzee archaeology'[55] and 'cultural panthropology'[66] have been used previously, they exclude non-ape primates; an inclusive term for these endeavours promotes common methods and encourages cross-taxa comparisons.

Non-human primates display stone and plant-material selection, processing and accumulation behaviours that challenge the conventional view of hominins as the sole creators of archaeological sites. If geographical and palaeoecological data suggest that a site was habitable by both hominins and other (potentially extinct) tool-using primates, then ambiguity arises over the extent to which either taxon contributed to any recovered assemblage. The only dated chimpanzee archaeological site faces this problem, human and chimpanzee stone tools being present in the same deposit[54]. The known innovative abilities of tool-using primates should therefore act as a cautionary brake in automatically assigning assemblages from such sites to any one taxon[40].

Whereas there is only one extant species of hominin and two of panin, there are eight living capuchin species[67], only one of which

(*C. libidinosus*) is known to use tools frequently in the wild[9]. Primate archaeological research into the contexts of *Cebus* long-term tool selection, tool manipulation and spatial and social organizational strategies therefore offers one of the clearest opportunities to uncover important variables in the emergence of material culture. For example, we need to quantify the energetic costs of being a habitual tool user, and assess how these costs affect differential survival. As mentioned above, acceptance of the convergent evolution of tool use among Old and New World monkeys means that we should reasonably expect other species to have discovered, employed and then lost tool use throughout the past. On the basis of extant primate behaviour, we can expect some of these instances to have involved stone tools, and for those artefacts still to be recoverable today. Primate archaeology enables us to discover whether or not the long record of stone-tool use in the hominin line (at present covering around half the period since the common ancestor with *Pan*) is exceptional.

Sharp-edged tools are not the only stone artefacts produced during the Oldowan and later periods, and more material was used for percussion than for flaking at some early sites[25]. A perspective that emphasizes hominin systematic flaking obscures the potential contribution of non-hominin primates to the known archaeological record, and inhibits meaningful discussion of the factors behind cross-taxa parallels evident in other areas of technology. These factors include energy trade-offs between diet and tool manufacture, inter- and intraspecies variability in resource exploitation and tool function, intentionality in tool material and attribute selection, and modes of social transmission. From a longer-term perspective, primate archaeology also contributes essential data to debates over the necessity and sufficiency of tool use in the evolution of bipedalism and terrestriality.

Comparisons between taxa that have close morphological and genetic affinities with humans form a naturally strong research basis. However, the rationale behind primate archaeology applies also to non-primates, because any animal that accumulates or modifies durable materials will leave a signature in the archaeological record. Animals as diverse as crows[68], dolphins[69], beavers[70] and bowerbirds[71] selectively manipulate material objects in regular patterns for extractive foraging, niche construction and sexual display. Animals that use stones to break open resources, for example the portable anvils of Californian sea otters[72] or fixed nut-cracking anvils of New Caledonian crows[73], may leave recoverable impact signatures that inform on the time depth and adaptive value of tool use in these species. Animal behavioural studies typically concentrate on current abilities rather than evidence for the evolutionary development of object manipulation, and even those field sites with multidecadal research data (for example Gombe Stream National Park for chimpanzees) have recorded only a tiny fraction of the lifespan of a species. A clearer understanding of the behavioural remnants of non-human animals provides a 'control experiment' on the uniqueness of human artefacts.

The future of primate archaeology

Primate archaeology places the entirety of human behavioural evolution into its wider comparative biological context, and its establishment clarifies the need for refinement of existing research programs. First, there is a need for systematic collaboration: few archaeologists have seen a wild primate use a tool, and few primatologists have taken part in archaeological excavations. Integrated involvement of disciplines outside these areas, including comparative anatomy, cognitive science and evolutionary ecology, will also be essential. Second, standardization of recording procedures will combine an archaeological focus on material attributes and actions with the comprehensive spatially, temporally and socially patterned behavioural data generated by primatology. Third, creation of a comprehensive database comparing cross-taxa tool making and tool use in their environmental, biological and social contexts will enable exploration of the selective costs and benefits involved, with the ultimate aim of identifying those circumstances driving the uptake and continuance of tool use among primates, including human ancestors. Finally, direct documentation of the use of both human- and non-human-primate tools will allow us to tie specific actions to resultant residues and wear patterns, in turn helping us to understand why archaeologically recovered artefacts were created or chosen, producing valuable data on cognitive evolution.

Greater recognition of the value of primate tool use and site creation to human evolutionary studies can only help in the protection of rapidly declining primate populations. Field observation of the frequency of site visitation and the processes of site accumulation produces clues to the timescales represented at known hominin sites. Linking data to identified individuals within a primate group provides rich contextual information on cultural transmission and demographic variation in tool use within social and kinship networks[74], as well as emphasizing the roles of age, sex and experience in phenomena such as tool standardisation, connecting artefacts to social roles and ontogeny. All such data help clarify those characteristics of technology that differentiate primates from other tool-using species. Crucially, excavation of primate activity areas adds time depth to these observations; other than for the later hominins, the span of time over which any species has used tools is currently unknown.

Recent years have seen the first observations of wild tool use by populations of orangutans, capuchins and macaques, and a dramatic increase in our knowledge of the range, variability and time depth of chimpanzee material culture. We have also seen the announcement of the world's oldest hominin tools and the revelation of unexpected skill in their manufacture. With the likelihood of even earlier hominin tool use, the proposed addition of new taxa to the hominin family tree[42,75,76] has added new complexity to our evolutionary history. Primate archaeology provides the necessary methodology and scope to incorporate these developments into a coherent framework for the systematic location and interpretation of evolutionarily significant material culture, opening up new avenues in our understanding of the place of technology in primate and human societies.

1. Beck, B. *Animal Tool Behavior: The Use and Manufacture of Tools* (Garland STPM, 1980).
2. Washburn, S. Tools and human evolution. *Sci. Am.* **203,** 62–75 (1960).
3. McGrew, W. C. *Chimpanzee Material Culture: Implications for Human Evolution* (Cambridge Univ. Press, 1992).
4. Kumar, S., Filipski, A., Swarna, V., Walker, A. & Hedges, S. B. Placing confidence limits on the molecular age of the human–chimpanzee divergence. *Proc. Natl Acad. Sci. USA* **102,** 18842–18847 (2005).
5. Semaw, S. *et al.* 2.5-million-year-old stone tools from Gona, Ethiopia. *Nature* **385,** 333–336 (1997).
6. Delagnes, A. & Roche, H. Late Pliocene hominid knapping skills: the case of Lokalalei 2C, West Turkana, Kenya. *J. Hum. Evol.* **48,** 435–472 (2005).
This paper provides the best evidence so far for the advanced technological skill involved in making some of the world's oldest known stone artefacts, leading to suggestions that earlier tool manufacture is likely.
7. Foley, R. Hominid species and stone-tool assemblages: how are they related? *Antiquity* **61,** 380–392 (1987).
8. Panger, M., Brooks, A., Richmond, B. G. & Wood, B. Older than the Oldowan? Rethinking the emergence of hominin tool use. *Evol. Anthropol.* **11,** 235–245 (2002).
9. Ottoni, E. & Izar, P. Capuchin monkey tool use: overview and implications. *Evol. Anthropol.* **17,** 171–178 (2008).
This paper is the first study to summarize the burgeoning results from observations of wild-capuchin tool use, placing these into their taxonomic, regional and ecological contexts.
10. Hohmann, G. & Fruth, B. Culture in bonobos? Between-species and within-species variation in behavior. *Curr. Anthropol.* **44,** 563–570 (2003).
11. Teleki, G. Chimpanzee subsistence technology: materials and skills. *J. Hum. Evol.* **3,** 575–594 (1974).
12. St Amant, R. & Horton, T. E. Revisiting the definition of animal tool use. *Anim. Behav.* **75,** 1199–1208 (2008).
13. Boesch, C. & Boesch, H. Tool use and tool making in wild chimpanzees. *Folia Primatol. (Basel)* **54,** 86–99 (1990).
14. McGrew, W. C. *The Cultured Chimpanzee: Reflections on Cultural Primatology* (Cambridge Univ. Press, 2004).
15. Hernandez-Aguilar, R. A., Moore, J. & Pickering, T. R. Savanna chimpanzees use tools to harvest the underground storage organs of plants. *Proc. Natl Acad. Sci. USA* **104,** 19210–19213 (2007).

16. Pruetz, J. D. & Bertolani, P. Savanna chimpanzees, *Pan troglodytes verus*, hunt with tools. *Curr. Biol.* **17,** 412–417 (2007).
17. van Schaik, C. P., Fox, E. A. & Sitompul, A. F. Manufacture and use of tools in wild Sumatran orangutans. Implications for human evolution. *Naturwissenschaften* **83,** 186–188 (1996).
18. Whiten, A. *et al.* Cultures in chimpanzees. *Nature* **399,** 682–685 (1999).
19. Mannu, M. & Ottoni, E. The enhanced tool-kit of two groups of wild bearded capuchin monkeys in the Caatinga: tool making, associative use, and secondary tools. *Am. J. Primatol.* **71,** 242–251 (2009).
20. Breuer, T., Ndoundou-Hockemba, M. & Fishlock, V. First observation of tool use in wild gorillas. *PLoS Biol.* **3,** e380 (2005).
21. Goren-Inbar, N., Werker, E. & Feibel, C. S. *The Acheulian Site of Gesher Benot Ya'aqov, Israel: The Wood Assemblage* (Oxbow Books, 2002).
22. Keeley, L. H. & Toth, N. Microwear polishes on early stone tools from Koobi Fora, Kenya. *Nature* **293,** 464–465 (1981).
23. Inizan, M.-L., Reduron-Balliner, M., Roche, H. & Tixier, J. *Technologie de la Pierre Taillée* (CREP, 1995).
24. Leakey, M. D. *Olduvai Gorge. Excavations in Beds I and II, 1960–1963* 7 (Cambridge Univ. Press, 1971).
25. Mora, R. & de la Torre, I. Percussion tools in Olduvai Beds I and II (Tanzania): implications for early human activities. *J. Anthropol. Archaeol.* **24,** 179–192 (2005).
26. de Heinzelin, J. *et al.* Environment and behavior of 2.5-million-year-old Bouri hominids. *Science* **284,** 625–629 (1999).
27. Copeland, S. R. Vegetation and plant food reconstruction of lowermost Bed II, Olduvai Gorge, using modern analogs. *J. Hum. Evol.* **53,** 146–175 (2007).
28. Backwell, L. R. & d'Errico, F. Evidence of termite foraging by Swartkrans early hominids. *Proc. Natl Acad. Sci. USA* **98,** 1358–1363 (2001).
29. McGrew, W. C. in *Meat-Eating and Human Evolution* (eds Stanford, C. & Bunn, H.) 160–178 (Oxford Univ. Press, 2001).
30. Visalberghi, E. *et al.* Characteristics of hammer stones and anvils used by wild bearded capuchin monkeys (*Cebus libidinosus*) to crack open palm nuts. *Am. J. Phys. Anthropol.* **132,** 426–444 (2007).
This paper records physical characteristics of capuchin stone tools, providing the first systematic baseline data for comparison with the better-known chimpanzee and archaeological hominin records.
31. Boinski, S. *et al.* in *The Biology of Traditions: Models and Evidence* (eds Fragaszy, D. & Perry, D.) 365–390 (Cambridge Univ. Press, 2003).
32. Marchant, L. F. & McGrew, W. C. in *Stone Knapping: The Necessary Conditions for a Uniquely Hominin Behaviour* (eds Roux, V. & Bril, B.) 341–350 (McDonald Institute for Archaeological Research, 2005).
This paper reports on fruit-smashing behaviour by chimpanzees, and outlines a proposal (testable by means of the archaeological record) for the series of steps early hominins may have taken to move from simple anvil use to sophisticated flaked-stone manufacture.
33. Malaivijitnond, S. *et al.* Stone-tool usage by Thai long-tailed macaques (*Macaca fascicularis*). *Am. J. Primatol.* **69,** 227–233 (2007).
34. de A, Moura, A. & Lee, P. Capuchin stone tool use in caatinga dry forest. *Science* **306,** 1909 (2004).
35. Joulian, F. in *Modelling the Early Human Mind* (eds Mellars, P. & Gibson, K.) 173–189 (McDonald Institute for Archaeological Research, 1996).
36. Goren-Inbar, N., Sharon, G., Melamed, Y. & Kislev, M. Nuts, nut cracking, and pitted stones at Gesher Benot Ya'aqov, Israel. *Proc. Natl Acad. Sci. USA* **99,** 2455–2460 (2002).
37. Wood, B. & Lonergan, N. The hominin fossil record: taxa, grades and clades. *J. Anat.* **212,** 354–376 (2008).
38. Leakey, L. S., Tobias, P. V. & Napier, J. R. A new species of the genus *Homo* from Olduvai Gorge. *Nature* **202,** 7–9 (1964).
39. Plummer, T. Flaked stones and old bones: biological and cultural evolution at the dawn of technology. *Yearb. Phys. Anthropol.* **125,** 118–164 (2004).
40. Wynn, T. G. & McGrew, W. C. An ape's view of the Oldowan. *Mankind* **24,** 383–398 (1989).
41. Kappelman, J. The evolution of body mass and relative brain size in fossil hominids. *J. Hum. Evol.* **30,** 243–276 (1996).
42. Asfaw, B. *et al. Australopithecus garhi*: a new species of early hominid from Ethiopia. *Science* **284,** 629–635 (1999).
43. Wright, R. V. S. Imitative learning of a flaked tool technology: the case of an orangutan. *Mankind* **8,** 296–306 (1972).
44. Toth, N., Schick, K. & Semaw, S. in *The Oldowan: Case Studies into the Earliest Stone Age* (eds Toth, N. & Schick, K.) 155–222 (Stone Age Institute, 2006).
45. Westergaard, G. C. The stone-tool technology of capuchin monkeys: possible implications for the evolution of symbolic communication in hominids. *World Archaeol.* **27,** 1–9 (1995).
46. Schick, K. D. *et al.* Continuing investigations into the stone tool-making and tool-using capabilities of a bonobo (*Pan paniscus*). *J. Archaeol. Sci.* **26,** 821–832 (1999).
47. Tocheri, M. W., Orr, C. M., Jacofsky, M. C. & Marzke, M. W. The evolutionary history of the hominin hand since the last common ancestor of *Pan* and *Homo*. *J. Anat.* **212,** 544–562 (2008).
48. Nishida, T. The bark-eating habits in primates, with special reference to their status in the diet of wild chimpanzees. *Folia Primatol. (Basel)* **25,** 277–287 (1976).
49. Boesch, C. & Boesch, H. Optimization of nut-cracking with natural hammers by wild chimpanzees. *Behaviour* **83,** 265–286 (1983).
50. Aiello, L. & Dean, C. *An Introduction to Human Evolutionary Anatomy* 191 (Academic, 1990).
51. McBrearty, S. & Jablonski, N. First fossil chimpanzee. *Nature* **437,** 105–108 (2005).
52. Schrago, C. G. & Russo, C. A. Timing the origin of New World monkeys. *Mol. Biol. Evol.* **20,** 1620–1625 (2003).
53. Mercader, J., Panger, M. & Boesch, C. Excavation of a chimpanzee stone tool site in the African rainforest. *Science* **296,** 1452–1455 (2002).
54. Mercader, J. *et al.* 4,300-year-old chimpanzee sites and the origins of percussive stone technology. *Proc. Natl Acad. Sci. USA* **104,** 3043–3048 (2007).
This paper reports the oldest dated chimpanzee activity area, and establishes a series of test criteria for distinguishing human from non-human artefacts that includes the presence of flaking features and distinct use residues.
55. Carvalho, S., Cunho, E., Sousa, C. & Matsuzawa, T. *Chaînes opératoires* and resource-exploitation strategies in chimpanzee (*Pan troglodytes*) nut cracking. *J. Hum. Evol.* **55,** 148–163 (2008).
This is so far the most comprehensive study bringing archaeological methodology to the study of primate tools, applying the theoretical concept of operational sequences to chimpanzee nut cracking and highlighting variability in tool function and regional diversity in tool use.
56. Visalberghi, E. *et al.* Distribution of potential suitable hammers and transport of hammer tools and nuts by wild capuchin monkeys. *Primates* **50,** 95–104 (2009).
57. Liu, Q. *et al.* Kinematics and energetics of nut-cracking in wild capuchin monkeys (*Cebus libidinosus*) in Piauí, Brazil. *Am. J. Phys. Anthropol.* **138,** 210–220 (2009).
58. Visalberghi, E. *et al.* Selection of effective stone tools by wild bearded Capuchin monkeys. *Curr. Biol.* **19,** 213–217 (2009).
59. Matsuzawa, T. in *Chimpanzee Cultures* (eds Wrangham, R., McGrew, W. C., de Waal, F. & Heltne, P.) 351–370 (Harvard Univ. Press, 1994).
60. Blumenschine, R. J. & Selvaggio, M. M. Percussion marks on bone surfaces as a new diagnostic of hominid behaviour. *Nature* **333,** 763–765 (1988).
61. Sept, J. M. Was there no place like home? A new perspective on early hominid archaeological sites from the mapping of chimpanzee nests. *Curr. Anthropol.* **33,** 187–207 (1992).
62. Sept, J. Shadows on a changing landscape: comparing nesting patterns of hominids and chimpanzees since their last common ancestor. *Am. J. Primatol.* **46,** 85–101 (1998).
63. Hernandez-Aguilar, A. Chimpanzee nest distribution and site re-use in a dry habitat: implications for early hominin ranging. *J. Hum. Evol.* (in the press).
64. Anderson, J. R. Sleep-related behavioural adaptations in free-ranging anthropoid primates. *Sleep Med. Rev.* **4,** 355–373 (2000).
65. McGrew, W. C., McKee, J. & Tutin, C. Primates in caves: two new reports of *Papio* spp. *J. Hum. Evol.* **44,** 521–526 (2003).
66. Whiten, A., Horner, V. & Marshall-Pescini, S. Cultural panthropology. *Evol. Anthropol.* **12,** 92–105 (2003).
67. Fragaszy, D., Visalberghi, E. & Fedigan, L. *The Complete Capuchin: The Biology of the Genus Cebus* (Cambridge Univ. Press, 2004).
68. Hunt, G. R. Manufacture and use of hook-tools by New Caledonian crows. *Nature* **379,** 249–251 (1996).
69. Mann, J. *et al.* Why do dolphins carry sponges? *PLoS One* **3,** e3868 (2008).
70. Rybczynski, N. Woodcutting behavior in beavers (Castoridae, Rodentia): estimating ecological performance in a modern and a fossil taxon. *Paleobiology* **34,** 389–402 (2008).
71. Madden, J. R. Do bowerbirds exhibit cultures? *Anim. Cogn.* **11,** 1–12 (2008).
72. Hall, K. R. L. & Schaller, G. B. Tool-using behavior of the California sea otter. *J. Mamm.* **45,** 287–298 (1964).
73. Hunt, G. R., Sakuma, F. & Shibata, Y. New Caledonian crows drop candle-nuts onto rock from communally used forks on branches. *Emu* **102,** 283–290 (2002).
74. Biro, D. *et al.* Cultural innovation and transmission of tool use in wild chimpanzees: evidence from field experiments. *Anim. Cogn.* **6,** 213–223 (2003).
75. Leakey, M. G., Spoor, F. & Brown, F. New hominin genus from eastern Africa shows diverse middle Pliocene lineages. *Nature* **410,** 433–440 (2001).
76. Brunet, M. *et al.* A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418,** 145–151 (2002).

Acknowledgements We thank R. Foley and the US National Science Foundation Revealing Hominid Origins Initiative. The University of Cambridge, the Galton Institute and King's College, Cambridge, UK, sponsored the symposium 'Primatology Meets Palaeoanthropology 2: The Origins of Percussive Technology', held in October 2008 at the Leverhulme Centre for Human Evolutionary Studies in Cambridge.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence should be addressed to M.H. (mah66@cam.ac.uk).

ARTICLES

The *Schistosoma japonicum* genome reveals features of host–parasite interplay

The Schistosoma japonicum Genome Sequencing and Functional Analysis Consortium*

***Schistosoma japonicum* is a parasitic flatworm that causes human schistosomiasis, which is a significant cause of morbidity in China and the Philippines. Here we present a draft genomic sequence for the worm. The genome provides a global insight into the molecular architecture and host interaction of this complex metazoan pathogen, revealing that it can exploit host nutrients, neuroendocrine hormones and signalling pathways for growth, development and maturation. Having a complex nervous system and a well-developed sensory system, *S. japonicum* can accept stimulation of the corresponding ligands as a physiological response to different environments, such as fresh water or the tissues of its intermediate and mammalian hosts. Numerous proteases, including cercarial elastase, are implicated in mammalian skin penetration and haemoglobin degradation. The genomic information will serve as a valuable platform to facilitate development of new interventions for schistosomiasis control.**

Schistosomiasis is an ancient scourge of mankind, depicted graphically in papyri from Pharaonic Egypt and known from human remains over 2,000 years old from China[1,2]. Blood-dwelling trematodes (phylum Platyhelminthes) of the genus *Schistosoma* cause this chronic and debilitating disease, which afflicts more than 200 million people in 76 tropical and subtropical countries. Morbidity is high and schistosomiasis contributes to several hundreds of thousands of deaths annually[3–5]. Three principal species can infect humans: *Schistosoma japonicum*, *Schistosoma mansoni* and *Schistosoma haematobium*. The first of these is prevalent in the Philippines and parts of Indonesia, and is a major disease risk for 66 million people living in southern China[2]. It remains a major public health concern in China despite over 50 years of concerted campaigns for its control[2,6]. Approximately one million people in China, and more than 1.7 million bovines and other mammals, are currently infected[2]. Control measures include community-based praziquantel chemotherapy, health education, improved sanitation, environmental modification and snail control. However, additional approaches, such as the development and deployment of new drugs and anti-schistosome vaccines are urgently needed to meet the prevailing challenges, which include the spectre of praziquantel-resistant parasites[7,8].

During their complex developmental cycle, schistosomes alternate between a mammalian host and a snail host through the medium of fresh water. After burrowing out of the snail host, free-swimming cercariae penetrate the skin of the mammalian host, travel through the blood to the liver via the lungs, and transform into schistosomula. These mature in the hepatic portal vein, mate and, in the case of *S. japonicum*, migrate to their final destination in the mesenteric venous plexus. Female worms release thousands of eggs daily, which are discharged in the faeces after a damaging passage through the intestinal wall. If they reach fresh water, eggs hatch to release free-swimming ciliated miracidia, which, guided by light and chemical stimuli, seek amphibious snails of the genus *Oncomelania*. Within the hemocoel of the snail, miracidia give rise asexually to numbers of sporocysts, in which further asexual propagation produces numerous cercariae.

Eggs deposited by adult female schistosomes embolize in the liver, intestines and other tissue sites and are the key contributors to the pathology and associated morbidity of schistosomiasis. Notably, the highly adapted relationship between schistosomes and their snail intermediate and mammalian definitive hosts appears to involve exploitation by the parasite of host endocrine and immune signals[9,10]. The evasion strategies that underpin avoidance of the host immune system, allowing schistosomes to survive for years despite strong host immune responses, have long interested investigators intent on development of an efficacious vaccine.

Unlike most other platyhelminths, schistosomes are dioecious. The genome is arrayed on eight pairs of chromosomes, seven pairs of autosomes and one pair of sex chromosomes. Females are the heterogametic sex (ZW); males are homogametic (ZZ)[11,12]. No other lophotrochozoan[13] has yet been sequenced.

Genome features and evolution

General information. The whole-genome shotgun (WGS) sequencing strategy was used to decode the 397-megabase-pair (Mb) sequences, covering most (>90%) of the *S. japonicum* genome (Supplementary Tables 1 and 2 and Supplementary Fig. 1). A total of 13,469 protein-coding genes were identified, comprising about 4% of the draft *S. japonicum* genome (Supplementary Figs 2 and 3). Of the protein-coding genes, 6,972 (52%) were mapped to categories established by the Gene Ontology project (Fig. 1a and Supplementary Fig. 4) and an orthologue relationship existed between 2,516 (19%) of them and 1,546 Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology groups (Supplementary Fig. 5). *Schistosoma japonicum* has a relatively large genome and low gene density in comparison with other invertebrates, including *Brugia malayi* (Table 1). On the basis of the outbred source of the genomic libraries, high-quality discrepancies found during assembly were used to identify 557,739 single nucleotide polymorphisms (SNPs) (Supplementary Table 3), with an average density of 1.4 SNPs per kilobase pair, and the insertion and deletion (indel) rates were much lower.

Repeat sequences. A total of 657 different repeat families/elements, constituting 159 Mb (40.1%) of the *S. japonicum* genome were revealed by comparing known repetitive sequences and using the software REPEATSCOUT (version 1.0.3)[14] (Fig. 2 and Supplementary Table 4). Among them, 29 kinds of retrotransposon were found,

*Lists of participants and their affiliations appear at the end of the paper.

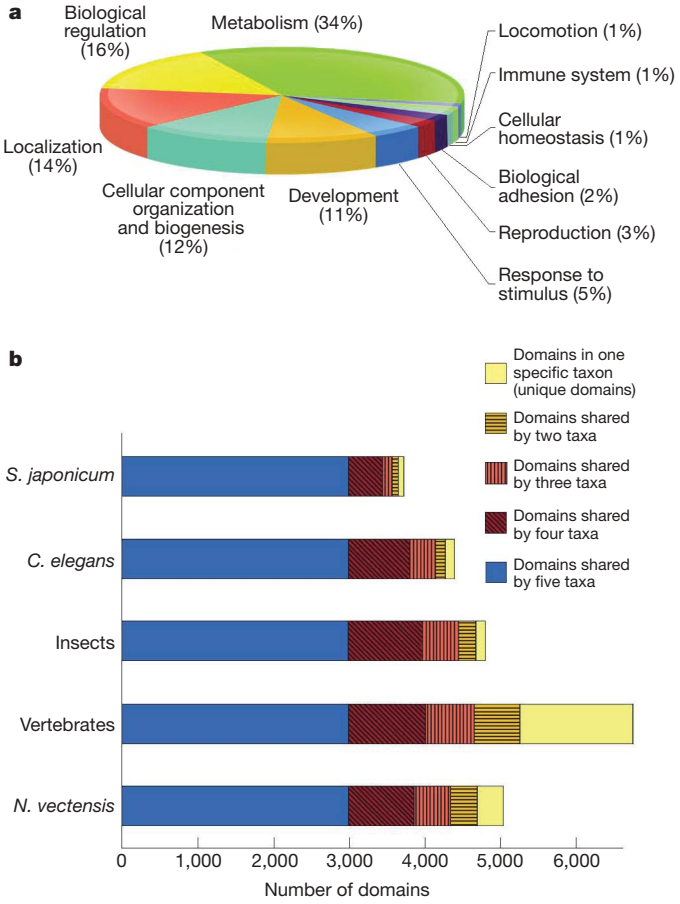


Figure 1 | Functional categorization of *S. japonicum* genes and protein-domain analysis. a, Proportion of the 6,972 *S. japonicum* proteins with functional information in different Gene Ontology categories. **b**, In *S. japonicum*, vertebrates (*H. sapiens*, *G. gallus* and *D. rerio*), insects (*D. melanogaster* and *A. gambiae*), *C. elegans* and *Nematostella vectensis*, a total of 7,562 domains were detected. The majority of *S. japonicum* domains are shared with other taxa, having the fewest unique domains, whereas vertebrates evolved significant numbers of unique protein domains.

including known *Gulliver*, *SjR1*, *SjR2* and *Sj-pido* elements as well as 25 novel elements, together constituting 19.8% of the genome (Supplementary Table 5). Of the 25 novel retrotransposons, 18 are long terminal repeat (LTR) forms, four are non-LTR forms and three are *Penelope*-like elements—enigmatic retroelements that retain introns[15]. Each type of retrotransposon was represented by one to 793 intact copies or hundreds to thousands of partial copies. The non-LTR retrotransposons have significantly higher copy numbers, constituting 12.6% of the genome.

Gene loss/duplication. It was intriguing to observe that schistosomes share more orthologues with the vertebrates (Supplementary Table 6), such as *H. sapiens* (4,324 pairs), than they do with the ecdysozoans, for example *C. elegans* (3,292), despite the Ecdysozoa and Lophotrochozoa being phylogenetically adjacent[13]. Similarly, the cnidarians and vertebrates have been shown to share more orthologous genes with each other than either does with the ecdysozoans[16]. One possible reason for this is that a higher evolutionary rate in the Ecdysozoa causes an apparently larger orthologue divergence, although the scenario of functional selection of orthologue patterns in the context of parasite–host interplay is also worth consideration.

To test possible consequences of parasitism at the genome level, we investigated gene family and domain variations between schistosomes and other metazoans. It is clear that there was minor variation in total numbers of protein families among *S. japonicum* (6,322) and the other species, such as *C. elegans* (6,669), *D. melanogaster* (5,184) and *H. sapiens* (6,877) (Supplementary Table 7). However, a major reduction in number, or even the elimination, of protein domains was apparent in the *S. japonicum* genome, in that the great majority (3,654) of the 3,728 protein domains from the flatworm were shared with other species (Fig. 1b) and can thus be considered ubiquitous among metazoans, whereas 3,834 domains found in at least one of the other species were not detected in schistosomes. Of these 3,834 domains, 1,140 were shared by more than three taxa of vertebrates, insects, a nematode and sea anemones (Supplementary Fig. 6). Notably, domain-loss events seem to be more widespread in *S. japonicum* than in any other species studied so far, including *C. elegans*, a model organism well known for rapid evolutionary rates and a high degree of gene loss[17]. Roughly 1,000 protein domains have been abandoned by *S. japonicum*, including some involved in basic metabolic pathways and defence, implying that loss of these domains could be, at least partly, a consequence of the adoption of a parasitic way of life.

Against the background of extensive gene/domain loss, the finding of expanded gene families in schistosomes might provide clues to the requirements for a parasitic lifestyle. Among the most expanded gene families in schistosomes (Supplementary Tables 8 and 9), that encoding leishmanolysin (a major surface protease, also called gp63), a member of the metallopeptidase M8 family, has 12 putative family members in *S. japonicum*, but there is only one in human, fruit fly and nematode (*C. elegans*), and only three putative counterparts in the free-living flatworm *Schmidtea*[18] (Supplementary Information). In addition to elastase (see later), leishmanolysin-like proteases may contribute to tissue invasion by schistosome cercariae[19].

Development and metabolism

Cellular signalling pathways in development. To investigate regulatory networks involved in embryonic development and organogenesis, we undertook comparative genomics analysis of well-characterized signalling pathways, including those for Wnt, notch, hedgehog and transforming growth factor β (TGF-β). Notably, the *S. japonicum* genome encodes these growth factors, receptors and essential components to regulate many cellular processes during organogenesis and tissue development (Fig. 3 and Supplementary Tables 10 and 11). *Schistosoma japonicum* also encodes endogenous epidermal growth factor (EGF)-like and fibroblast growth factor (FGF)-like peptides (Fig. 3). The intact downstream cascade composed of the Ras→Raf→mitogen-activated protein kinase (MAPK) and TGF-β→SMAD signalling pathways, including FGF- and EGF-receptors, has components sharing high identity with mammalian orthologues, which implies that schistosomes, in addition to using their own pathways, can exploit host growth factors as developmental signals. Indeed, we have identified an insulin receptor with high sequence similarity with those of mammals[20], whereas no insulin growth factor or insulin molecules were found, further supporting the notion that schistosomes exploit key signalling pathways of their hosts for growth and metabolism.

Metabolic pathways. Analysis of the KEGG pathways assigned to metabolic process (Supplementary Table 12 and Supplementary Figs 7 and 8) indicates that *S. japonicum* can use carbohydrates as energy/carbon sources. It is unable to *de novo* synthesize fatty acids, sterols, purines, nine human essential amino acids, arginine or tyrosine (Supplementary Figs 9–11). Loss or degeneracy of fatty acid, sterol and purine synthesis pathways in schistosomes is probably a consequence of the adoption of a parasitic lifestyle; notably, the genes encoding all the key enzymes for both the *de novo* fatty acid and purine syntheses are complete in the free-living flatworm, *Schmidtea mediterranea*[18] (Supplementary Information). To obtain essential lipid nutrients, the *S. japonicum* genome indeed encodes many transporters, including apolipoproteins, low-density lipoprotein receptor, scavenger receptor, fatty-acid-binding protein, ATP-binding-cassette transporters and cholesterol esterase (Supplementary Table 13), to exploit fatty acids and cholesterol from host blood and plasma.

Table 1 | Summary of *S. japonicum* genomic features in comparison with other organisms

Genome features	Lophotrochozoa	Ecdysozoa				Deuterostomia		
	Schistosoma japonicum	*Caenorhabditis elegans*	*Brugia malayi*	*Drosophila melanogaster*	*Anopheles gambiae*	*Gallus gallus*	*Danio rerio*	*Homo sapiens*
Total genome size (Mb)	398	100	88†	169	265	1,051	1,527	3,255
Total GC content (%)	34.1	35.4	30.5†	40.2	42.2	41.2	34.4	36.2
GC content in coding regions (%)	36	42.7	39.6†	53.3	44.3	42.9	34.4	37.5
GC content in intron regions (%)	33.8	32.1	27.6†	39.5	43.9	41.3	34.1	36.6
GC content in intergenic regions (%)	34.7	35.4	30.9†	40.2	42.2	41.2	34.4	36.2
Repeat rate (%)	40.1	18.3	~15†	24.7	16.6	10.8	46.6	44
Total coding size (Mb)/ratio (%)	15.9/4	24.8/25	12.9/15	21.7/13	14.8/6	25.7/2	47.9/3	35.9/1
Number of coding genes	13,469	20,077	11,515	14,144	12,527	18,107	35,321	25,077
Gene density (genes per Mb)	34	200	130	84	47	17	23	8
Average CDS size (kb)	1.18	1.23	1.12	1.54	1.18	1.42	1.36	1.43
Average gene size ± s.d. (kb)	10.5 ± 16.0	2.8 ± 3.2	2.8 ± 2.9	4.3 ± 10.6	4.6 ± 10.0	21.6 ± 46.8	19.5 ± 35.1	41.3 ± 98.0
Number of transfer RNAs	153	608 (~590*)	~233†	314	450	189	2,010	129
Number of ribosomal RNAs	184	19 (~275*)	~400†	161	N/A	14	N/A	675

All coding sequence (CDS)-related features were calculated from the KEGG database, version 46 (ftp://ftp.genome.jp/pub/kegg/release/archive/kegg/46/). *Caenorhabditis elegans* and *B. malayi* genome feature files were downloaded from Wormbase (ftp://ftp.wormbase.org/pub/). Other genome-feature files and genome sequences were downloaded from Ensembl (ftp://ftp.ensembl.org/pub/release-49).
* The features calculated from the feature file are different from those in the reference[50].
† Data obtained from previous publications[51].

Nervous system and neuroendocrine system

Platyhelminths possess a central nervous system with a variety of sensory structures that can transduce a wide range of stimuli, and use a neuroendocrine system to regulate growth, metabolism and homeostasis.

Neurotransmitters and receptors. We characterized a number of receptors and transporters of neurotransmitters (Supplementary Table 14) that may be required, for example, by miracidia and cercariae to navigate through water to locate new hosts and for the schistosomula and adult flukes to establish and reproduce within the human vasculature. In addition to known neurotransmitters and receptors, we have identified a receptor for octopamine (Supplementary Table 14) and two key enzymes for synthesis of octopamine (Supplementary Fig. 12).

The nervous systems of flatworms can be considered to be predominantly peptidergic[21]. We found additional putative neuropeptide receptors for opioids, galanin and melatonin. Thus, it appears that schistosomes can accept stimulation of the corresponding ligands as a physiological response to different environments, such as fresh water or the tissues of their snail and mammalian hosts. There are genes encoding receptors predicted to accept gastrointestinal neuropeptide hormone signals including cholecystokinin, secretin, gastric inhibitory polypeptide and xenin, all of which are involved in functions promoting the release of alimentary tract fluids containing digestive enzymes.

We also identified receptors for urotensin, angiotensin II and neuromedin (types U and B), which have an important role in physiological regulation of the cardiovascular system, the hypothalamus and other vertebrate organs. Although schistosomes do not have these organs, these components could have other effects on the cells or tissues of the blood fluke, such as the regulation of cell growth or in tissue remodelling. In addition, we found receptors for hypocretin (orexin), leptin and hypothalamic neuropeptides. Together, these features suggest that schistosomes have many advanced physiological features regarded as more characteristic of higher metazoans.

Unexpectedly, a myokinin-like receptor was also observed (Supplementary Table 14). Myokinins are invertebrate neuropeptides with myotropic and diuretic activities for which a receptor, called lymnokinin receptor, was first identified in the tick *Boophilus microplus*[22]. The discovery of such a receptor in schistosomes supports the notion that they might synthesize myokinin because their vertebrate hosts do not produce this neuropeptide. Additional examples of receptors found for other invertebrate neuropeptides included FMRFamide and myosuppressin[23,24], both belonging to the FMRFamide-like peptide superfamily.

Complex sensory system. Schistosomes have a variety of sensory structures using which they, during their different life stages, presumably respond to a myriad of environmental stimuli. Free-living cercariae and miracidia can sense light, mechanical stimuli and temperature[25], facilitating the finding of hosts, whereas the parasitic adult

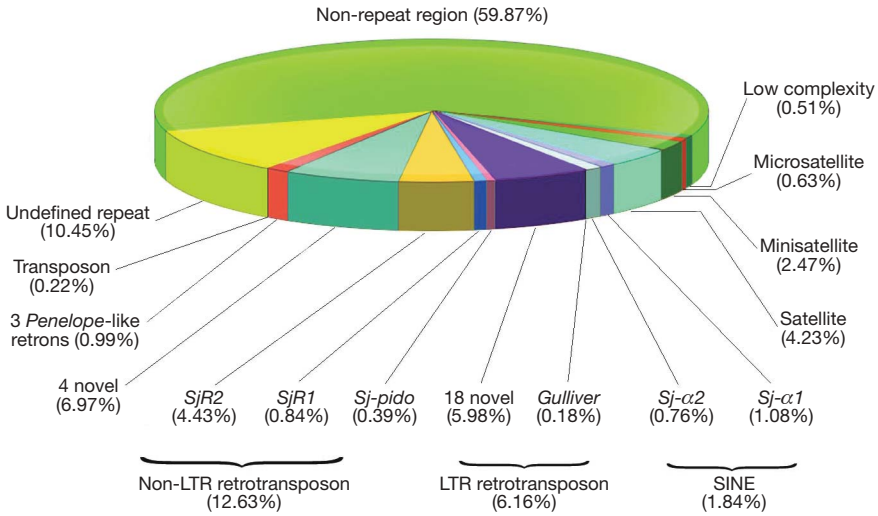


Figure 2 | The distribution of categories and composition of repeat elements in the *S. japonicum* genome. Retrons, retrotransposons; SINE, short interspersed nuclear element.

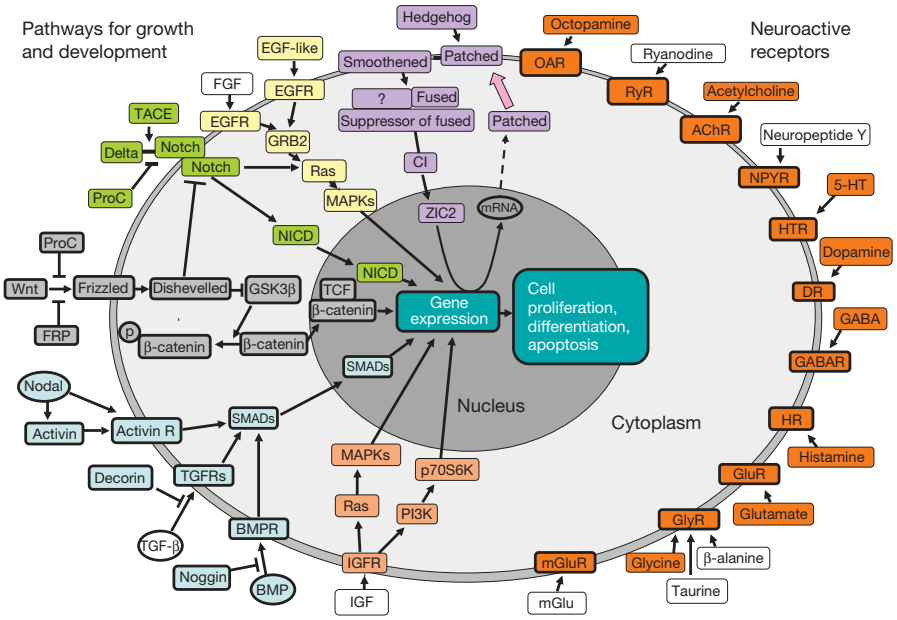


Figure 3 | Putative signalling pathways for growth, development and neuroactive ligand-receptor interaction in *S. japonicum*. The pathways for growth and development (indicated with different colours), and the neuroactive ligand-receptor interactions in *S. japonicum* are shown on the left and right, respectively. TACE, tumour-necrosis-factor-α-converting enzyme; ProC, porcupine homologue (*Drosophila*); NICD, notch intracellular domain; FRP, frizzled-related protein 1; GSK3β, glycogen synthase kinase 3β; TCF, transcription factor 7; 'p' within cycle, phosphorylation on the proteins indicated; BMP, bone morphogenetic protein; IGF, insulin-like growth factor; mGlu, metabotropic glutamate; GlyR, glycine receptor; GluR, glutamate receptor; HR, histamine receptor; GABA, γ-aminobutyric acid; DR, dopamine receptor; 5-HT, 5-hydroxytryptamine; HTR, 5-hydroxytryptamine receptor; NPYR, neuropeptide Y receptor; AChR, acetylcholine receptor; RyR, ryanodine receptor; OAR, octopamine receptor; ZIC2, Zic family member 2; CI, cubitus interruptus; suffix 'R' denotes receptor.

worms are able to respond to changes in levels of chemicals and nutrients. Using a top-down Gene-Ontology-based strategy to facilitate the gene annotation (Supplementary Fig. 13), we identified 71 genes encoding receptors, membrane channels, enzymes and other components, such as rhodopsins/opsins[26], phosrestins/arrestins[27], transducins, cyclic nucleotide-gated channel, rhodopsin kinase and guanylate cyclase 2D (Supplementary Table 15). Both *S. japonicum* and *S. mansoni* have only two members of the rhodopsin family, unlike *Drosophila*, which possesses 13 members, and zebrafish, which has at least seven (Supplementary Fig. 14a). Phylogenetic analysis indicated that there are at least four schistosome transducins, each of which could represent a divergent subtype of transducin superfamily across chordates, echinoderms, molluscs and arthropods (Supplementary Fig. 14b), and could therefore mediate distinct responses of sensors to signals.

The genome sequence analysis also revealed an array of genes encoding sensory proteins that could interact with chemical ligands and other stimuli. These included guanine-nucleotide-binding protein, potassium-voltage-gated-channel protein Shaker, the glutamate receptor for umami taste and protein Prospero[28] (Supplementary Table 16 and Supplementary Fig. 15a). Notably, the genome encodes most components of four of the five human gustatory sensation pathways: the salty, sour, sweet and umami tastes. We also found several potential sensors for sound perception, a common characteristic of vertebrates and arthropods[29], in the genome (Supplementary Table 17).

We discovered an apparently intact olfaction pathway, including cyclic nucleotide-gated olfactory channel, guanine-nucleotide-binding protein and adenylyl cyclase type 3 (Supplementary Table 18 and Supplementary Fig. 15b). Moreover, mechanosensory perception mediated by mechanically gated ion channels represents the basis for the sensing of touch, balance, temperature and sound, and contributes essentially to the development and homeostasis of all Eumetazoa[30] (Supplementary Table 19). Putative sensory components for equilibrium/balance, mechanical stimulation, pain and temperature (Supplementary Tables 20 and 21) were also found in the *S. japonicum* genome, including two proteins that have similarities with the well-known mechanosensory protein, transient receptor potential cation channel[31], and several receptors such as metabotropic glutamate receptor 3, which participate in the sensory perception of pain, light and taste.

Neuroendocrine system. Schistosomes have receptors that apparently evolved to accept endogenous hormones as well as those of the parasitized mammalian host[20,32]. By surveying hormones and receptors related to the classical neuroendocrine axis in the genomic sequence of *S. japonicum*, we found (Fig. 4) putative receptors for hypothalamic hormones such as thyrotropin-releasing hormone (TRH), prolactin-releasing hormone, somatostatin, melanin-concentrating hormone and leptin, as well as transmembrane proteins that have some similarities with receptors for gonadotropin-releasing hormone, corticotropin-releasing hormone and growth-hormone-releasing hormone. Moreover, putative receptors are present that show weak similarity with those in mammals for the pituitary hormones thyroid-stimulating hormone (TSH), luteinizing hormone, follicle-stimulating hormone, arginine vasopressin and oxytocin.

Although a hypothalamus–pituitary-like organ has not been described in schistosomes, it is possible that some neurons, similar to those in the hypothalamus and pituitary of vertebrates, could fulfil similar functions in terms of modulating the behaviour of *S. japonicum* through peripheral endocrine tissues and cells. In this regard, it is noteworthy that the genomic information suggests the presence of an integral hypothalamic–pituitary–thyroid axis in *S. japonicum*. In addition to the superior TRH–TSH receptors, an intact system for synthesis of thyroxine and active triiodothyronine, as well as an inactivation mechanism of these hormones using deiodination, was identified. Nuclear receptors for triiodothyronine and thyroxine were revealed with identity to mammalian orthologues. Hence, *S. japonicum* may use an endogenous thyroid hormone/receptor signalling pathway for growth and development (Fig. 4 and Supplementary Table 22).

We confirmed that *S. japonicum* has receptors for steroid hormones such as progestin, progesterone and oestrogen[32,33]. In addition, it possesses intricate pathways for processing steroid hormones

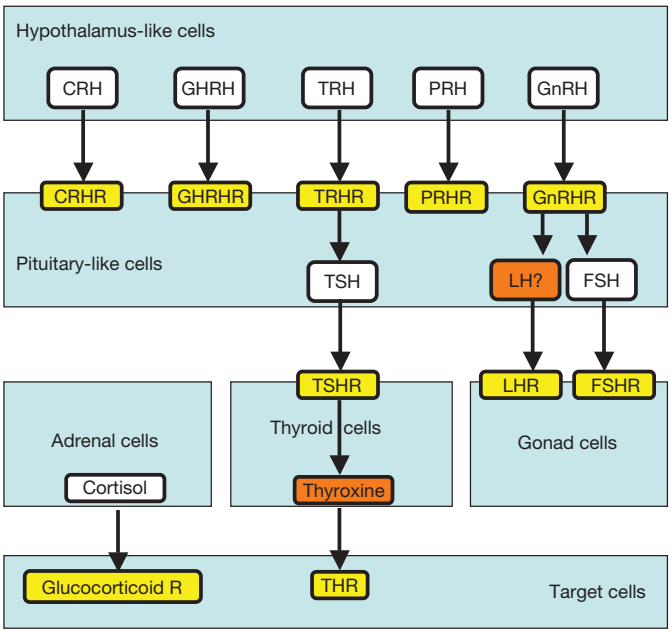


Figure 4 | Putative neuroendocrine system in *S. japonicum*. Structured according to the proposed hypothalamus–pituitary–peripheral-endocrine-glands axis with putative ligands found in *S. japonicum* coloured in orange and *S. japonicum* receptors in yellow. CRH, corticotrophin-releasing hormone; GHRH, growth-hormone-releasing hormone; TRH, thyrotropin-releasing hormone; PRH, prolactin-releasing hormone; GnRH, gonadotropin-releasing hormone; TSH, thyroid-stimulating hormone; FSH, follicle-stimulating hormone; LH, luteinizing hormone; suffix 'R' denotes receptor.

to form other sex hormones. For example, there are putative enzymes present that could convert the female hormones progesterone and pregnenolone to estriol, oestrone, androsterone and testosterone. Hence, schistosomes might use these pathways during their parasitic existence. *Schistosoma japonicum* also encodes enzymes to catabolize excessive or used steroid hormones such as aldosterone.

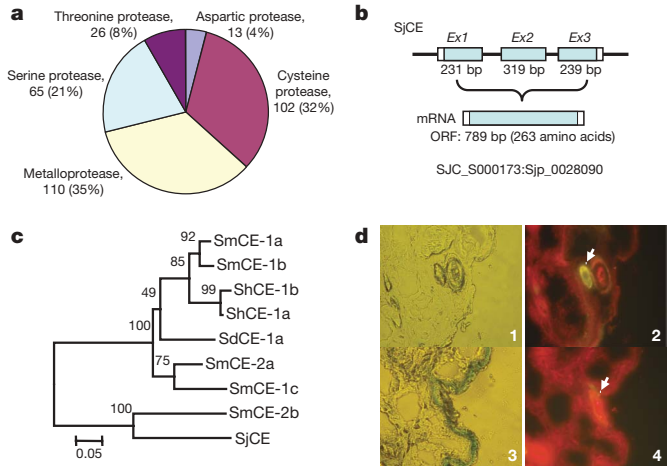


Figure 5 | *S. japonicum* proteases and elastase. **a**, The pie chart shows the distribution of the five kinds of protease. **b**, The genomic structure of *S. japonicum* cercarial elastase (SjCE). **c**, A phylogeny of the elastase family in schistosomes using the neighbour-joining method. Bootstrap values are provided above the branches. SmCE, *S. mansoni* cercarial elastase; ShCE, *S. haematobium* cercarial elastase; SdCE, *Schistosomatium douthitti* elastase. **d**, Immunofluorescence assay showing the presence (white arrow) of SjCE around a schistosomulum following its penetration through mouse skin (panel 2). A naive rabbit serum was used as negative control (panel 4). The location of the cercaria is indicated (white arrow). Panels 1 and 3 show the skin tissue slices under the optical microscope.

With regard to the process of glycolysis for essential energy supply, receptors for adiponectin, an insulin-sensitizing hormone[34], and leptin, a suppressor of the secretion of insulin[35], are also encoded by the genome of *S. japonicum* (Supplementary Table 22), providing further support for the notion that the blood fluke modulates its energy metabolism in response to either its own insulin-like hormones or those of its mammalian host.

The schistosomulum renews its tegument during maturation into an adult schistosome under the effects of ecdysone[36,37]. In concordance, we identified an ecdysone-like receptor and its downstream effector ecdysone-induced protein 78C. In addition, allatostatin, a polypeptide hormone that suppresses the secretion of juvenile hormone, was previously reported to be found throughout the schistosome nervous system[38,39]. An allatostatin-like receptor sequence that has high similarity with that of the cockroach was also identified (Supplementary Table 22).

Disease pathogenesis

Cercarial elastase and protease superfamily. Schistosome proteases have key roles in invasion[40], migration[41] and feeding/nutrition[42]. We identified 314 putative proteases, including metallo-, cysteine, serine, threonine and aspartic proteases, in the *S. japonicum* genome data set (Fig. 5a and Supplementary Tables 23–27) by searching in the MEROPS database of peptidases. We classified 108 *S. japonicum* metalloproteases into 21 subtypes, 16 belonging to the aminopeptidases (Supplementary Table 23). Notably, the leucine aminopeptidase of the M17 family was reported as a major egg antigen[43,44] and a putative anti-fluke vaccine[45]. The second largest assemblage comprised the cysteine proteases, of which 102 members were assigned to 17 subtypes (Supplementary Table 24). Among them, the cathepsins B, C, F and L have pivotal roles in schistosome feeding and nutrition[42], as well as in migration through human tissues[41]. The cysteine proteases cathepsins K and S, as well as the cathepsin A serine protease, have not previously been recognized in schistosomes, and may contribute to catabolism of haemoglobin and other host proteins.

Among the 65 serine proteases (Supplementary Table 25), we discovered a *S. japonicum* cercarial elastase (SjCE), an enzyme that in *S. mansoni* is vital in the penetration by cercariae of mammalian skin to initiate infection[40,46]. The elastase locus predicted from the *S. japonicum* genome spans three exons and two introns, similar to the known *S. mansoni* elastases[47] (Fig. 5b); however, unlike for *S. mansoni*, only a single elastase was identified in *S. japonicum*. Phylogenetic analysis of available schistosome elastases (Supplementary Table 28) suggested that the elastase genes in *S. mansoni* have expanded through at least two rounds of gene duplication, whereas SjCE is an orthologue of *S. mansoni* cercarial elastase 2b (Fig. 5c). Moreover, by re-examination of mass spectra data that we collected previously[33], we identified a unique peptide (IAFLALSDFDHR) of SjCE in cercariae (Supplementary Fig. 16a). We also confirmed the existence of SjCE gene products in both the sporocyst and cercarial stages of *S. japonicum* by immunoblot and immunofluorescence assays (Fig. 5d and Supplementary Fig. 16b). In addition, the native protease was recognized by anti-recombinant SjCE antibodies in infected mouse skin, indicating that this cercarial elastase is secreted/released by the parasite during invasion of mammalian skin.

Immune system and inflammatory factors. The immune system of *S. japonicum* has to face both invading microbial pathogens and the immune statuses of both its molluscan and mammalian hosts. Although adaptive immune molecules such as immunoglobulin are lacking in *S. japonicum* and a classical Toll-like receptor was not found, putative Toll-interacting protein or proteins containing Toll/interleukin-1 resistance motif or leucine-rich repeats appear to be present (Supplementary Table 29). Therefore, schistosomes, like nematodes, appear to possess a primordial Toll pathway as a first line of defence against microbial infections. The identification of the downstream components of a Toll-related pathway, including putative interleukin-1-receptor-associated kinases, toll-like receptor

adaptors, TNF-receptor-associated factor 6 (TRAF6), inhibitor of nuclear factor κB kinase subunit epsilon (IKK-ε) and p38 MAPK, further support the view that this primitive innate immune system could be crucial for the worm (Supplementary Table 30).

On the other hand, factors and metabolites in *S. japonicum* that could contribute to stimulation and regulation of mammalian immunity were discovered. It is well accepted that glycans and lipids synthesized by adult schistosomes or eggs may regulate secondary signals through corresponding receptors on effector cells and accessory cells of the mammalian host, thus compromising host immunological defences targeting the parasite. We therefore searched for enzymes involved in the metabolism of various glycans or lipid antigens by interrogating this worm genome. It turned out that, with the rare exception of enzymes such as α1,3-mannosyltransferase, a complete set of enzymatic machinery for biosynthesis and modification of glycans and lipids exists (Supplementary Table 31).

In addition, prostaglandins, which are well-known mediators of inflammation, can be synthesized by *S. japonicum* as a result of arachidonic-acid metabolism. It is feasible that *S. japonicum* synthesizes arachidonate by using lecithin, converting the arachidonate into leukotriene A4 using arachidonate 5-lipoxygenase, followed by the conversion of unstable leukotriene A4 into the active chemical leukotriene B4 through leukotriene A4 hydrolase. The *S. japonicum* genome also encodes putative receptors for leukotriene B4, cysteinyl leukotriene and prostaglandins E2 and F2, suggesting that prostaglandins could have an important role in the physiology of schistosomes and also in the host–parasite interplay. Unexpectedly, *S. japonicum* possesses proteins paralogous to mammalian autoimmune-disease-related autoantigens; these include 69 kDa islet cell autoantigen (ICA1), islet antigen-2 (PTPRN) and glutamate decarboxylase (GAD), known autoantigens related to type-I diabetes in β-cells, which raises the possibility that these autoantigen-mimicking molecules could induce chemokine-receptor-mediated cell migration and initiate leukocyte migration into inflamed tissue, which ultimately contribute to the granuloma formation that promotes parasite survival.

Concluding remarks

Lophotrochozoa, of which *S. japonicum* is a member, is a large taxon that includes ~50% of all animal phyla. Our work provides a model for evaluating the genomic architecture, biology and evolution in this major taxon. Although the genome of *S. japonicum* has undergone significant protein-domain-loss events, a detailed molecular repertoire exists to permit the pathogen to locate and penetrate hosts, nourish itself and interact with the environment and its host. With the release and analysis of the *S. mansoni* genome[48], a comparative-genomics approach elucidating the similarities and differences between these two closely related parasites will provide more clues regarding these important pathways. Further functional analysis, using approaches such as RNA interference and translational studies are essential to resolve uncertainties in the molecular physiology of schistosomes and to illuminate mechanisms of pathogenesis in schistosomiasis, efforts that may lead to the development of new interventions for its control and eventual elimination.

METHODS SUMMARY

We obtained adult worms and eggs of *S. japonicum* from infected rabbits. The genomic DNA was extracted from ~1,000 mixed, outbred adult male and female *S. japonicum*, perfused from rabbits infected with cercariae released by naturally infected snails. Genomic libraries, including bacterial artificial chromosome (BAC), fosmid and plasmid libraries, were constructed. We performed WGS sequencing on capillary sequencers, and then used a modified PHUSION (version 2.1c) package to assemble the reads. Protein-encoding genes were predicted using EXONHUNTER (version 2.0)[49]. We used a stepwise method to predict the gene functions. The metabolic and regulatory pathway of *S. japonicum* was reconstructed with reference to the KEGG pathway database. Proteins were first clustered using a Markov cluster algorithm and then merged according to protein-domain information to establish protein-family clusters. We used immunoblot and immunofluorescence assays to detect cercarial elastase.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 February; accepted 8 May 2009.

1. Adamson, P. B. Schistosomiasis in antiquity. *Med. Hist.* **20,** 176–188 (1976).
2. Zhou, X. N. *et al.* The public health significance and control of schistosomiasis in China - then and now. *Acta Trop.* **96,** 97–105 (2005).
3. King, C. H., Dickman, K. & Tisch, D. J. Reassessment of the cost of chronic helmintic infection: a meta-analysis of disability-related outcomes in endemic schistosomiasis. *Lancet* **365,** 1561–1569 (2005).
4. Steinmann, P., Keiser, J., Bos, R., Tanner, M. & Utzinger, J. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *Lancet Infect. Dis.* **6,** 411–425 (2006).
5. Finkelstein, J. L., Schleinitz, M. D., Carabin, H. & McGarvey, S. T. Decision-model estimation of the age-specific disability weight for *Schistosomiasis japonica*: a systematic review of the literature. *PLoS Negl. Trop. Dis.* **2,** e158 (2008).
6. Utzinger, J., Zhou, X. N., Chen, M. G. & Bergquist, R. Conquering schistosomiasis in China: the long march. *Acta Trop.* **96,** 69–96 (2005).
7. Li, Y. S. *et al.* Large water management projects and schistosomiasis control, Dongting Lake region, China. *Emerg. Infect. Dis.* **13,** 973–979 (2007).
8. Bergquist, R., Utzinger, J. & McManus, D. P. Trick or treat: the role of vaccines in integrated schistosomiasis control. *PLoS Negl. Trop. Dis.* **2,** e244 (2008).
9. Amiri, P. *et al.* Tumour necrosis factor α restores granulomas and induces parasite egg-laying in schistosome-infected SCID mice. *Nature* **356,** 604–607 (1992).
10. Davies, S. J. *et al.* Modulation of blood fluke development in the liver by hepatic CD4+ lymphocytes. *Science* **294,** 1358–1361 (2001).
11. Hirai, H. *et al.* Chromosomal differentiation of the *Schistosoma japonicum* complex. *Int. J. Parasitol.* **30,** 441–452 (2000).
12. Moné, H. & Boissier, J. Sexual biology of schistosomes. *Adv. Parasitol.* **57,** 89–189 (2004).
13. Halanych, K. M. The new view of animal phylogeny. *Annu. Rev. Ecol. Evol. Syst.* **35,** 229–256 (2004).
14. Price, A. L., Jones, N. C. & Pevzner, P. A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21,** i351–i358 (2005).
15. Arkhipova, I. R., Pyatkov, K. I., Meselson, M. & Evgen'ev, M. B. Retroelements containing introns in diverse invertebrate taxa. *Nature Genet.* **33,** 123–124 (2003).
16. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317,** 86–94 (2007).
17. Gamulin, V., Muller, I. M. & Muller, W. E. Sponge proteins are more similar to those of *Homo sapiens* than to *Caenorhabditis elegans. Biol. J. Linn. Soc.* **71,** 821–828 (2000).
18. Robb, S. M., Ross, E. & Sanchez Alvarado, A. SmedGD: the *Schmidtea mediterranea* genome database. *Nucleic Acids Res.* **36,** D599–D606 (2008).
19. Curwen, R. S., Ashton, P. D., Sundaralingam, S. & Wilson, R. A. Identification of novel proteases and immunomodulators in the secretions of schistosome cercariae that facilitate host entry. *Mol. Cell. Proteomics* **5,** 835–844 (2006).
20. Hu, W. *et al.* Evolutionary and biomedical implications of a *Schistosoma japonicum* complementary DNA resource. *Nature Genet.* **35,** 139–147 (2003).
21. Mousley, A., Maule, A. G., Halton, D. W. & Marks, N. J. Inter-phyla studies on neuropeptides: the potential for broad-spectrum anthelmintic and/or endectocide discovery. *Parasitology* **131,** S143–S167 (2005).
22. Holmes, S. P., Barhoumi, R., Nachman, R. J. & Pietrantonio, P. V. Functional analysis of a G protein-coupled receptor from the southern cattle tick *Boophilus microplus* (Acari: Ixodidae) identifies it as the first arthropod myokinin receptor. *Insect Mol. Biol.* **12,** 27–38 (2003).
23. Egerod, K. *et al.* Molecular cloning and functional expression of the first two specific insect myosuppressin receptors. *Proc. Natl Acad. Sci. USA* **100,** 9808–9813 (2003).
24. Scholler, S. *et al.* Molecular identification of a myosuppressin receptor from the malaria mosquito *Anopheles gambiae. Biochem. Biophys. Res. Commun.* **327,** 29–34 (2005).
25. Cohen, L. M., Neimark, H. & Eveland, L. K. *Schistosoma mansoni*: response of cercariae to a thermal gradient. *J. Parasitol.* **66,** 362–364 (1980).
26. Hoffmann, K. F., Davis, E. M., Fischer, E. R. & Wynn, T. A. The guanine protein coupled receptor rhodopsin is developmentally regulated in the free-living stages of *Schistosoma mansoni. Mol. Biochem. Parasitol.* **112,** 113–123 (2001).
27. Matsumoto, H. & Yamada, T. Phosrestins I and II: arrestin homologs which undergo differential light-induced phosphorylation in the *Drosophila* photoreceptor *in vivo. Biochem. Biophys. Res. Commun.* **177,** 1306–1312 (1991).
28. Grosjean, Y., Lacaille, F., Acebes, A., Clemencet, J. & Ferveur, J. F. Taste, movement, and death: varying effects of new prospero mutants during *Drosophila* development. *J. Neurobiol.* **55,** 1–13 (2003).
29. Robert, D. & Gopfert, M. C. Acoustic sensitivity of fly antennae. *J. Insect Physiol.* **48,** 189–196 (2002).
30. Walker, R. G., Willingham, A. T. & Zuker, C. S. A *Drosophila* mechanosensory transduction channel. *Science* **287,** 2229–2234 (2000).
31. Mutai, H. & Heller, S. Vertebrate and invertebrate TRPV-like mechanoreceptors. *Cell Calcium* **33,** 471–478 (2003).

32. Hu, W., Brindley, P. J., McManus, D. P., Feng, Z. & Han, Z. G. Schistosome transcriptomes: new insights into the parasite and schistosomiasis. *Trends Mol. Med.* **10,** 217–225 (2004).
33. Liu, F. *et al.* New perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. *PLoS Pathog.* **2,** e29 (2006).
34. Heilbronn, L. K., Smith, S. R. & Ravussin, E. The insulin-sensitizing role of the fat derived hormone adiponectin. *Curr. Pharm. Des.* **9,** 1411–1418 (2003).
35. Kieffer, T. J., Heller, R. S., Leech, C. A., Holz, G. G. & Habener, J. F. Leptin suppression of insulin secretion by the activation of ATP-sensitive K+ channels in pancreatic beta-cells. *Diabetes* **46,** 1087–1093 (1997).
36. Foster, J. M., Mercer, J. G. & Rees, H. H. Analysis of ecdysteroids in the trematodes, *Schistosoma mansoni* and *Fasciola hepatica*. *Trop. Med. Parasitol.* **43,** 239–244 (1992).
37. Basch, P. F. Immunocytochemical localization of ecdysteroids in the life history stages of *Schistosoma mansoni*. *Comp. Biochem. Physiol. Comp. Physiol.* **83,** 199–202 (1986).
38. Smart, D. *et al.* Localization of *Diploptera punctata* allatostatin-like immunoreactivity in helminths: an immunocytochemical study. *Parasitology* **110,** 87–96 (1995).
39. Smart, D. *et al.* Peptides related to the *Diploptera punctata* allatostatins in nonarthropod invertebrates: an immunocytochemical survey. *J. Comp. Neurol.* **347,** 426–432 (1994).
40. Dvorak, J. *et al.* Differential use of protease families for invasion by schistosome cercariae. *Biochimie* **90,** 345–358 (2008).
41. Dvorak, J. *et al.* Multiple cathepsin B isoforms in schistosomula of *Trichobilharzia regenti*: identification, characterisation and putative role in migration and nutrition. *Int. J. Parasitol.* **35,** 895–910 (2005).
42. Koehler, J. W., Morales, M. E., Shelby, B. D. & Brindley, P. J. Aspartic protease activities of schistosomes cleave mammalian hemoglobins in a host-specific manner. *Mem. Inst. Oswaldo Cruz* **102,** 83–85 (2007).
43. Abouel-Nour, M. F., Lotfy, M., El-Kady, I., El-Shahat, M. & Doughty, B. L. Localization of leucine aminopeptidase in the *Schistosoma mansoni* eggs and in liver tissue from infected mice. *J. Egypt. Soc. Parasitol.* **35,** 147–156 (2005).
44. Xu, Y. Z., Shawar, S. M. & Dresden, M. H. *Schistosoma mansoni*: purification and characterization of a membrane-associated leucine aminopeptidase. *Exp. Parasitol.* **70,** 124–133 (1990).
45. Hillyer, G. V. Fasciola antigens as vaccines against fascioliasis and schistosomiasis. *J. Helminthol.* **79,** 241–247 (2005).
46. Newport, G. R. *et al.* Cloning of the proteinase that facilitates infection by schistosome parasites. *J. Biol. Chem.* **263,** 13179–13184 (1988).
47. Salter, J. P. *et al.* Cercarial elastase is encoded by a functionally conserved gene family across multiple species of schistosomes. *J. Biol. Chem.* **277,** 24618–24624 (2002).
48. Berriman, M. *et al.* The genome of the blood fluke *Schistosoma mansoni*. *Nature* doi:10.1038/nature08160 (this issue).
49. Brejova, B. *et al.* Finding genes in *Schistosoma japonicum*: annotating novel genomes with help of extrinsic evidence. *Nucleic Acids Res.* **37,** e52 (2009).
50. Stricklin, S. L., Griffiths-Jones, S. & Eddy, S. R. *C. elegans* noncoding RNA genes. *WormBook* **25,** 1–7 (2005).
51. Ghedin, E. *et al.* Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* **317,** 1756–1760 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This investigation was mainly supported by the Chinese National High-Tech Program (863 Program) (2004AA2Z1010, 2006AA02Z335, 2006AA02Z318, 2007AA02Z153), the Chinese National Key Project for Basic Research (973 Project) (2006CB708510, 2007CB513100), the Chinese Academy of Sciences, the Shanghai Municipal Commission for Science and Technology (04DZ14010, 055407031, 06JC14059, 07QA14043, 07DZ22915), and the National Natural Science Foundation of China. Support from the US National Institute of Allergy and Infectious Diseases (award number AI39461), the National Science and Engineering Research Council of Canada (OGP0046506), a International Collaborative Research Grants award from the National Health and Medical Research Council of Australia, and the Wellcome Trust, UK, is also gratefully acknowledged. The Shanghai Supercomputer Center kindly provided computational facilities for some of the data analysis. The authors wish to thank P. De Jong and his colleagues for the BAC libraries construction of *S. japonicum* and N. M. El-Sayed for his contribution on the collaboration between the *S. japonicum* and *S. mansoni* sequencing consortia.

Author Contributions Y. Zhou, H.Z., F.L., W. Hu, Z.-Q.W., G.L. and S.R. contributed equally to this work.

Author Information The sequences of *S. japonicum* WGS assembly contigs and scaffolds, BACs, full-length complementary DNAs and retrotransposons have been deposited in the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database (http://www.ebi.ac.uk/embl/) and the Shanghai Center for Life Science & Biotechnology Information (LSBI; http://lifecenter.sgst.cn/schistosoma/en/schistosomaCnIndexPage.do), and can be freely downloaded. The EMBL accession numbers are CABF01000001–CABF01095265 (contigs) FN330975–FN356022 (scaffolds), FN293020–FN293041 (BACs), FN313573–FN330973 (full-length cDNAs), FN356203–FN356227 (retrotransposons). The LSBI accession numbers are CNUS0000108051–CNUS0000203315 (contigs), CCON0000096785–CCON0000121832 (scaffolds), CNUS0000095394–CNUS0000108050 (predicted genes), CPRT0000000001–CPRT0000012657 (predicted proteins), CNUS0000203316–CNUS0000203337 (BACs), CNUS0000203338–CNUS0000220738 (full-length cDNAs), CNUS0000220739–CNUS0000220763 (retrotransposons). The sequences of *S. japonicum* integrated protein-coding genes are available on the Chinese National Human Genome Center at Shanghai website (http://www.chgc.sh.cn/japonicum). The BAC library (CHORI-108) is available from the laboratory of P. De Jong at the BACPAC Resources Center, Children's Hospital Oakland Research Institute, California (http://bacpac.chori.org/library.php?id=168). Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.W. (wangsy@chgc.sh.cn), Z.-G.H. (hanzg@chgc.sh.cn) or Z. Chen (zchen@stn.sh.cn).

The *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium

Genome annotation and evolution analysis Yan Zhou[1,2], Huajun Zheng[1,2], Yangyi Chen[1], Lei Zhang[1], Kai Wang[1], Jing Guo[1], Zhen Huang[1], Bo Zhang[1], Wei Huang[1], Ke Jin[2], Tonghai Dou[2], Masami Hasegawa[2], Li Wang[2,3], Yuan Zhang[2], Jie Zhou[2], Lin Tao[3], Zhiwei Cao[3], Yixue Li[3], Tomas Vinar[4], Brona Brejova[4], Dan Brown[4], Ming Li[4], David J. Miller[5], David Blair[5], Yang Zhong (Principal Investigator)[2,3], Zhu Chen (Principal Investigator)[1,6]; **Functional genomics analysis** Feng Liu[1,2], Wei Hu[7], Zhi-Qin Wang[1], Qin-Hua Zhang[8], Huai-Dong Song[6], Saijuan Chen[6], Xuenian Xu[7], Bin Xu[7], Chuan Ju[7], Yucheng Huang[7], Paul J. Brindley[9], Donald P. McManus[10], Zheng Feng (Principal Investigator)[7], Ze-Guang Han (Principal Investigator)[1]; **Sequencing and assembly** Gang Lu[1,6], Shuangxi Ren[1], Yuezhu Wang[1], Wenyi Gu[1], Hui Kang[1], Jie Chen[1], Xiaoyun Chen[1], Shuting Chen[1], Lijun Wang[1], Jie Yan [1], Biyun Wang[1], Xinyan Lv[1], Lei Jin[1], Bofei Wang[1], Shiyin Pu[1], Xianglin Zhang[1], Wei Zhang[1], Qiuping Hu[1], Genfeng Zhu[1], Jun Wang[11], Jun Yu[11], Jian Wang[11], Huanming Yang[11], Zemin Ning[12], Matthew Beriman[12], Chia-Lin Wei[13], Yijun Ruan[13], Guoping Zhao (Principal Investigator)[1,2,14], Shengyue Wang (Principal Investigator)[1]; **Paper writing** Feng Liu[1,2], Yan Zhou[1,2], Zhi-Qin Wang[1], Gang Lu[1,6], Huajun Zheng[1,2], Paul J. Brindley[9], Donald P. McManus[10], David Blair[5], Qin-hua Zhang[8], Yang Zhong[2,3], Shengyue Wang[1], Ze-Guang Han[1], Zhu Chen[1,6]; **Project leaders** Shengyue Wang[1], Ze-Guang Han[1], Zhu Chen[1,6]

[1]Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China. [2]School of Life Science/Institutes of Biomedical Sciences, Fudan University, 220 Han Dan Road, Shanghai 200433, China. [3]Shanghai Center for Bioinformation Technology, 100 Qinzhou Road, Shanghai 200235, China. [4]Cheriton School of Computer Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada. [5]Comparative Genomics Centre/School of Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia. [6]State Key Laboratory of Medical Genomics and Shanghai Institute of Hematology, RuiJin Hospital, School of Medicine, Shanghai Jiao Tong University, 197 Rui Jin Road II, Shanghai 200025, China. [7]National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, 207 Rui Jin Er Road, Shanghai 200025, China. [8]Shanghai Center for Biochip Engineering, 151 Li Bing Road, Shanghai 201203, China. [9]Department of Microbiology, Immunology & Tropical Medicine, George Washington University Medical Center, Ross Hall, Room 448, 2300 I Street, NW, Washington DC 20037, USA. [10]Molecular Parasitology Laboratory, Queensland Institute of Medical Research, 300 Herston Road, Brisbane, Queensland 4006, Australia. [11]Beijing Institute of Genomics, Chinese Academy of Sciences/Beijing Genomics Institute, B-6 Beijing Airport Industrial Zone, Beijing 101300, China. [12]Pathogen Sequencing Unit, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. [13]Genome Institute of Singapore, 60 Biopolis Street, Genome #02-01, 138672, Singapore. [14]Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue Yang Road, Shanghai 200031, China.

METHODS

***Schistosoma japonicum* genomic and full-length cDNA library construction.** Genomic DNA was extracted from ~1,000 mixed, outbred adult male and female *S. japonicum*, perfused from rabbits infected with cercariae released by naturally infected snails collected from an endemic focus in Anhui Province, as described[20]. Four genomic libraries with different insert sizes were constructed, one of bacterial artificial chromosomes (inserts, 80–120 kb), one of fosmids (36–42 kb) and two of plasmids (6–10 kb and 1.6–4 kb) (Supplementary Table 1). Total RNAs from *S. japonicum* adults and eggs were isolated using Trizol (Invitrogen), after which mRNA was purified using the Poly(A) Purist mRNA Purification Kit (Ambion). Two full-length cDNA libraries, from adults and eggs were constructed using a modified biotinylated CAP-trapper approach[52,53].

WGS sequencing and assembly. After the clone ends of four discrete genomic libraries were sequenced by capillary DNA sequencers ABI3700 (Applied Biosystems) and MegaBACE 1000 or MegaBACE 4000 (General Electric), PHRED (version 0.020425.c)[54,55] was used for base calling. All reads were qualified by removing clone vector and bacterial host sequences, as well as the host rabbit (*Oryctolagus cuniculus*) DNA sequences (http://www.ensembl.org/Oryctolagus_cuniculus/index.html). A modified PHUSION (version 2.1c) package[56] was used for assembly.

Repeat and retrotransposon identification. A repetitive sequence library of *S. japonicum* was generated by the method of consensus seed extending using REPEATSCOUT (version 1.0.3)[14], with the k-mer size of 16. Tandem repeats in the genome were identified using TANDEM REPEATS FINDER (version 4.00)[57] and categorized using the tandem repeats analysis program TRAP (version 1.0)[58]. Microsatellites, minisatellites and satellites are classically defined as repeat units of 1–6 bp, 11–100 bp and more than 100 bp, respectively. Polyprotein and reverse transcriptase from GenBank were used as queries to search genome sequences of *S. japonicum* using tBLASTN (e-value $\leq 10^{-10}$). The best hit sequences were then used to query the genome, and those yielding multiple hits in the genome were categorized as candidate retrotransposons. All candidate retrotransposons were assembled to establish complete CDSs encoding polyprotein or reverse transcriptase. Once the complete CDS was determined, sequences upstream and downstream of this CDS in the genome were analysed to identify LTRs which flank the left and right termini of LTR retrotransposons and retroviruses.

Prediction and integration of protein-coding genes. Protein encoding genes were predicted using EXONHUNTER (version 2.0)[49]. The prediction program combined *ab initio* gene prediction with supporting evidence from *S. japonicum* and *S. mansoni* expressed sequence tags, *S. japonicum* pair-end ditags, the Swiss-Prot protein database[59] and the Pfam protein-domain database (version 22.0)[60]. Because there were few training sets available for *S. japonicum* or for any other closely related species, we developed an iterative method that started from the distantly related species *C. elegans*, and progressively improved parameters of the gene finder on the basis of well-supported predicted gene fragments. The predicted genes were merged with putative expressed sequence tags and full-length cDNA-derived CDSs (proteins), yielding an integrated protein-coding gene set for further functional analysis. These genes were classified into categories established by the Gene Ontology project through the encoding proteins or domains matched to the Gene Ontology index provided by UniProt[61] and InterPro[62] (iprscan_DATA_17.0 and iprscan_PTHR_DATA_14.0).

Genome variation analysis. The PHUSION assembler[56] does not provide alignment information of reads to its contig consensus, so BLASTN was used to relocate reads to contig consensus, with overall identity of over 95%, and to provide alignment information. We established a locally developed SNP pipeline based on neighbourhood quality standard, with the following rules: for each candidate SNP on shotgun reads, the 5-bp flanking sequences should be the same as the contig consensus, the base quality on the SNP site should be no less than 23 and the base quality of the flanking 5 bp should be not less than 15 (refs 63, 64).

Pathway mapping. The metabolic and regulatory pathway of *S. japonicum* was reconstructed on the basis of the KEGG pathway database[65]. The KEGG orthology identifier was used as a linkage between genes and pathways. The assignment of *S. japonicum* genes to KEGG orthologues was implemented with a modified bidirectional-best-BLAST-hits method, which was adjusted using phylogenetic information. The pathway mapping results for the *S. japonicum* genome are available at http://chgc.sh.cn/japonicum.

Gene-family analysis. Proteins of *S. japonicum*, *C. elegans*, *D. melanogaster*, *A. gambiae*, *D. rerio*, *G. gallus*, *H. sapiens* and *N. vectensis* were first clustered using a Markov cluster algorithm[66] and then merged according to protein-domain information to establish protein-family clusters. The *S. japonicum* protein domains were scanned using INTERPROSCAN[62]. Protein-domain information on other species was sourced from the KEGG database[65].

Analysis of *S. japonicum* proteases. Putative proteases in the *S. japonicum* data set were identified by comparing *S. japonicum* cDNA and predicted genes with the MEROPS database[67]. The results were manually checked and compared with annotations generated by BLAST searches against more comprehensive databases as above. Results with inconsistent annotations from MEROPS and BLAST were removed. For phylogenetic and evolutionary analyses of gene families, deduced amino-acid sequences were aligned using CLUSTAL W (version 1.83)[68]. Phylogenetic trees were generated using MEGA (version 3.1)[69] with the neighbour-joining method and tested with 1,000 bootstrap replicates.

Immunofluorescence assay of *S. japonicum* cercarial elastase. A mouse anaesthetized with pentobarbital was infected with *S. japonicum* cercariae. After 10 min, the skin was excised, finely diced, and embedded in OCT fixative. The prepared 7-μm-thick frozen sections were incubated for 30 min in a solution of 20% goat serum in Tris-HCl-buffered saline. The sections were incubated with the rabbit primary antiserum raised against purified recombinant SjCE or normal rabbit serum, followed by a FITC-conjugated second antibody. Fluorescence was visualized using a Leica DM-2500 fluorescence microscope.

52. Seki, M., Carninci, P., Nishiyama, Y., Hayashizaki, Y. & Shinozaki, K. High-efficiency cloning of Arabidopsis full-length cDNA by biotinylated CAP trapper. *Plant J.* **15,** 707–720 (1998).
53. Wei, C. L. *et al.* 5′ long serial analysis of gene expression (LongSAGE) and 3′ LongSAGE for transcriptome characterization and genome annotation. *Proc. Natl Acad. Sci. USA* **101,** 11701–11706 (2004).
54. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8,** 175–185 (1998).
55. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8,** 186–194 (1998).
56. Mullikin, J. C. & Ning, Z. The phusion assembler. *Genome Res.* **13,** 81–90 (2003).
57. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27,** 573–580 (1999).
58. Sobreira, T. J., Durham, A. M. & Gruber, A. TRAP: automated classification, quantification and annotation of tandemly repeated sequences. *Bioinformatics* **22,** 361–362 (2006).
59. Gasteiger, E. *et al.* ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31,** 3784–3788 (2003).
60. Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **36,** D281–D288 (2008).
61. UniProt Consortium. The universal protein resource (UniProt). *Nucleic Acids Res.* **36,** D190–D195 (2008).
62. Zdobnov, E. M. & Apweiler, R. InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17,** 847–848 (2001).
63. Mullikin, J. C. *et al.* An SNP map of human chromosome 22. *Nature* **407,** 516–520 (2000).
64. Altshuler, D. *et al.* An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407,** 513–516 (2000).
65. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36,** D480–D484 (2008).
66. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30,** 1575–1584 (2002).
67. Rawlings, N. D., Morton, F. R. & Barrett, A. J. MEROPS: the peptidase database. *Nucleic Acids Res.* **34,** D270–D272 (2006).
68. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680 (1994).
69. Kumar, S., Tamura, K. & Nei, M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5,** 150–163 (2004).

ARTICLES

The genome of the blood fluke *Schistosoma mansoni*

Matthew Berriman[1], Brian J. Haas[3]†, Philip T. LoVerde[4], R. Alan Wilson[5], Gary P. Dillon[5], Gustavo C. Cerqueira[6,7,8], Susan T. Mashiyama[9,10], Bissan Al-Lazikani[11], Luiza F. Andrade[12], Peter D. Ashton[4], Martin A. Aslett[1], Daniella C. Bartholomeu[3]†, Gaelle Blandin[3], Conor R. Caffrey[9], Avril Coghlan[13], Richard Coulson[2], Tim A. Day[14], Art Delcher[7], Ricardo DeMarco[5,15,16], Appolinaire Djikeng[3], Tina Eyre[1], John A. Gamble[1], Elodie Ghedin[3]†, Yong Gu[1], Christiane Hertz-Fowler[1], Hirohisha Hirai[17], Yuriko Hirai[17], Robin Houston[1], Alasdair Ivens[1]†, David A. Johnston[18]†, Daniela Lacerda[3]†, Camila D. Macedo[6,8], Paul McVeigh[14], Zemin Ning[1], Guilherme Oliveira[12], John P. Overington[2], Julian Parkhill[1], Mihaela Pertea[7], Raymond J. Pierce[19], Anna V. Protasio[1], Michael A. Quail[1], Marie-Adèle Rajandream[1], Jane Rogers[1]†, Mohammed Sajid[9]†, Steven L. Salzberg[7,8], Mario Stanke[20], Adrian R. Tivey[1], Owen White[3]†, David L. Williams[21]†, Jennifer Wortman[3]†, Wenjie Wu[4]†, Mostafa Zamanian[14], Adhemar Zerlotini[11], Claire M. Fraser-Liggett[3]†, Barclay G. Barrell[1] & Najib M. El-Sayed[3,6,7,8]

***Schistosoma mansoni* is responsible for the neglected tropical disease schistosomiasis that affects 210 million people in 76 countries. Here we present analysis of the 363 megabase nuclear genome of the blood fluke. It encodes at least 11,809 genes, with an unusual intron size distribution, and new families of micro-exon genes that undergo frequent alternative splicing. As the first sequenced flatworm, and a representative of the Lophotrochozoa, it offers insights into early events in the evolution of the animals, including the development of a body pattern with bilateral symmetry, and the development of tissues into organs. Our analysis has been informed by the need to find new drug targets. The deficits in lipid metabolism that make schistosomes dependent on the host are revealed, and the identification of membrane receptors, ion channels and more than 300 proteases provide new insights into the biology of the life cycle and new targets. Bioinformatics approaches have identified metabolic chokepoints, and a chemogenomic screen has pinpointed schistosome proteins for which existing drugs may be active. The information generated provides an invaluable resource for the research community to develop much needed new control tools for the treatment and eradication of this important and neglected disease.**

Schistosomiasis is a neglected tropical disease that ranks with malaria and tuberculosis as a major source of morbidity affecting approximately 210 million people in 76 countries, despite strenuous control efforts[1]. It is caused by blood flukes of the genus *Schistosoma* (phylum Platyhelminthes), which exhibit dioecy and have complex life cycles comprising several morphologically distinct phenotypes in definitive human and intermediate snail hosts. *Schistosoma mansoni*, one of the three major human species, occurs across much of sub-Saharan Africa, parts of the Middle East, Brazil, Venezuela and some West Indian islands. The mature flukes dwell in the human portal vasculature, depositing eggs in the intestinal wall that either pass to the gut lumen and are voided in the faeces, or travel to the liver where they trigger immune-mediated granuloma formation and peri-portal fibrosis[2]. Approximately 280,000 deaths per annum are attributable to schistosomiasis in sub-Saharan Africa alone[3]. However, the disease is better known for its chronicity and debilitating morbidity[4]. A single drug, praziquantel, is almost exclusively used to treat the infection but this does not prevent reinfection, and with the large-scale control programmes in place, there is concern about the development of drug resistance. Indeed, resistance can be selected for in the laboratory and there are reports of increased drug tolerance in the field[5].

[1]Wellcome Trust Sanger Institute, [2]European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [3]The Institute for Genomic Research/The J. Craig Venter Institute, 9712 Medical Center Drive, Rockville, Maryland 20850, USA. [4]Departments of Biochemistry and Pathology, Mail Code 7760, University of Texas, Health Science Center, San Antonio, Texas 78229-3900, USA. [5]Department of Biology, University of York, PO Box 373, York YO10 5YW, UK. [6]Department of Cell Biology and Molecular Genetics, [7]Center for Bioinformatics and Computational Biology, and [8]Maryland Pathogen Research Institute, University of Maryland, College Park, Maryland 20742, USA. [9]Sandler Center for Basic Research in Parasitic Diseases, [10]Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, California Institute for Quantitative Biomedical Research (QB3), Byers Hall, 1700 4th Street, University of California, San Francisco, California 94158-2330, USA. [11]Cancer Research UK Centre for Cancer Therapeutics, The Institute of Cancer Research, Haddow Laboratories, 15 Cotswold Road, Belmont, Sutton, Surrey SM2 5NG, UK. [12]Centro de Pesquisas René Rachou (CPqRR)—FIOCRUZ, Av Augusto de Lima 1715, Belo Horizonte, MG 30190002, Brazil. [13]Department of Microbiology, University College Cork, Western Road, Cork, Ireland. [14]Department of Biomedical Sciences, Iowa State University, Ames, Iowa 50011, USA. [15]Instituto de Química, [16]Instituto de Física de São Carlos, Universidade de São Paulo, Brazil. [17]Primate Research Institute, Kyoto University, Inuyama, Aichi 484–8506, Japan. [18]Biomedical Parasitology Division, The Natural History Museum, London SW7 5BD, UK. [19]Inserm, U 547, Université Lille 2, Institut Pasteur de Lille, IFR 142, Lille, France. [20]Institut für Mikrobiologie und Genetik, Abteilung Bioinformatik, Universität Göttingen, Goldschmidtstraße 1, Göttingen 37077, Germany. [21]Department of Biological Sciences, Illinois State University, Normal, Illinois 61790-4120, USA. †Present addresses: The Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA (B.J.H.); Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil (D.C.B. and D.L.); Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA (E.G.); Fios Genomics Ltd, ETTC, King's Buildings, Edinburgh EH9 3JL, UK (A.I.); Biomedical Imaging Unit, School of Medicine, University of Southampton, Southampton SO16 6YD, UK (D.A.J.); John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK (J.R.); Leiden University Medical Centre, Parasitologie, Albinusdreef, 2333 ZA Leiden, The Netherlands (M.S.); Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA (O.W., J.W. and C.M.F.-L.); Immunology/Microbiology, Rush University Medical Center, 1735 West Harrison Street, Chicago, Illinois 60612-3824, USA (D.L.W.); Department of Biochemistry, School of Medicine and Biomedical Research, State University of New York at Buffalo, Buffalo, New York 14214, USA (W.W.); Developmental Genomics Group, New York State Center of Excellence in Bioinformatics and Life Sciences, 701 Ellicott Street, Buffalo, New York 14203, USA (W.W.).

In this study we present the sequence and analysis of the *S. mansoni* genome. Previous metazoan projects have been restricted to Deuterostomia (for example, *Homo*, *Mus* and *Ciona*) and the ecdysozoan clade of the Protostomia (for example, *Drosophila*, *Caenorhabditis* and *Brugia*). Together with the accompanying article on *S. japonicum*[6], we present, to our knowledge, the first descriptions of metazoan genomes from the lophotrochozoan clade. The genome reveals features that aid our understanding of the evolution of complex body plans. We have mined the genome to predict new drug targets, on the basis of searches involving traditional areas for drug discovery, metabolic reconstruction, and bioinformatics screens that exploit shared pharmacology. It is hoped that these and other targets will accelerate drug discovery, generating the much needed new treatments for the control and eradication of schistosomiasis.

Genome structure and content

The nuclear genome sequence of *S. mansoni* was determined by whole-genome shotgun sequencing and assembled into 5,745 scaffolds greater than 2 kilobases (kb) (Supplementary Table 1), totalling 363 megabases (Mb). Although 40% of the genome is repetitive, 50% is assembled into scaffolds of at least 824.5 kb. Furthermore, 43% of the genome assembly (distributed over 153 scaffolds) was unambiguously assigned to chromosomes (seven autosomal, plus ZW sex-determination pairs) using fluorescence *in situ* hybridization (FISH; Fig. 1, Supplementary Fig. 1 and Supplementary Table 2).

We identified 72 families of both long-terminal repeat (LTR) and non-LTR transposons, comprising 15% and 5% of the genome, respectively, and containing 63 and 60 new families each (Supplementary Table 3). The LTR transposons are from the Ty3/*Gypsy* and BEL clades, whereas the non-LTR transposons are restricted to the RTE, CR1 and R2 clades. Two previously described non-LTR retrotransposon families from the RTE clade (SR2 and Perere-3)[7,8] seem to have undergone a burst of transposition events after divergence of *S. mansoni* and *S. japonicum*, and contribute to an overall higher representation of non-LTR retrotransposons in *S. mansoni* (15%, around 8% in *S. japonicum*). A new DNA transposon belonging to the Mu family was also found, which represents the first instance in a flatworm. The presence of target site duplications in some copies indicates recent transposition, and suggests that active copies may still exist in the genome. A lack of terminal inverted repeats, a feature of Mu family members, suggests a peculiar mechanism for recognition of this element by the transposition apparatus.

We identified 11,809 putative genes encoding 13,197 transcripts. Considering genes that do not span a gap, the average gene size is 4.7 kb, typically with large introns (the average is 1,692 base pairs (bp)) and much smaller exons (the average is 217 bp). Moreover, the introns show a markedly skewed size distribution that has not been observed in other eukaryotes, whereby 5′ introns are smaller than 3′ introns (Fig. 2, Supplementary Information and Supplementary Table 5). In multi-exon genes, the first few introns can be as small as 26 bp, whereas introns towards the 3′ end are typically kilobases in length (the largest is 33.8 kb). The reason for this is unclear but it suggests unusual transcriptional control. However, a survey of conserved transcription factor domains shows *S. mansoni* to be

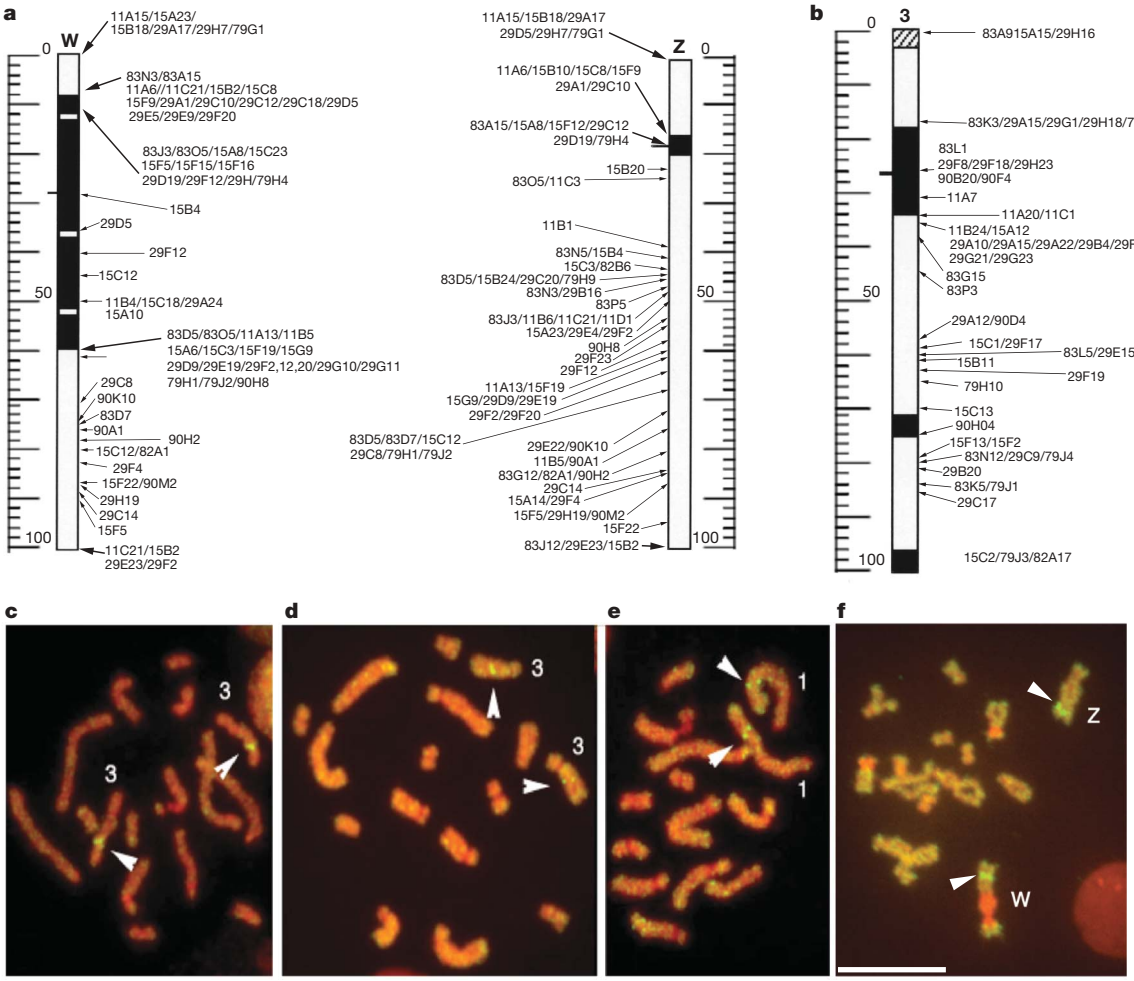


Figure 1 | Physical map of *S. mansoni*. **a**, **b**, Idiogram of *S. mansoni* chromosomes W, Z (**a**) and 3 (**b**). *S. mansoni* BAC clones were mapped to the karyotype of *S. mansoni* by FISH. The solid black areas are heterochromatin, the open areas are euchromatin. The BAC clones are identified by BAC numbers. **c**–**f**, Chromosome spreads with FISH-mapped BACS are shown. FISH-mapped BACS are identified by arrowheads on labelled chromosomes. Scale bar, 10 μm. See Supplementary Fig. 1 for idiograms of all *S. mansoni* chromosomes.

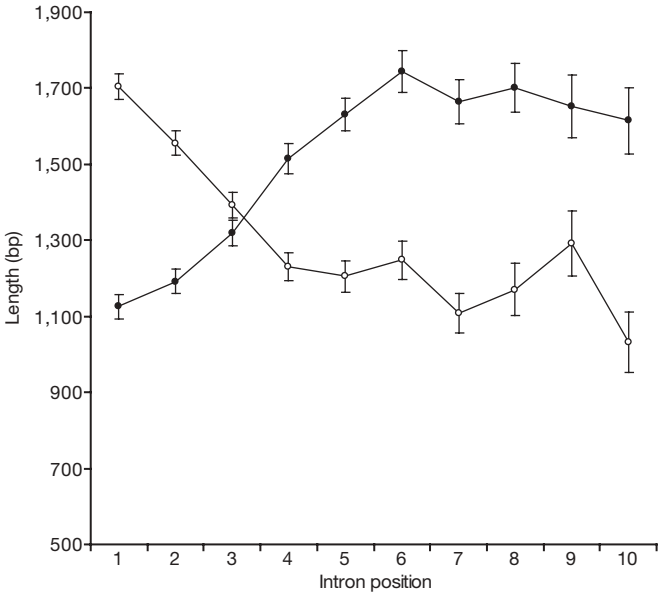


Figure 2 | Intron size distribution. The length of introns varies according to their position in a transcript, counting from the 5′ end (solid circles) and the 3′ end (open circles). Mean lengths ± standard errors are shown. After about five introns, the length difference is no longer apparent owing to the variation in the number of introns per transcript (see Supplementary Information).

broadly similar to other eukaryotes (Supplementary Information, Supplementary Fig. 2 and Supplementary Table 6). It is noteworthy that 43% of transcription factor families with schistosome representatives also contained vertebrate sequences, nearly twice the number that matched nematode worms, emphasizing their evolutionary distance.

Micro-exon genes

At least 45 genes have an unusual micro-exon structure. Individual micro-exons have been described in other genomes, dispersed among several normal exons[9]. However, *S. mansoni* is notable in containing micro-exon genes (MEGs) that comprise 75% of the coding sequence, are flanked at the 5′ and 3′ extremes by conventional exons, and have lengths that are multiples of three bases (from 6 to 36).

Other than having shared gene structure, no similarity could be detected between 14 MEG families (each with up to 23 members; Fig. 3 and Supplementary Table 7). Moreover, they showed no similarity with annotated genes from outside *Schistosoma* spp., nor any identifiable motifs or functional domains. Comparisons between MEG family members and related proteins from *S. japonicum* suggest that some gene duplication events preceded the divergence of the two species. Almost all encode a signal peptide at the 5′ end and three have membrane anchors, so most are probably secreted. Examination of the large expressed-sequence tag (EST) data set from across the life cycle shows that genes from all MEG families are transcribed in the intramammalian stages of the life cycle, and the germ balls of daughter sporocysts that develop into infective cercariae, but probably not in miracidia that infect the snail intermediate host (Fig. 3).

Sequencing of transcripts from three MEG families revealed the occurrence of several alternative splice variants formed by exon skipping. In one of the families analysed, all internal exons except those coding for the signal peptide were missing in at least one transcript sampled, and a gene from a second family presented different

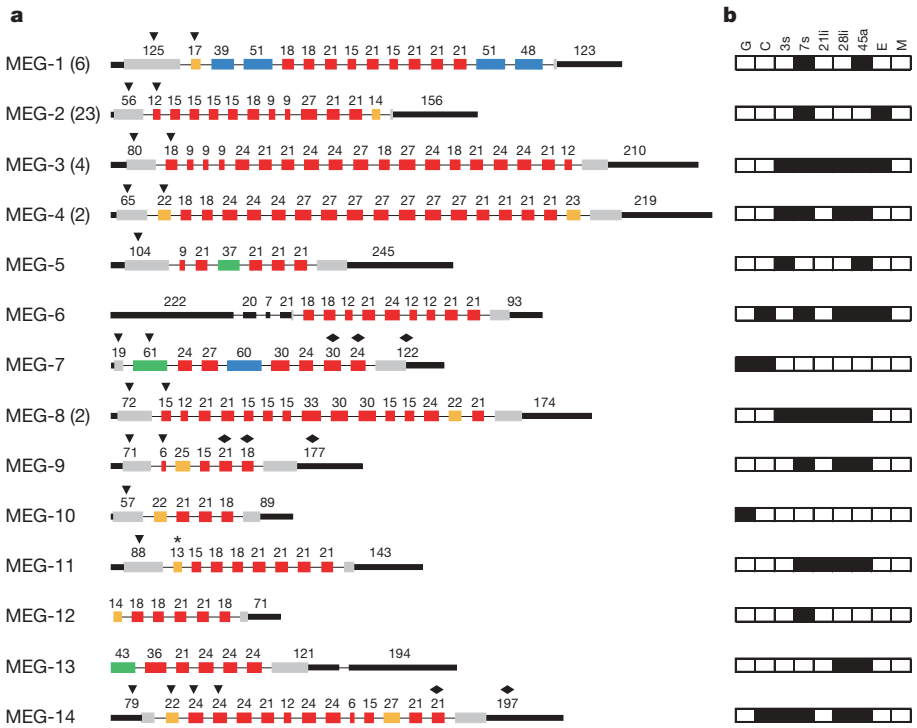


Figure 3 | Schematic representation of gene structure from MEG family members. **a**, Structure of a representative member from each MEG family. Where several members were found, the total number detected is indicated in parentheses. Each box represents an exon drawn to scale, and the number above it indicates the exon size in nucleotides. For illustrative purposes, the introns are shown with fixed length. Black triangles and diamonds indicate exons encoding predicted signal peptides and transmembrane helices, respectively. Other characteristics associated with exons are indicated by colour and grouped as follow: micro-exons having lengths that either are multiples of 3 bp (red) or are indivisible by 3 bp (orange); exons longer than 36 bp and having lengths that either are multiples of 3 bp (blue) or are indivisible by 3 bp (green); putative initiation and termination exons (grey); untranslated region (UTR) (black). The asterisk indicates an exon deduced from transcript data, which did not match the sequenced genome. MEG-12 and MEG-13 structures were only partially predicted owing to the lack of transcripts containing the 5′ end of these genes. **b**, PCR with reverse transcription (RT–PCR) or EST-based evidence of transcription (black box) for each family across different life-cycle stages. C, cercaria; E, egg; G, germball; M, miracidium; 3s and 7s, 3- and 7-day schistosomula; 21li and 28li, 21- and 28-day liver worms; 45a, 45-day adult worm pairs.

transcripts with extended exons produced by the use of alternative splicing sites. These observations suggest that a 'pick and mix' strategy is used to create protein variation

Evolution of triploblasty, parasitism and tissues

Schistosomes are the first Platyhelminthes to be fully sequenced, and provide insights into the evolution of 'simple' animals. Using Treefam to make comparisons with the sea anemone *Nematostella vectensis*, a representative of the Radiata, we sought gene families restricted to, or expanded in, the Bilateria (Supplementary Table 8). The advent of a third germ layer in flatworms is paralleled by the expansion of genes encoding cell adhesion molecules such as cadherins. Similarly, tissue-patterning developmental cues (for example, Notch/Delta) and histone-modifying enzymes (for example, histone acetyltransferases) have proliferated. Some genes, such as the tetraspanins that encode membrane structural proteins, have greatly proliferated in schistosomes, suggesting a critical role in worm physiology/parasitism. The large array of paralogues for fucosyl and xylosyltransferases involved in the generation of new glycans expressed at the host–parasite interface may be important for subverting the immune system. The expansion of proteases in schistosomes also seems to be directly related to parasitism, because it includes families involved in host invasion (invadolysins) and blood feeding (cathepsins). Furthermore, G-protein-coupled receptors (GPCRs) show varying levels of contraction in schistosomes, whereas several classes (for example, peropsins) are greatly expanded in *Nematostella*, indicating functions associated with the free-living lifestyle.

Although schistosomes are acoelomate, they possess tissues approaching the sophistication of organs—such as gut, nephridia, nerve and muscle—that are concerned with discrete physiological processes, such as feeding, excretion and locomotion. However, as lophotrochozoans they are evolutionarily distant from the previously sequenced parasitic nematodes *Brugia*[10] and *Meloidogyne*[11,12] (both ecdysozoans). Compartmentalisation of schistosome tissues and the formation of epithelial barriers are crucial for life in the hostile environment of the host bloodstream. Schistosomes possess the typical machinery of higher metazoa to interact with the cytoskeleton and control cell polarity (Supplementary Information and Supplementary Table 9), organize epithelia and denote tissue boundary lines.

S. mansoni possess a nervous system that includes an anterior brain and longitudinal nerve cords, which extend from the brain to run the length of the worm body. Furthermore, a variety of sensory structures (at least six types in the cercaria[13]) are able to transduce a wide range of stimuli that assist in host location, penetration and navigation through the vasculature. In common with more complex organisms, schistosomes possess the tools needed to mediate neurogenesis and control axon growth cones and migration of neural cells (Supplementary Information and Supplementary Table 9), supporting the ancient origins of neural complexity.

Insights into possible new drug targets

Historically, anti-schistosomiasis agents were identified by *in vivo* screening in animal models. The *S. mansoni* genome project makes a more target-based approach to drug discovery feasible, and some promising leads have already emerged. These include a family of nuclear receptors[14] (Supplementary Information) and a redox enzyme, thioredoxin glutathione reductase, recently validated as a drug target[15]. The condensed redox biochemistry of *S. mansoni*, relative to its human host, may offer further drug development targets (Supplementary Information). In the context of drug discovery, we have explored other potential areas of vulnerability, including: lipid metabolism, GPCRs, ligand- and voltage-gated ion channels, kinases, proteases and neuropeptides. We also undertook two bioinformatics-led approaches: metabolic reconstruction to identify chokepoints, and sequence searches for structures related to known drug targets.

Lipid metabolism

S. mansoni contains a full complement of genes required for most core metabolic processes, such as glycolysis, tricarboxylic acid cycle and the pentose phosphate pathway. However, schistosomes are incapable of *de novo* synthesis of sterols or free fatty acids and must use complex precursors from the host[16]. An extensive lipid-carrying protein repertoire could be identified, but despite producing precursors for fatty acid synthesis, fatty acid synthase could not be identified. An inability to use isoprene products of the mevalonate pathway probably accounts for the lack of sterol biosynthesis (Supplementary Table 11 and Supplementary Information). The genes necessary for a complete β-oxidation pathway are present, and this usually inactive pathway might operate in reverse to perform syntheses[17]. Despite constituting 40% or more of the lipid content of adult worms[16], triacylglycerol has an uncertain role in the schistosome's life cycle—it is slow to turn over, does not contribute to the formation of other lipids[16] and its use as an energy store is doubtful[17]. Nevertheless, *S. mansoni* possesses lipases capable of breaking down triacylglycerol, so it may have functions other than preventing too high concentrations of intracellular fatty acids[16]. Pathways responsible for synthesizing the phospholipid components of membranes are well represented, except that phosphatidylcholine must be derived from diacylglycerol[18] and the parasite must depend on its host as a source of inositol.

GPCRs, ligand-gated and voltage-gated ion channels

GPCRs, ligand-gated and voltage-gated ion channels are targets for 50% of all current pharmaceuticals[19]. At least 92 putative GPCR-encoding genes are present (Supplementary Table 12), the bulk (82) of which are from the rhodopsin family. The largest groups are the α-subfamily (30), which includes amine receptors, and the β-subfamily (24), which contains neuropeptide and hormone receptors. The diversity of the former subfamily underlines the wide range of potential amine/neurotransmitter reactivities of schistosomes, but the tentative identities assigned need to be confirmed by functional studies, as has already been performed for a histamine receptor[20]. Schistosomes detect chemosensory cues, but a large, unique clade of the mediating receptors was not found. However, the 26 'orphan' rhodopsin family GPCRs may include proteins with this role. Outside the large rhodopsin family, representatives from each of the smaller families of GPCRs, glutamate family (2), frizzled family (3), and the secretin/adhesion family (4) are present.

Each of the three major ligand-gated ion channel families—the Cys-loop family, glutamate-activated cation channels, and ATP-gated ion channels—are represented in the schistosome genome. Of the 13 Cys-loop family ligand-gated ion channels, nine encode nicotinic acetylcholine receptor subunits (Supplementary Fig. 4 and Supplementary Table 13). The remaining four anion channel subunits group among GABA (γ-aminobutyric acid), glycine and glutamate receptors, but it is not possible to assign precise identities. The seven schistosome glutamate-activated cation channels comprise at least two sequences from each of the three common subgroupings. The presence of a functional P2X receptor for ATP-mediated signalling in schistosomes was already known[21], and the data here show at least four more.

Voltage-gated ion channels generate and control membrane potential in excitable cells, and are central to ionic homeostasis. There are examples of successful drugs targeting voltage-gated sodium, potassium and calcium channels[22]. Although voltage-gated sodium channels were not found, at least 41 members from each of the major six transmembrane (6TM) and four transmembrane (4TM) families of potassium channels (Supplementary Table 14) are present. The 6TM voltage-gated potassium channel family (20 members) is the largest, including the well-characterized Kv1.1 channel found in nerve and muscle of adult schsitosomes[23]. Other classes of 6TM potassium channels include the KQT channels, large calcium-activated channels, small calcium-activated channels, and cyclic-nucleotide-gated groups. This last group, comprising eight members, is most often associated

with signal transduction in primary olfactory and visual sensory cells (*Caenorhabditis elegans* has only five; ref. 24). *S. mansoni* possesses six 4TM inward-rectifying TWIK-related potassium channels (about 46 in *C. elegans*). There are four α and two β subunits of voltage-gated calcium channels in schistosomes, and a β subunit is implicated as a molecular target of the anti-schistosomal praziquantel[25].

The kinome

Protein kinases are important regulators of many different cellular functions. Both they and their inhibitors have entered the drug development pipeline in recent years[26] but few schistosome kinases have been characterized to date. The *S. mansoni* genome encodes 249 kinases, including 22 genes with alternative splicing (Supplementary Information). This corresponds to 1.9% of the total coding proteins in the genome, a figure comparable to that found in other species[27] (Supplementary Fig. 6). *S. mansoni* possesses representatives of all of the main kinase groups (Supplementary Fig. 7), the largest of which is the CMGC (cyclin-dependent kinases, mitogen-activated protein kinases, glycogen synthase kinase 3 and CK2-related kinases) group, in contrast to other analysed eukaryotic genomes. However, a single class (RCK) is absent from the CMGC family, a deficiency shared with yeast but not nematodes or mammals.

The least represented groups are the casein kinase (CK1) and receptor guanylate cyclase families with only seven and three members, respectively, contrasting with *C. elegans*, in which casein kinase is the largest group and receptor guanylate cyclase has 27 members. CK1 (and CMGC) group members that are expressed in sperm or during spermatogenesis in *C. elegans* are missing in *S. mansoni*.

The degradome

Proteolytic enzymes (proteases), making up an organism's 'degradome'[28], operate in virtually every biological and pathological phenomenon[29] and are proven drug targets in diverse biomedical contexts[30,31]. All five major classes of proteases (aspartic, cysteine, metallo-, serine and threonine) are represented as various clans (mechanistically related groups) in the parasite genome (Supplementary Table 17). The percentage distribution of the major clans is generally similar to that of the human host with some notable exceptions, mainly owing to the expansion of constituent protease families in humans. Of the 73 protease families, 61 are found in humans and in *S. mansoni*, and 60 families are shared. With 335 sequences, proteases comprise 2.5% of the putative proteome (Supplementary Table 18), consistent with the proportion in other organisms (1–5%), but this is only one-third of that in humans (945 sequences, if A2 family retrovirus and retrotransposon proteases are included).

The greatest difference between host and parasite is in the paucity of chymotrypsin-like S1 family enzymes in the latter (22 versus 135 human sequences). This reflects the evolution and diversification of family S1 for complex and highly regulated proteolysis cascades in vertebrates and some invertebrates, such as innate immunity, development, blood coagulation and complement activation[32–34]. From a therapeutic standpoint, the reduced complexity may prove valuable with fewer parasite proteases available for essential life-sustaining functions. For example, robust drug discovery programmes are in place for chymotrypsin-like S1 families[35] and peptidase C14 (caspases)[36], on which anti-schistosomal drug discovery could 'piggy-back'[37]. It is also notable that a smaller number of schistosome protease families (for example, C1, M8 and M13) have more members than the respective families in humans. C1 proteases are involved in nutrient digestion by the parasite, which contrasts with the S1 enzymes used in the host. This disparity has already been exploited for a promising anti-schistosome therapy[38]. One protease family (C83) is apparently unique to *S. mansoni*.

Apart from the degradome, but involved in its modulation, 34 protease inhibitors were found (Supplementary Table 19). Most of these are serine protease inhibitors belonging to families I2 (Kunitz-type) and I4 (serpins). Two inhibitors of cysteine proteases (cystatins[39,40]) and two α-2-macroglobulin homologues (I39) were also identified, as were three inhibitor of apoptosis proteins (I32), one of which is highly expressed in adults, where it may function to regulate one or more of the four schistosome caspases.

Neuropeptides

Thirteen putative neuropeptides were identified (Supplementary Table 20), indicating that schistosomes may have much greater diversity than the two described previously. Apart from the neuropeptide Fs (NPFs), most are apparently restricted to the Platyhelminthes—their absence from humans making them a credible source of anthelmintic drug leads. The predicted product of *npp-6* (the amidated heptapeptide AVRLMRLamide) resembles molluscan myomodulin, whereas the two NPP-13 peptides show 100% carboxy-terminal identity with vertebrate neuropeptide-FF-like peptides (peptides ending with a C-terminal sequence PQRFamide); neither of these has previously been reported in any non-vertebrate organism. The discovery of a second NPF (NPP-21b) as well as the known NPP-21a[41] is reminiscent of the vertebrate neuropeptide Y (NPY) superfamily, and strengthens the argument that NPFs and NPYs have a common ancestry.

Metabolic chokepoints

A chokepoint analysis of metabolic pathways reconstructed from the *S. mansoni* genome was used to identify further targets. A total of 607 enzymatic reactions could be placed in pathways, and 120 of these enzymes were identified as chokepoints (Supplementary Table 21). The list of chokepoints includes many that are drug targets in other organisms as well as target reactions already characterized in *S. mansoni*, validating the approach (Supplementary Information). The list also contains new candidate targets and comprises approximately 1% of the *S. mansoni* proteome.

Chemogenomics screening

In the context of neglected tropical diseases and with constrained investment in drug discovery, piggy-backing[37] or 'drug-repositioning' strategies[42] that re-use existing drugs offer potential time-saving and cost benefits. We adopted a twofold strategy to find significant matches between proteins from the parasite and known 'druggable' protein targets of the human host and human-infective pathogens. Using conservative parameters of >50% sequence identity over >80% of the target, we first performed a similarity search against a database of targets curated from medicinal chemistry literature. This revealed 240 distinct *S. mansoni* transcripts with matches to targets against which there are high quality compounds (Supplementary Table 22). Given the need for short-course, oral therapies against schistosomiasis, this list was further reduced to 94 *S. mansoni* targets by filtering for potency and predicted bioavailability. A second search, against a database of the targets for human-directed drugs, showed 66 significant matches with pharmaceuticals marketed at present (Supplementary Table 23), corresponding to 34 *S. mansoni* targets (26, after representing multicopy genes as a single instance; Table 1). For instance, disulfiram, for controlling substance abuse, was highlighted as a potential anti-schistosomal drug; its anti-parasite properties have already been investigated[43]. Manual inspection of the list for compounds with side effects and toxicity can further refine choices—for example, by eliminating the immunosuppressants, cyclosporin and rapamycin. The remaining known drugs could be directly tested in animal models, and either applied unmodified in anti-schistosomal therapy, or could serve as leads for further optimisation. Widening the search beyond the initial strict criteria would expand opportunities, for example, topoisomerase 1 is retrieved below our initial threshold, at 71% identity but only 58% overlap.

Conclusion

A century after Louis Sambon first named the species in 1907 (ref. 44), the sequencing of the *S. mansoni* genome is a landmark event. The

Table 1 | *S. mansoni* genes that match a human gene with marketed drugs

Gene identifier	Protein description	Potential drugs
Smp_005210	Histone deacetylase 1 (HDAC1)	Vorinostat‡
Smp_009030	Ribonucleoside-diphosphate reductase, α subunit, putative	Fludarabine phosphate‡
Smp_012930	Inosine-5-monophosphate dehydrogenase, putative	Mycophenolate mofetil†, mycophenolic acid†, ribavirin§
Smp_015020	Na^+,K^+-ATPase α subunit (SNaK1)	Digoxin‡, digitoxin‡, acetyldigitoxin§, deslanoside§
Smp_016780*	Tubulin α chain, putative	Vinblastine†, colchicine†, vincristine†
Smp_022960	Aldehyde dehydrogenase, putative	Disulfiram‡
Smp_026560	Calmodulin, putative	Bepridil§
Smp_030730*	Tubulin β chain, putative	Colchicine†, vinblastine†, vincristine†, albendazole‡, mebendazole‡, paclitaxel‡, thiabendazole‡, vinorelbine‡, docetaxel§
Smp_040130	Cyclophilin (p17.7)	Cyclosporine†
Smp_040790	Cyclophilin B	Cyclosporine†
Smp_044440	Alcohol dehydrogenase, putative	Fomepizole†
Smp_048430	Thioredoxin glutathione reductase	Auranofin‡
Smp_050390	Aldehyde dehydrogenase, putative	Disulfiram‡
Smp_053220	Aldo-keto reductase, putative	Tolrestat†
Smp_055890	Ribonucleoside-diphosphate reductase small chain, putative	Hydroxyurea†, gemcitabine‡
Smp_065120	Deoxyhypusine synthase, putative	Ciclopirox‡
Smp_069160	Cyclophilin, putative	Cyclosporine†
Smp_079230	Immunophilin FK506 binding protein FKBP12, putative	Pimecrolimus†, temsirolimus†, sirolimus‡, tacrolimus‡
Smp_093280	Histone deacetylase 3 (HDAC3)	Vorinostat‡
Smp_094810	Cyclophilin E	Cyclosporine†
Smp_121920	Vesicular amine transporter, putative	Rauwolfia serpentina‡, reserpine‡, deserpidine§, rescinnamine§, alseroxylon§
Smp_135460	Bifunctional dihydrofolate reductase-thymidylate synthase, putative	Pemetrexed‡, flucytosine‡, floxuridine‡, capecitabine‡, gluorouracil‡
Smp_136300	Tyrosine kinase 5	Dasatinib‡
Smp_147050	ATP synthase α subunit vacuolar, putative	Tiludronate§
Smp_171580	Aromatic amino acid decarboxylase, putative	Carbidopa‡
Smp_173280	Cyclophilin, putative	Cyclosporine†

Gene identifier is the genome project systematic name for each gene. It corresponds to the locus tag in the DDBJ/EMBL/GenBank record and to the main accession numbers for GeneDB.
* There are several copies of tubulin (alpha, Smp_027920, Smp_090120 and Smp_103140; beta, Smp_192110, Smp_079960, Smp_079970, Smp_078040 and Smp_035760).
The potential drugs are classified according to the confidence with which the efficacy of the drug in human can be attributed to the target.
† Direct and clear evidence that this interaction is primarily responsible for the therapeutic action of the drug.
‡ Direct and clear evidence that this interaction represents one mechanism for the drug, other targets/mechanisms may also exist.
§ Indirect or inferred evidence of the association of the drug, target and therapeutic action, although the exact mechanism is still speculative.

sequence provides the scientific community with several avenues to study this under-researched human pathogen, and will drive future evolutionary, genetic and functional genomic research. Not least, given that just one drug is widely available to treat schistosomiasis at present, the genome sequence, including the genome-mining analysis presented, offers the possibility that new drug candidates will soon be identified.

METHODS SUMMARY

Mixed-sex cercariae from the Puerto Rico isolate of *S. mansoni*[45], released from infected *Biomphalaria glabrata* snails, were placed in low-melting agarose plugs and genomic DNA was prepared by standard methods. Approximately sixfold coverage of the nuclear genome was obtained using a whole-genome shotgun sequencing approach, in which libraries of different cloned insert sizes (in plasmid, fosmid and bacterial artificial chromosomes (BAC) vectors) were randomly sequenced by Sanger technology from either end. Sequence reads were assembled, and scaffolds were FISH-mapped to individual chromosomes where possible (Supplementary Table 2). The output of several gene prediction algorithms, trained using 409 manually curated gene structures, were integrated into a single set of gene predictions (version 4), which were used for subsequent analyses. Data were accessed from GeneDB (http://www.genedb.org), and Artemis was used for manual annotation and curation of a further 958 genes during subsequent analyses (as described previously[46]).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 January; accepted 22 May 2009.

1. Steinmann, P., Keiser, J., Bos, R., Tanner, M. & Utzinger, J. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *Lancet Infect. Dis.* **6,** 411–425 (2006).
2. Gryseels, B., Polman, K., Clerinx, J. & Kestens, L. Human schistosomiasis. *Lancet* **368,** 1106–1118 (2006).
3. van der Werf, M. J. *et al.* Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. *Acta Trop.* **86,** 125–139 (2003).
4. King, C. H., Dickman, K. & Tisch, D. J. Reassessment of the cost of chronic helmintic infection: a meta-analysis of disability-related outcomes in endemic schistosomiasis. *Lancet* **365,** 1561–1569 (2005).
5. Doenhoff, M. J. & Pica-Mattoccia, L. Praziquantel for the treatment of schistosomiasis: its use for control in areas with endemic disease and prospects for drug resistance. *Expert Rev. Anti Infect. Ther.* **4,** 199–210 (2006).
6. The Schistosoma japonicum Genome Sequencing and Functional Analysis Consortium. The *Schistosoma japonicum* genome reveals features of host–parasite interplay. *Nature* doi:10.1038/nature08140 (this issue).
7. Drew, A. C., Minchella, D. J., King, L. T., Rollinson, D. & Brindley, P. J. SR2 elements, non-long terminal repeat retrotransposons of the RTE-1 lineage from the human blood fluke *Schistosoma mansoni*. *Mol. Biol. Evol.* **16,** 1256–1269 (1999).
8. DeMarco, R., Machado, A. A., Bisson-Filho, A. W. & Verjovski-Almeida, S. Identification of 18 new transcribed retrotransposons in *Schistosoma mansoni*. *Biochem. Biophys. Res. Commun.* **333,** 230–240 (2005).
9. Volfovsky, N., Haas, B. J. & Salzberg, S. L. Computational discovery of internal micro-exons. *Genome Res.* **13,** 1216–1221 (2003).
10. Ghedin, E. *et al.* Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* **317,** 1756–1760 (2007).
11. Abad, P. *et al.* Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnol.* **26,** 909–915 (2008).
12. Opperman, C. H. *et al.* Sequence and genetic map of *Meloidogyne hapla*: a compact nematode genome for plant parasitism. *Proc. Natl Acad. Sci. USA* **105,** 14802–14807 (2008).
13. Dorsey, C. H., Cousin, C. E., Lewis, F. A. & Stirewalt, M. A. Ultrastructure of the *Schistosoma mansoni* cercaria. *Micron* **33,** 279–323 (2002).
14. Wu, W., Niles, E. G., Hirai, H. & LoVerde, P. T. Evolution of a novel subfamily of nuclear receptors with members that each contain two DNA binding domains. *BMC Evol. Biol.* **7,** 27 (2007).
15. Sayed, A. A. *et al.* Identification of oxadiazoles as new drug leads for the control of schistosomiasis. *Nature Med.* **14,** 407–412 (2008).
16. Brouwers, J. F., Smeenk, I. M., van Golde, L. M. & Tielens, A. G. The incorporation, modification and turnover of fatty acids in adult *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **88,** 175–185 (1997).
17. Barrett, J. *Biochemistry of Parasitic Helminths* (Macmillan Publishers, 1981).
18. de Kroon, A. I. Metabolism of phosphatidylcholine and its implications for lipid acyl chain composition in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* **1771,** 343–352 (2007).
19. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nature Rev. Drug Discov.* **5,** 993–996 (2006).
20. Hamdan, F. F. *et al.* A novel *Schistosoma mansoni* G protein-coupled receptor is responsive to histamine. *Mol. Biochem. Parasitol.* **119,** 75–86 (2002).
21. Agboh, K. C., Webb, T. E., Evans, R. J. & Ennion, S. J. Functional characterization of a P2X receptor from *Schistosoma mansoni*. *J. Biol. Chem.* **279,** 41650–41657 (2004).
22. Kaczorowski, G. J., McManus, O. B., Priest, B. T. & Garcia, M. L. Ion channels as drug targets: the next GPCRs. *J. Gen. Physiol.* **131,** 399–405 (2008).
23. Kim, E., Day, T. A., Bennett, J. L. & Pax, R. A. Cloning and functional expression of a *Shaker-related* voltage-gated potassium channel gene from *Schistosoma mansoni* (Trematoda: Digenea). *Parasitology* **110,** 171–180 (1995).
24. Salkoff, L. *et al.* Potassium channels in *C. elegans*. *WormBook* **Dec 30,** 1–15 (2005).

25. Jeziorski, M. C. & Greenberg, R. M. Voltage-gated calcium channel subunits from platyhelminths: potential role in praziquantel action. *Int. J. Parasitol.* **36,** 625–632 (2006).
26. Boyle, S. N. & Koleske, A. J. Dissecting kinase signaling pathways. *Drug Discov. Today* **12,** 717–724 (2007).
27. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298,** 1912–1934 (2002).
28. Lopez-Otin, C. & Overall, C. M. Protease degradomics: a new challenge for proteomics. *Nature Rev. Mol. Cell Biol.* **3,** 509–519 (2002).
29. Rawlings, N. D. & Morton, F. R. The MEROPS batch BLAST: a tool to detect peptidases and their non-peptidase homologues in a genome. *Biochimie* **90,** 243–259 (2008).
30. Abbenante, G. & Fairlie, D. P. Protease inhibitors in the clinic. *Med. Chem.* **1,** 71–104 (2005).
31. Fear, G., Komarnytsky, S. & Raskin, I. Protease inhibitors and their peptidomimetic derivatives as potential drugs. *Pharmacol. Ther.* **113,** 354–368 (2007).
32. Page, M. J. & Di Cera, E. Serine peptidases: classification, structure and function. *Cell. Mol. Life Sci.* **65,** 1220–1236 (2008).
33. Krem, M. M. & Di Cera, E. Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem. Sci.* **27,** 67–74 (2002).
34. Zou, Z., Lopez, D. L., Kanost, M. R., Evans, J. D. & Jiang, H. Comparative analysis of serine protease-related genes in the honey bee genome: possible involvement in embryonic development and innate immunity. *Insect Mol. Biol.* **15,** 603–614 (2006).
35. Ieko, M. *et al.* Factor Xa inhibitors: new anti-thrombotic agents and their characteristics. *Front. Biosci.* **11,** 232–248 (2006).
36. Okun, I., Balakin, K. V., Tkachenko, S. E. & Ivachtchenko, A. V. Caspase activity modulators as anticancer agents. *Anticancer Agents Med. Chem.* **8,** 322–341 (2008).
37. Caffrey, C. R. & Steverding, D. Recent initiatives and strategies to developing new drugs for tropical parasitic diseases. *Expert Opin. Drug Discov.* **3,** 173–186 (2008).
38. Abdulla, M. H., Lim, K. C., Sajid, M., McKerrow, J. H. & Caffrey, C. R. *Schistosomiasis mansoni*: novel chemotherapy using a cysteine protease inhibitor. *PLoS Med.* **4,** e14 (2007).
39. Cao, M., Chao, H. & Doughty, B. L. A cDNA from *Schistosoma mansoni* eggs sharing sequence features of mammalian cystatin. *Mol. Biochem. Parasitol.* **57,** 175–176 (1993).
40. Morales, F. C., Furtado, D. R. & Rumjanek, F. D. The N-terminus moiety of the cystatin SmCys from *Schistosoma mansoni* regulates its inhibitory activity *in vitro* and *in vivo*. *Mol. Biochem. Parasitol.* **134,** 65–73 (2004).
41. Humphries, J. E. *et al.* Structure and bioactivity of neuropeptide F from the human parasites *Schistosoma mansoni* and *Schistosoma japonicum*. *J. Biol. Chem.* **279,** 39880–39885 (2004).
42. Ashburn, T. T. & Thor, K. B. Drug repositioning: identifying and developing new uses for existing drugs. *Nature Rev. Drug Discov.* **3,** 673–683 (2004).
43. Nash, T. & Rice, W. G. Efficacies of zinc-finger-active drugs against *Giardia lamblia*. *Antimicrob. Agents Chemother.* **42,** 1488–1492 (1998).
44. Sambon, L. W. New or little known African Entozoa. *J. Trop. Med. Hyg.* **10,** 117 (1907).
45. Fletcher, M., LoVerde, P. T. & Woodruff, D. S. Genetic variation in *Schistosoma mansoni*: enzyme polymorphisms in populations from Africa, Southwest Asia, South America, and the West Indies. *Am. J. Trop. Med. Hyg.* **30,** 406–421 (1981).
46. Berriman, M. & Harris, M. Annotation of parasite genomes. *Methods Mol. Biol.* **270,** 17–44 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The genome sequencing and annotation work was funded by the Wellcome Trust (grant number WT085775/Z/08/Z) and the National Institutes of Health (NIH) National Institute of Allergy and Infectious Diseases (NIAID) grant AI48828 to N.M.E.-S. We thank N. D. Rawlings of the MEROPS database team at the Wellcome Trust Sanger Institute for his help, J. C. Illes for discussions on polarity complexes, and F. Prosdocimi and M. R. D. Sananes for early discussions and analyses in the project. FISH chromosome mappings were partially supported by Oyama Health Foundation (H.H.), Japan Society for the Promotion of Science (13557021) (H.H.), 21st century Centers of Exellence and global Centers of Excellence of Japan's Ministry of Education, Culture, Sports, Science and Technology. Additional support was by The Sandler Foundation (C.R.C. and M.S.), NIH-Fogarty 5D43TW006580 (P.T.L.), NIH-Fogarty 5D43TW007012-03, NIH grant AI054711-01A2 (R.A.W. and G.P.D.), FAPEMIG REDE-281/05 (G.O.), the PhRMA Foundation (Postdoctoral Fellowship in Informatics to S.T.M.), The Burroughs Wellcome Fund (P.T.L.) and the United Nations Children's Fund (UNICEF)/United Nations Development Program (UNDP)/World bank/World Health Organization (WHO) Special program for research and training in tropical diseases (TDR) (P.T.L.). R.D. was a recipient of CAPES and FAPESP fellowships.

Author Contributions A.I., R.A.W., C.M.F.-L., D.A.J., N.M.E.-S. and P.T.L. initiated the project; M.A.Q. constructed DNA libraries and J.P. and J.R. directed sequencing; Y.G. and Z.N. assembled the genome sequence data; H.H., P.T.L., R.J.P. and Y.H. produced the mapping data; A.De., A.Dj., A.R.T., B.J.H., D.C.B., D.L., G.C.C., J.W., M.A.A., M.-A.R., M.Sa., O.W., P.D.A., R.H., S.L.S. and T.E. provided computational and bioinformatic support; A.R.T., M.A.A. and R.H. set up and maintained the genome database; C.D.M., D.C.B., G.B., G.C.C. and J.A.G. produced the gene finding training set; B.J.H., M.P. and M.St. trained genefinding software; A.V.P., B.G.B. and B.J.H. annotated the genome data; A.C., A.Z., B.A.-L., C.R.C., D.L.W., G.O., G.P.D., J.P.O., L.F.A., M.Sa., M.Z., P.M., R.C., R.D., S.T.M., T.A.D. and W.W. contributed specific analysis topics presented in this manuscript; C.H.-F. and E.G. contributed to general project and sequencing management; B.G.B., C.H.-F., C.R.C. and J.P. commented on the manuscript drafts; G.C.C. performed data submission to GenBank; R.A.W., M.B., N.M.E.-S. and P.T.L. drafted and edited the paper; R.A.W., D.A.J. and P.T.L. provided DNA resources for the sequencing; M.B. and N.M.E.-S. directed the project and assembled the manuscript.

Author Information The annotated genome sequence has been submitted to EMBL with the accession numbers FN357292–FN376313. All data are also available for browsing in the GeneDB database (http://www.genedb.org/genedb/smansoni/). The CHORI BAC clones used in this study are available from http://bacpac.chori.org/. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.B. (mb4@sanger.ac.uk) or N.M.E.-S. (elsayed@umd.edu).

METHODS

Genome sequencing, assembly and mapping. The most commonly used Puerto Rican strain of *S. mansoni*[45] was maintained in albino *B. glabrata* snail and NMRI mice and golden hamsters as laboratory hosts (*Mesocricetus auratus*). Cercariae released from infected snails were resuspended in PBS at a concentration of 5×10^5 cercariae ml^{-1}. The parasites were transferred to a 42 °C water bath, incubated for 5 min, and mixed with an equal volume of 1.2% low-melting point agarose (Gibco-BRL) in PBS at 42 °C. The agarose/cell mixture was transferred to a disposable plug mould (Bio-Rad), placed on ice, treated twice for 24 h at 50 °C with 1% *N*-lauroyl sarcosine, 0.5 M EDTA, pH 8.0, 2 mg ml^{-1} proteinase K (Boehringer Mannheim). Proteinase K was then inhibited by a 30-min treatment with PMSF (40 μg ml^{-1}), followed by three successive 30-min dialyses against 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA. Sequencing libraries were constructed using genomic DNA extracted from mixed-sex cercariae. Sequencing reads were produced from small insert plasmid clones containing a range of insert sizes. In addition, 12,305 BAC end sequences from *S. mansoni* BAC library Sm1 (DDBJ/EMBL/GenBank accession numbers BH199420–BH211620), 19,136 CHORI 103 BAC end sequences (DDBJ/EMBL/GenBank accession numbers DX983724–ED003998) and 16,628 fosmid end reads were included. After filtering out low quality reads, 85% of the remaining 3.19 million reads were assembled using Phusion[47] into 381 Mb, or 363 Mb after filtering small (<2 kb) contigs (see Supplementary Table 1). The size is greater than the previously estimated size of 270 Mb[48], although this size estimate can be revised to 300 Mb because the original measurements were made using the *E. coli* genome as a control, which has a length that is 10% greater than previously thought. From the assembly, a depth of coverage of ~sixfold was calculated.

A physical map was generated using FISH to localize *S. mansoni* BACs to the seven autosomal and sex pairs of chromosomes using previously published methods[49]. Clones from two BAC libraries, Sm1 (ref. 50) and CHORI-103 (http://bacpac.chori.org/schis103.htm), each constructed from cercarial DNA were randomly picked and subjected to FISH analysis. Owing to the repetitive nature of the schistosome genome, BACs would often hybridize to more than one chromosome. This was in spite of using sheared genomic DNA to block the repetitive sequences. Of the 500 clones analysed, 334 showed unique hybridization patterns (Fig. 1, Supplementary Fig. 1 and Supplementary Table 2). A total of 118 BACs that were FISH-mapped were among those end-sequenced, and 153 scaffolds were assigned to a specific chromosome.

Retroelements analysis. We performed an iterative search of retroelements using the conserved reverse transcriptase domain as previously described[51]. Elements with higher than 80% nucleotide identity in the reverse-transcriptase region were considered as members of the same family. To obtain an unbiased estimate of abundance for each element in the genome, all the identified families were mapped to the shotgun reads using BLASTN[52]. The number of bases spanned by the alignment for each element was counted and compared with the total number of bases in the shotgun data to determine their representation in the *S. mansoni* genome.

Genome annotation and repeat content analysis. A training set (for *ab initio* gene finding) of 409 genes was manually curated from *S. mansoni* sequences already within the Uniprot database and manual prediction of highly conserved genes. Further genome-wide gene predictions were made using both EVidenceModeller and PASA[53]. EVidenceModeller uses an evidence-combining strategy to compute an optimal set of protein-coding gene structures derived from several, often conflicting, sources of gene predictions. The sources of evidence for our annotation of the *S. mansoni* genome included the following: *ab initio* gene predictions derived from GlimmerHMM[54], TWINSCAN[55], and Augustus[56]; protein sequence homologies to a non-redundant protein database using AAT[57]; cross-genome sequence homologies between *S. mansoni* and *S. japonicum* using PROmer[58]; spliced genome alignments to ESTs using GMAP[59]; and repeat regions identified using RepeatScout[60] and RepeatMasker (A. F. A. Smit, R. Hubley and P. Green, unpublished observations, http://www.repeatmasker.org). Consensus gene predictions generated by EVidenceModeller were further modified to include annotations of untranslated regions and alternative splicing isoforms for 1,038 genes by applying PASA leveraging the earlier GMAP aligned ESTs. A total of 13,197 transcripts were predicted for 11,809 genes. Of the 30,110 previously described EST clusters, 24,373 map to contigs >1 kb in the current genome assembly. The true number of genes could therefore be as high as 17,500. By parsing BLAST description lines, putative products were assigned to each gene. During the course of subsequent analyses, 958 of these were manually edited using the Artemis annotation tool.

For an unusually large gene, encoding a putative ryanodine receptor spanning ~164 kb, 79 of its 93 intron–exon boundaries were confirmed by RT–PCR. Approximately 45% of the *S. mansoni* genome was found to be repetitive, computed by summing up genomic bases matching known *S. mansoni* mobile element sequences or repeat family consensus sequences derived from the RepeatScout *de novo* repeat library. The repeat content was also assessed on the basis of the distribution of random sequences 25 nucleotides in length, 104,028,213 out of 373,600,457 or 28% of bases were repetitive. Note, this value is significantly lower than that of RepeatMasker because the latter allows sequence divergence of up to 20%.

Analysis of putative transcription factors. Profile hidden Markov models (HMMs) of domains present in the proteins that constitute the TRANSFAC eukaryotic transcriptional factor database[61] and the DBD DNA-binding domain database[62] were used to search the genome of *S. mansoni* in conjunction with 63 other eukaryotic genomes. The score threshold was defined as the lowest pairwise score among all members of the Pfam family associated to the HMM. The putative transcriptional activator proteins were then clustered on the basis of sequence similarities (BLASTP *E* value $\leq 10^{-6}$ considered significant) using the TRIBE-MCL algorithm[63] and an inflation value of 2.0 (ref. 64).

Micro-exon genes. MEGs were predicted as previously described[9] with further manual refinement using available *S. mansoni* EST data (including both published data[65] and unpublished data from GenBank/dbEST or ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/ESTs/). Further family members were identified by similarity searches against the available supercontigs in the assembly with long flanking MEG exons as query sequences. Signal peptides and transmembrane domains were detected using SignalP[66] and TMHMM 2.0 (ref. 67) programs, respectively.

Expression of MEG families at different stages throughout the life cycle was analysed by BLAST searching the sequences of all members of a family against the complete *S. mansoni* EST data set, which comprises ESTs from the following developmental stages: germball (28,497), cercaria (21,639), 3-day somule (6,122), 7-day somule (41,043), 21-day liver worm (6,044), 28-day liver worm (11,227), 45-day adult worm (59,552), egg (33,674) and miracidium (19,982).

Evolutionary analysis. To identify orthologues and paralogues of *S. mansoni* genes, we built a standalone version of the TreeFam database (version 7) of animal gene families[68,69]. For each *S. mansoni* predicted protein, we identified the top-matching TreeFam 'clean' family using HMMER[70] (with $E \leq 10^{-10}$ as a cutoff). Similarly, the top-matching family was identified for each *Nematostella vectensis* (release version 1.0)[71] and *S. japonicum* protein. Trees and alignments were built for the families as for the standard TreeFam pipeline. This resulted in trees for 5,829 families that contain *S. mansoni*, *S. japonicum* or *N. vectensis* genes. From these trees, we identified within-species paralogues in the three species, and identified the ancestral taxon in which the duplication that gave rise to each pair of paralogues occurred.

Kinome. A eukaryotic protein kinase domain HMM was built from a manually adjusted alignment of 68 diverse kinase domain sequences from yeast, worm, fly and human that share <50% sequence identity in the catalytic domain. To test the selectivity of the model, it was run against the Uniprot database. Using a $P < 0.1$ cutoff, the model detected 2,688 putative domains, all of which were annotated either as kinases or putative kinases in different description fields. Local and global HMM models were built with the HMMer package (http://hmmer.janelia.org/) from several sequence alignments generated by MAFFT software[72] and were used for sensitive searches against the *S. mansoni* database.

Identified genes were annotated using Artemis, integrating data from Interproscan and Reverse PSI-BLAST searches[73] and the size of the *S. mansoni* kinome was compared with those of: *Plasmodium falciparum*[74], *Homo sapiens*[27], *Trypanosoma cruzi*[75], *Trypanosoma brucei*[76], *C. elegans*[77], *Leishmania major*[78] and *Mus musculus*[79]. A dendogram was constructed using the kinase domains of the identified proteins with the CLC Main Workbench (CLC bio) using the neighbour-joining method with 1,000 replicates.

Identification of putative proteases and inhibitors. We used the MEROPS database[80] (http://merops.sanger.ac.uk) to identify active *S. mansoni* proteases and protease inhibitor homologues, using BLASTP[52,73] with $E \leq 10^{-9}$ as a cutoff. More distant relatives were identified through HMMER version 2.3.2 (ref. 70) searches of Pfam models[81] that corresond to MEROPS families (Pfam version 22.0 (ref. 82), http://pfam.sanger.ac.uk/), using the same *E*-value cutoff. This initial data set contained 656 provisional homologues, having removed predicted proteases <80 residues in length as well as provisional inhibitors <50 residues long. A secondary screen against the NCBI non-redundant protein database retained a total of 369 *S. mansoni* sequences, which overlapped in at least 50% of their MEROPS hit or Pfam domain with an experimentally characterized protease or inhibitor homologue. False positives were removed by comparing nonspecific MEROPS description lines (for example, 'non-peptidase homologues') to the top non-redundant BLAST hits with an *E*-value at least 3 logs greater than the top MEROPS or Pfam hit but lacking associated experimental validation. This approach removed MEROPS proteins that are not functional proteases but are structurally related (such as hormone-sensitive lipases in the family S9; flagged as homologues of proteins that are inactive protease homologues in

Supplementary Table 18). Similarly, the Pfam database models domains found not only in proteases and inhibitors but also in a wide range of other proteins (for example, PF00047, PF00059, PF00561, PF01476 and PF0764) were also removed as false positives in the absence of further evidence.

We next predicted which of the putative protease homologues were likely to be active or inactive. BLAST alignments of proteins against putative homologues classified in MEROPS predicted active site positions and residues in the *S. mansoni* query sequence, followed by manual inspection of sequence alignments to refine active site residue predictions. In a few cases, in which an acceptable alignment was not produced by BLASTP of MEROPS, a non-redundant sequence was used. In more difficult cases, involving two closely related *S. mansoni* sequences, active site residues were identified from multiple alignments of *S. mansoni* sequences, a representative sequence for the corresponding MEROPS family, and the seed alignment sequences for the relevant Pfam model.

Metabolic chokepoint analysis. An *S. mansoni* metabolic pathways database, SchistoCyc (http://schistocyc.schistodb.net/ptools), was created using the Pathway Tools software[83], which contains algorithms to predict an organism's metabolic pathways from its genome by comparison to MetaCyc, a reference pathways database[84]. From the pathway database, potential chokepoint reactions[85] were identified (those that uniquely consume a specific substrate or produce a specific product). Chokepoint reactions are probably critical to normal cellular physiology and therefore represent potential drug targets.

Chemogenomics. To identify, in *S. mansoni*, putative proteins for which therapeutic compound or high quality chemical tools may already be available, sequence similarity searching was performed using BLASTP against the Target Dictionary from Drugstore[19] (database of Food and Drug Administration approved drugs) and StARlite (a database of Structure Activity Relationship data abstracted and indexed manually from the primary literature and at present containing 440,055 unique compounds, directed against approximately 3,500 distinct molecular targets from the primary literature). The results were stringently filtered for significance: ≥50% identity, ≥80% overlap of the target and a BLAST $E \geq 10^{-10}$. To prioritise 755 hits to StARlite, we applied filters for potency/affinity against the matched target, combined with an estimate of the likelihood that the compound could be orally absorbed. The potency cutoff was set at a half-maximal inhibitory concentration (IC_{50}), inhibition constant (K_i), or dissociation constant (K_d) of 100 nM or better, and oral bioavailability was estimated using the 'rule of five' (molecular weight of no more than 500 Da, clogP less than five, no more than five hydrogen bond donors and no more than ten nitrogen or oxygen atoms)[86]. The drugs associated with matches in the DrugStore database were classified according to a broad range of current therapeutic categories: (1) direct and clear evidence that this interaction is primarily responsible for the therapeutic action of the drug; (2) direct and clear evidence that this interaction is one mechanism for the drug but other targets or mechanisms may exist; and (3) indirect or inferred evidence of the association of the drug, target and therapeutic action.

47. Mullikin, J. C. & Ning, Z. The Phusion assembler. *Genome Res.* **13,** 81–90 (2003).
48. Simpson, A. J., Sher, A. & McCutchan, T. F. The genome of *Schistosoma mansoni*: isolation of DNA, its size, bases and repetitive sequences. *Mol. Biochem. Parasitol.* **6,** 125–137 (1982).
49. Hirai, H. & Hirai, Y. FISH mapping for helminth genome. *Methods Mol. Biol.* **270,** 379–394 (2004).
50. Le Paslier, M. C. *et al.* Construction and characterization of a *Schistosoma mansoni* bacterial artificial chromosome library. *Genomics* **65,** 87–94 (2000).
51. Biedler, J. & Tu, Z. Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity. *Mol. Biol. Evol.* **20,** 1811–1825 (2003).
52. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
53. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9,** R7 (2008).
54. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics* **20,** 2878–2879 (2004).
55. Korf, I., Flicek, P., Duan, D. & Brent, M. R. Integrating genomic homology into gene structure prediction. *Bioinformatics* **17** (suppl. 1), S140–S148 (2001).
56. Stanke, M., Schoffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7,** 62 (2006).
57. Huang, X., Adams, M. D., Zhou, H. & Kerlavage, A. R. A tool for analyzing and annotating genomic sequences. *Genomics* **46,** 37–45 (1997).
58. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5,** R12 (2004).
59. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21,** 1859–1875 (2005).
60. Price, A. L., Jones, N. C. & Pevzner, P. A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21** (suppl. 1), i351–i358 (2005).
61. Wingender, E., Dietze, P., Karas, H. & Knuppel, R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res.* **24,** 238–241 (1996).
62. Kummerfeld, S. K. & Teichmann, S. A. DBD: a transcription factor prediction database. *Nucleic Acids Res.* **34,** D74–D81 (2006).
63. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30,** 1575–1584 (2002).
64. Coulson, R. M. & Ouzounis, C. A. The phylogenetic diversity of eukaryotic transcription. *Nucleic Acids Res.* **31,** 653–660 (2003).
65. Verjovski-Almeida, S. *et al.* Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nature Genet.* **35,** 148–157 (2003).
66. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2,** 953–971 (2007).
67. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305,** 567–580 (2001).
68. Li, H. *et al.* TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* **34,** D572–D580 (2006).
69. Ruan, J. *et al.* TreeFam: 2008 update. *Nucleic Acids Res.* **36,** D735–D740 (2008).
70. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14,** 755–763 (1998).
71. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317,** 86–94 (2007).
72. Katoh, K. & Toh, H. Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics* **9,** 212 (2008).
73. 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).
74. Anamika, K., Martin, J. & Srinivasan, N. Comparative kinomics of human and chimpanzee reveal unique kinship and functional diversity generated by new domain combinations. *BMC Genomics* **9,** 625 (2008).
75. El-Sayed, N. M. *et al.* The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309,** 409–415 (2005).
76. Berriman, M. *et al.* The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309,** 416–422 (2005).
77. Manning, G. Genomic overview of protein kinases. *WormBook* **Dec 13,** 1–19 (2005).
78. Parsons, M., Worthey, E. A., Ward, P. N. & Mottram, J. C. Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*. *BMC Genomics* **6,** 127 (2005).
79. Caenepeel, S., Charydczak, G., Sudarsanam, S., Hunter, T. & Manning, G. The mouse kinome: discovery and comparative genomics of all mouse protein kinases. *Proc. Natl Acad. Sci. USA* **101,** 11707–11712 (2004).
80. Rawlings, N. D., Morton, F. R., Kok, C. Y., Kong, J. & Barrett, A. J. MEROPS: the peptidase database. *Nucleic Acids Res.* **36,** D320–D325 (2008).
81. Sonnhammer, E. L., Eddy, S. R. & Durbin, R. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* **28,** 405–420 (1997).
82. Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **36,** D281–D288 (2008).
83. Karp, P. D., Paley, S. & Romero, P. The Pathway Tools software. *Bioinformatics* **18** (suppl. 1), S225–S232 (2002).
84. Caspi, R. *et al.* MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* **34,** D511–D516 (2006).
85. Yeh, I., Hanekamp, T., Tsoka, S., Karp, P. D. & Altman, R. B. Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery. *Genome Res.* **14,** 917–924 (2004).
86. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46,** 3–26 (2001).

ARTICLES

The active form of DNA polymerase V is UmuD′$_2$C–RecA–ATP

Qingfei Jiang[1], Kiyonobu Karata[2], Roger Woodgate[2], Michael M. Cox[3] & Myron F. Goodman[1]

DNA-damage-induced SOS mutations arise when *Escherichia coli* DNA polymerase (pol) V, activated by a RecA nucleoprotein filament (RecA*), catalyses translesion DNA synthesis. Here we address two longstanding enigmatic aspects of SOS mutagenesis, the molecular composition of mutagenically active pol V and the role of RecA*. We show that RecA* transfers a single RecA–ATP stoichiometrically from its DNA 3′-end to free pol V (UmuD′$_2$C) to form an active mutasome (pol V Mut) with the composition UmuD′$_2$C–RecA–ATP. Pol V Mut catalyses TLS in the absence of RecA* and deactivates rapidly upon dissociation from DNA. Deactivation occurs more slowly in the absence of DNA synthesis, while retaining RecA–ATP in the complex. Reactivation of pol V Mut is triggered by replacement of RecA–ATP from RecA*. Thus, the principal role of RecA* in SOS mutagenesis is to transfer RecA–ATP to pol V, and thus generate active mutasomal complex for translesion synthesis.

Pol V is a low-fidelity DNA polymerase[1,2] induced as part of the SOS regulon in *E. coli* in response to DNA damage[3]. The replicative polymerase, pol III, typically stalls when it encounters a DNA template lesion, arresting movement of the replication fork. One pathway that enables restoration of fork movement involves pol V, which replaces pol III on the sliding β-clamp[4–6] and catalyses translesion DNA synthesis (TLS). Pol V copies numerous types of lesions[7], but in a mutagenic manner[8,9]. After TLS, pol III resumes normal replication.

The pol V complex consists of UmuD′$_2$C[10,11]. Both *in vitro*[1,2,12,13] and *in vivo*[14–16], pol V activity requires the assembly of an active RecA filament on single-stranded (ss) DNA, termed RecA*[17]. The biological functions of RecA* in strand exchange during homologous recombination and in mediating cleavage of the repressor protein LexA and UmuD during the SOS response are well understood[17]. In contrast, the biochemical role of RecA* in pol-V-dependent mutagenic TLS remains poorly characterized.

Proposals for the role of RecA* in TLS have evolved from positioning UmuD′$_2$C on primer/template (p/t) DNA proximal to a lesion[18–20], to a dynamic interaction involving displacement of RecA* filaments on the template by an advancing pol V[4], to a model in which RecA* need not be located *in cis* on the template strand being copied, but can instead assemble on a separate ssDNA strand to transactivate pol V for TLS[21]. What had not been contemplated in previous models, and what previous experimental designs were unable to detect, is whether a RecA* filament might function remotely from the site of translesion synthesis. In this unexplored scenario, RecA* would transfer a RecA monomer and ATP from the 3′-tip of RecA* to free UmuD′$_2$C, thereby converting pol V (UmuD′$_2$C) to an activated form that can be used elsewhere for TLS.

Formation of activated pol V Mut

In earlier studies, either free RecA protein or RecA filament was added to pol V and p/t DNA substrate. Here, we first form RecA* by incubating RecA with biotinylated ssDNA bound to a streptavidin-agarose resin matrix in the presence of ATPγS (adenosine 5′[γ-thio] triphosphate; see Methods). Pol V (UmuD′$_2$C) is then incubated with RecA* in the absence of p/t DNA, forming an activated pol V species that can be isolated, pol V Mut (Supplementary Figs 1 and 2). Only once RecA* is removed by centrifugation is p/t DNA added, allowing DNA synthesis in the absence of RecA* (Fig. 1a, b). In these studies, the p/t template is a hairpin with a 3-nucleotide overhang[21] instead of, for example, an oligonucleotide annealed to a ssDNA circle; this prevents formation of activated RecA* on the p/t DNA itself because the hairpin lacks free ssDNA on which RecA* can assemble (the 3-nucleotide overhang is too short to permit this). Throughout, we use the properties of two key RecA mutant proteins, highly active RecA E38KΔC17 (also known as RecA730ΔC17; ref. 22), and mutagenically inactive RecA F117S (also known as RecA1730; refs 15, 23).

Removal of the RecA*-resin by centrifugation allows activated pol V Mut to remain in solution (Supplementary Fig. 2). Filaments containing RecA E38KΔC17 activate pol V with approximately two-fold higher efficiency (Fig. 1a) than wild-type RecA* (Fig. 1b). Pol V avidly copies p/t DNA when RecA* is present *in trans*, as reported previously[21] (Fig. 1c, d). Thus, exposure of UmuD′$_2$C to RecA* is required for pol V activity. Most importantly, pol V remains highly active after removal of RecA* from the reaction (Fig. 1a, b).

To investigate how long pol V Mut remains active in the absence of RecA*, we incubated the enzyme in standard reaction buffer at 37 °C for various times before adding the p/t DNA and dNTP substrates to initiate DNA synthesis. Polymerase activity decreases as the delay time increases (Fig. 1e, f). The decrease in activity is approximately exponential, and is similar for pol V Mut activated with either RecA E38KΔC17* or wild-type RecA* (Fig. 1g). The loss in activity is reversible; addition of RecA* *in trans* at each time point restores pol V Mut to a fully activated state (Fig. 1g). ATP hydrolysis has no role in pol V activation or in the activity decay, because synthesis by pol V Mut and the loss of activity during the time delay is identical when RecA* is assembled with ATP, ATPγS or AMP-PNP (β-γ-imidoadenosine 5′-triphosphate; Supplementary Figs 3 and 4).

Each active pol V Mut complex can promote only one round of DNA synthesis (Fig. 1h, i). A plot of primer use against time shows that, after the addition of one or more nucleotides, pol V Mut cannot

[1]Departments of Biological Sciences and Chemistry, University of Southern California, University Park, Los Angeles, California 90089-2910, USA. [2]Laboratory of Genomic Integrity, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892-3371, USA. [3]Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.

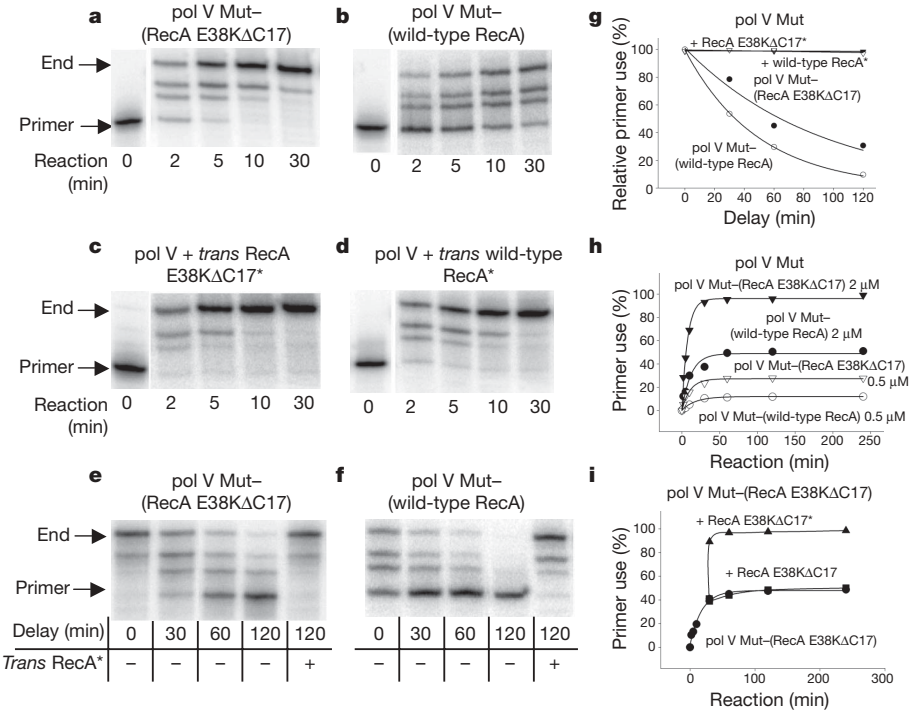


Figure 1 | DNA synthesis by pol V Mut or pol V transactivated by RecA*. The p/t DNA is a hairpin containing a 3-nucleotide template overhang. **a**, **b**, DNA synthesis by pol V Mut–(RecA E38KΔC17) (**a**) or pol V Mut–(wild-type RecA) (**b**) in the absence of RecA*. **c**, **d**, DNA synthesis by pol V transactivated by RecA E38KΔC17* (**c**) or wild-type RecA* (**d**). **e**, **f**, DNA synthesis by pol V Mut–(RecA E38KΔC17) (**e**) or pol V Mut–(wild-type RecA) (**f**) undergoing deactivation as a function of time. **g**, Deactivation of pol V Mut–(RecA E38KΔC17) and pol V Mut–(wild-type RecA) measured by quantifying DNA synthesis obtained from data in panels **e** and **f**, respectively. Complete reactivation of deactivated pol V Mut is observed by addition of *trans* RecA* at each delay time. **h**, pol V Mut performs one round of DNA synthesis and cannot reinitiate synthesis on separate p/t DNA substrate. The concentration of p/t DNA is 1 μM. We estimate that the fraction of active polymerase in the reaction is 50% for pol V Mut–(RecA E38KΔC17) and 25% for pol V Mut–(wild-type RecA). **i**, Addition of *trans* RecA E38KΔC17* enables pol V Mut–(RecA E38KΔC17) to reinitiate DNA synthesis on a separate p/t DNA substrate. Cycling also occurs using pol V Mut–(wild-type RecA) in conjunction with *trans* wild-type RecA* (data not shown). *Trans* RecA* is at 1 μM when present.

reinitiate synthesis on a different p/t DNA (Fig. 1h, i). Deactivated pol V Mut is reactivated by adding RecA* *in trans* (Fig. 1i), but not by adding either RecA (Fig. 1i) or ssDNA alone (data not shown). After synthesis, deactivated pol V Mut dissociates into solution (Supplementary Fig. 5). Pol V Mut catalyses TLS in the absence of RecA*, but without pre-activation, pol V is inactive even on undamaged DNA (Supplementary Fig. 6).

Pol V Mut consists of UmuD′$_2$C–RecA–ATP

Pol V Mut activated with RecA E38KΔC17* synthesizes DNA equally well with or without a nucleotide cofactor, ATP or ATPγS, added post-activation (Fig. 2a). In marked contrast, pol V Mut activated with wild-type RecA* is inactive in the absence of added ATP or ATPγS (Fig. 2b). These properties indicate that activation entails formation of a pol V–RecA complex. A denaturing gel of activated pol V Mut provides visual evidence that UmuD′$_2$C forms a complex with RecA (Fig. 2c). A plot of RecA released from RecA* as a function of increasing pol V concentration shows a 1:1:1 ratio of RecA:UmuC:UmuD′$_2$ over a tenfold range of pol V Mut (Fig. 2d). When resin-bound RecA* is pre-incubated in the absence of pol V and centrifuged, an insignificant amount of RecA is released from the resin into the supernatant (Fig. 2c, lane 1) compared with the amount of RecA bound to pol V Mut (Fig. 2c, lanes 2–5). Quantification of ATPγS released into the supernatant also indicates that a RecA–ATP complex is transferred from RecA* to pol V to form pol V Mut, in a 1:1:1 stoichiometry of UmuD′$_2$C–RecA–ATP (Fig. 2c, d).

To obtain direct evidence that RecA is part of pol V Mut, we formed pol V Mut using fluorescein-labelled (Fl-) wild-type RecA (Fig. 2e); its behaviour is indistinguishable from pol V Mut activated with unlabelled wild-type RecA (Supplementary Fig. 7). We identified the components of the activated and deactivated pol V Mut species by affinity chromatography of pol V Mut–(Fl-wild-type RecA) either immediately after activation or after delay times of 1 or 2 h. After elution from the column, Fl-wild-type RecA and [γ-^{35}S]ATP are retained in a complex with UmuC and UmuD′ irrespective of delay time (Fig. 2e). Thus, deactivation of pol V Mut is neither caused nor accompanied by a concomitant loss of RecA or ATPγS. We verified the composition of reactivated pol V Mut by incubating pol V–(Fl-wild-type RecA) complexes isolated after a 2 h delay time with increasing amounts of unlabelled wild-type RecA* (Fig. 2f). After adsorption and elution of pol V–RecA from the column, the Fl-wild-type RecA and [γ-^{35}S]ATP are readily replaced with unlabelled wild-type RecA and ATPγS, with greater replacement occurring as the concentration of unlabelled wild-type RecA* is increased (Fig. 2f). These data confirm that UmuD′$_2$C, RecA and ATP are present in 1:1:1 stoichiometry (Fig. 2d).

We used laser multi-angle light scattering (MALS)[24] to evaluate the molecular mass of pol V Mut–(wild-type RecA). Pol V alone was activated by incubating UmuD′$_2$C with *trans* wild-type RecA* bound to resin. After centrifugation, the supernatant containing a mixture of pol V Mut–(wild-type RecA) and non-activated pol V was resolved by size-exclusion chromatography. The column effluent was passed directly on-line into a MALS detection system. Two light scattering peaks were observed (Fig. 3a, upper trace). The peak eluting at 17.3 min indicates a molecular mass of 113 kDa, in accord with the value predicted for UmuD′$_2$C–(wild-type RecA) (110 kDa). The peak eluting at 18.4 min indicates a molecular mass of 73 kDa, in accord with UmuD′$_2$C (72 kDa), and coinciding with the single peak exhibited by non-activated pol V (UmuD′$_2$C; Fig. 3a, lower trace). Analysis of the two peaks shows that UmuC, UmuD′ and wild-type RecA are present in the 113 kDa peak, whereas only UmuC and UmuD′ are in the 73 kDa peak (Fig. 3b). Thus, the MALS data

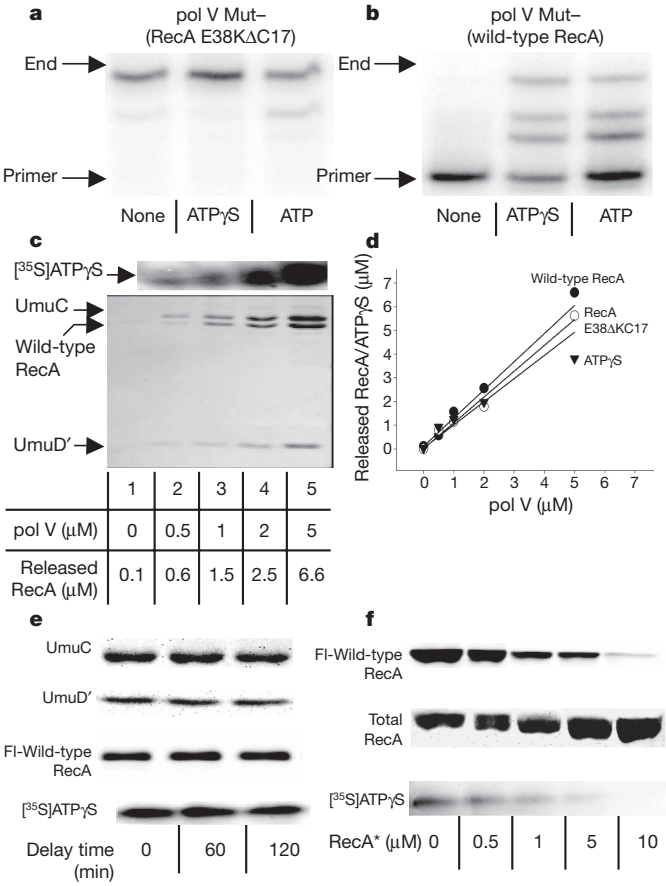


Figure 2 | Activated pol V Mut complex is composed of UmuD′$_2$C–RecA–ATP. **a**, **b**, DNA synthesis by pol V Mut–(wild-type RecA) (**b**), but not pol V Mut–(RecA E38KΔC17) (**a**), requires the presence of ATPγS or ATP. **c**, **d**, Activated pol V Mut–(wild-type RecA) complex contains UmuC, UmuD′ and wild-type RecA, in a 1:1:1 ratio. **e**, Fl-wild-type RecA and ATPγS remain bound in deactivated pol V Mut–(Fl-wild-type RecA) complexes as measured by affinity chromatography. **f**, Reactivation of pol V Mut–(Fl-wild-type RecA) is achieved by replacing Fl-wild-type RecA and [^{35}S]ATPγS bound to deactivated pol V Mut–(Fl-wild-type RecA) with unlabelled wild-type RecA and ATPγS from wild-type RecA*. Reactivated pol V Mut–(Fl-wild-type RecA) complexes were isolated as in panel **e**.

provide independent verification that the intact pol V Mut–(wild-type RecA) complex contains UmuD′$_2$C bound to wild-type RecA, in accord with the affinity column binding and elution analysis of pol V Mut–(wild-type RecA) (Fig. 2e, f).

Activation requires transfer from the 3′-proximal tip of RecA*

We have previously shown that pol V is activated only when it encounters a RecA molecule located at the 3′-proximal tip of RecA*[21]. We determined whether RecA is transferred to pol V from one or both RecA* filament ends by incubating pol V with RecA* having either the 3′- or 5′-proximal tip exposed (Supplementary Fig. 8a). After removing RecA* by centrifugation, the supernatant containing pol V Mut was separated by SDS–polyacrylamide gel electrophoresis (Fig. 4a). The data show that one molecule each of RecA and ATPγS is transferred to pol V from both ends of RecA* (Fig. 4a and Supplementary Fig. 8b). However, pol V becomes activated only if wild-type RecA or RecA E38KΔC17 is transferred from the 3′-proximal tip of RecA* (Fig. 4b). RecA F117S is also transferred to pol V from either the 3′- or 5′-tip, but fails to activate pol V in both cases (Fig. 4a, b).

Discussion

We propose a comprehensive biochemical model depicting pol V activation, deactivation and reactivation (Fig. 5). Pol V (UmuD′$_2$C) first migrates to a preformed RecA* filament. It is activated by the transfer of a 3′-RecA–ATP subunit from RecA* to pol V to form activated pol V Mut, that is: UmuD′$_2$C–RecA–ATP. Pol V Mut scans the genome for sites where replication has stalled. When a suitable primer/template is located, pol V Mut binds and promotes DNA synthesis in the absence of RecA*. Pol V Mut is inactivated by one round of primer extension. Subsequent RecA*-mediated reactivation of pol V Mut occurs when the complex again encounters a RecA* filament, and replaces the 'used' RecA–ATP, which is still bound, but in an inactive conformation, with 'new' RecA–ATP (Fig. 5). The model has the unique feature that there is no need for an active RecA* filament to be present during TLS; also, although ATP is required, ATP hydrolysis is not.

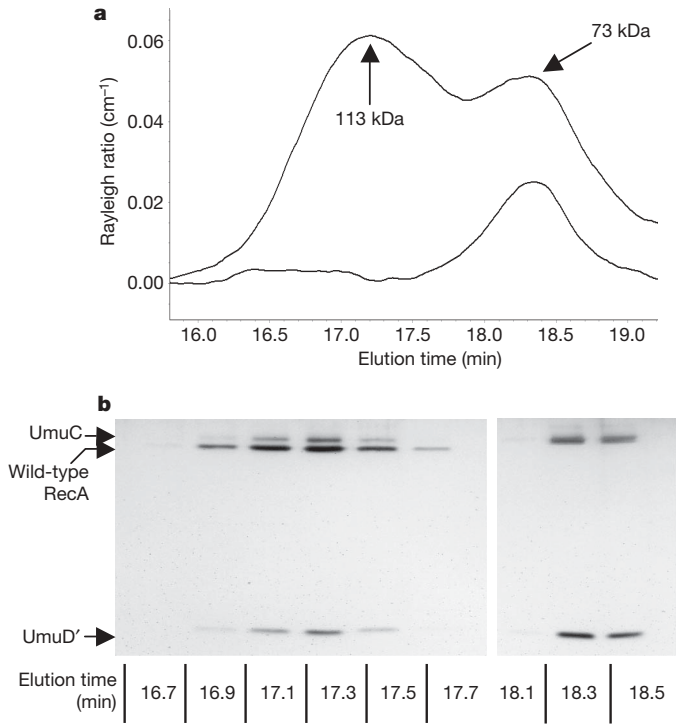


Figure 3 | Determination of the molecular mass of pol V Mut–(wild-type RecA) by MALS. **a**, After wild-type-RecA*-mediated transactivation of UmuD′$_2$C and removal of wild-type RecA*, the mixture of pol V Mut–(wild-type RecA) and non-activated pol V was resolved by size-exclusion chromatography (upper trace), and the molecular mass corresponding to each peak was measured by MALS. Non-activated pol V run separately on the silica gel elutes at 18.4 min (lower trace). **b**, Silver-stained SDS–polyacrylamide gel showing the protein composition from the two peaks contained in panel **a** (upper trace).

Pol-V-mediated TLS requires the presence of RecA* both *in vivo*[14–16] and *in vitro*[1,2,12,13]. Since the late 1980s the most important but refractory questions have been 'what is the molecular composition of the pol V mutasome?' and 'what is RecA* doing?'. The requirement for RecA* initially suggested that a RecA filament would form *in cis*[4,13]. However, RecA filaments formed *in cis* strongly impede pol-V-mediated DNA synthesis[4]. Activation *in trans* resolves the difficult situation of pol V replication of DNA with RecA bound to it[21]. However, transactivation introduces new structural and topological problems associated with a RecA filament bound to DNA in one genomic location interacting continuously with a pol V replicating in a DNA gap somewhere else. And finally, how is pol V inactivated when the SOS response is no longer needed?

The model in Fig. 5 answers these questions. The role of RecA* is both to promote the autocatalytic cleavage of UmuD to UmuD′, and to activate the resulting UmuD′$_2$C complex by transferring a single RecA subunit to pol V from its 3′-proximal tip. Activated pol V Mut can migrate to wherever it is needed; no RecA* filament needs be near the site where TLS actually takes place. In fact, no RecA nucleoprotein

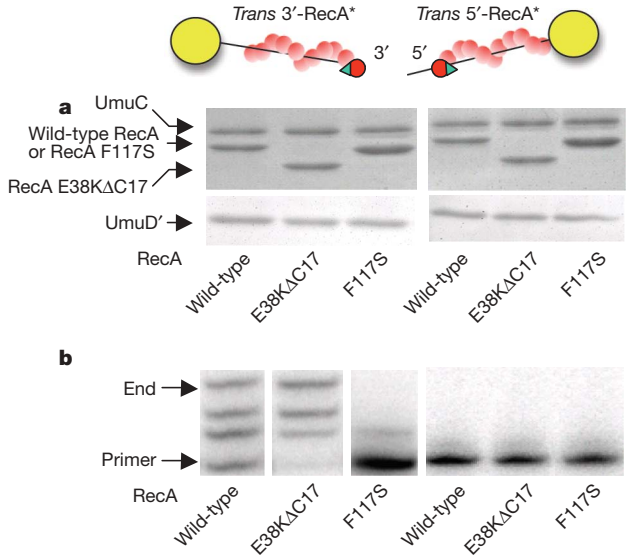


Figure 4 | Pol V activation requires transfer of a RecA molecule from the DNA 3′ end of RecA*. a, The amount of RecA transferred to pol V from exposed 3′- or 5′-tips of RecA* are similar for wild-type RecA*, RecA E38KΔC17* and RecA F117S*. **b**, DNA synthesis catalysed by pol V Mut–(wild-type RecA) or pol V Mut–(RecA E38KΔC17) occurs when RecA is transferred from the 3′-RecA* tip (left gel). Pol V Mut cannot synthesize DNA when RecA transfer takes place from the 5′-RecA* tip (right gel). Mutagenically inactive RecA F117S* fails to activate pol V.

filament needs to participate directly in TLS, either *in cis* or *in trans*, a fact obscured in previous studies because such filaments were always present when TLS was observed. Participation in TLS as a subunit of pol V Mut becomes the only filament-independent role documented for RecA. Rapid inactivation of pol V Mut after TLS, or slower inactivation if no suitable p/t is encountered, ensures that pol-V-catalysed error-prone DNA synthesis will cease soon after the RecA* filaments supporting the SOS response (and pol V reactivation) are gone. The mutational load associated with pol V function is thereby minimized, effectively restricting it to periods when SOS is induced.

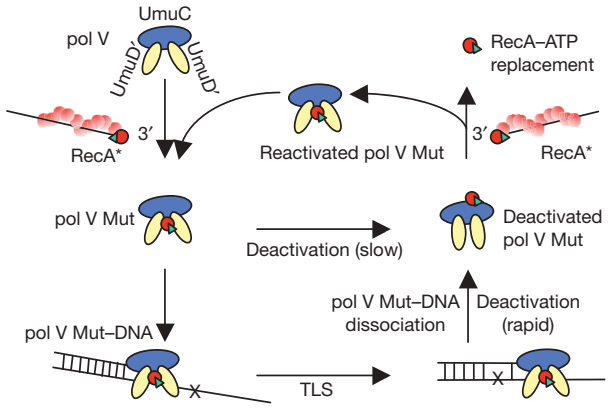


Figure 5 | Model depicting pol V activation, deactivation and reactivation. Pol V (UmuD′$_2$C) is barely active in the absence of RecA*. A molecule of RecA (red circle) and ATP (green triangle) is transferred from the DNA 3′-end of RecA* to form a mutasomal complex (pol V Mut) containing UmuD′$_2$C–RecA–ATP; this catalyses TLS in the absence of RecA*. After TLS, pol V Mut undergoes rapid deactivation upon dissociation from DNA, with RecA and ATP retained in the complex. When pol V Mut is activated, we surmise that RecA–ATP is bound to UmuD′$_2$, and that dissociation from p/t DNA triggers a repositioning of RecA–ATP to bind with UmuC to deactivate pol V Mut. Reactivation of pol V Mut involves a replacement of RecA–ATP from the RecA* 3′-tip. Free pol V Mut undergoes slow deactivation, with RecA and ATP retained in the complex. A damaged DNA base is shown as a black cross in the template strand.

We have shown that transfer from the 3′-tip of RecA* is an absolute requirement for pol V activation (Fig. 4b). Consistent with *in vivo* data[23], the mutagenically inactive RecA F117S is unable to activate pol V for DNA synthesis (Fig. 4b, left gel), despite the transfer of RecA F117S from the 3′-tip of RecA F117S* (Fig. 4a, left gel). Notably, the mutation in RecA F117S is located at the surface facing 3′ on a RecA filament end[20].

In the model (Fig. 5), we show RecA bound to UmuD′$_2$ when pol V Mut is either activated or reactivated, and to UmuC when pol V Mut is deactivated. This assignment is based on our previous identification of two distinct modes of RecA binding to pol V[25]. RecA and pol V form a stably bound complex (dissociation constant $K_d \approx 250$ nM) through the UmuC subunit in the absence of DNA; we have called this 'mode 1' binding[25]. RecA also binds to pol V, with a similar apparent K_d, through its UmuD′$_2$ subunit, requiring DNA and ATP, but not ATP hydrolysis, referred to as 'mode 2' binding[25]. Therefore, we speculate that RecA binds proximally to UmuD′$_2$ during DNA synthesis, and that dissociation from p/t DNA triggers a rapid conformational rearrangement, repositioning RecA to bind with UmuC. When a newly activated pol V Mut is free in solution, the presumed conformational rearrangement shifting RecA binding from UmuD′$_2$ to UmuC occurs much more slowly (Fig. 1g). Subsequent reactivation of pol V Mut (Fig. 1e–g) could occur by a RecA*-mediated replacement of RecA–ATP with the deactivated form of pol V Mut (Fig. 2f).

A decade before the discovery of pol V, the term ''mutasome'' was coined by Echols to define the protein components that assemble at a lesion in the presence of a stalled replication complex to mediate TLS[19]. We now propose that the mutasome envisioned by Echols is a 'stand alone' activated multiprotein complex, pol V Mut, composed of UmuD′$_2$C–RecA–ATP.

METHODS SUMMARY

Pol V was activated by incubation with RecA* immobilized on streptavidin-coated resin. RecA* was removed by centrifugation, leaving pol V Mut in the supernatant. The polymerase activity of pol V or pol V Mut was measured as the rate of deoxynucleotide addition to the 3′-OH end of a hairpin p/t DNA containing a ^{32}P label at its 5′-end. The sequences of template and biotinylated oligomers are shown in Supplementary Fig. 9. Gel band intensities were quantified by phosphorimaging with IMAGEQUANT software (Molecular Dynamics), and primer utilization was calculated from the integrated gel band intensities of extended p/t DNA divided by the total DNA intensity. Deactivated and reactivated His–pol V Mut were bound to a nickel-affinity column and washed twice with reaction buffer. Bound pol V Mut was eluted from the column by washing with buffer containing imidazole. The amount of Fl-RecA and [γ-^{35}S]ATP remaining bound to pol V Mut were visualized by SDS–PAGE or by urea PAGE, respectively. For MALS analysis, pol V Mut was subjected to size-exclusion chromatography using a silica-based column (KW-802.5, Shodex). Chromatography was performed at 0.5 ml min^{-1} and the column effluent was passed directly on-line into a MALS detection system (Dawn Heleos, Wyatt Technology) to determine the molecular masses of eluted species; fractions were collected manually after passing through the light scattering detector. Data analysis was performed with ASTRA software.

Received 6 April; accepted 15 May 2009.

1. Tang, M. J. *et al.* UmuD′$_2$C is an error-prone DNA polymerase, *Escherichia coli* pol V. *Proc. Natl Acad. Sci. USA* **96,** 8919–8924 (1999).
2. Reuven, N. B., Arad, G., Maor-Shoshani, A. & Livneh, Z. The mutagenesis protein UmuC is a DNA polymerase activated by UmuD′, RecA, and SSB and is specialized for translesion replication. *J. Biol. Chem.* **274,** 31763–31766 (1999).
3. Friedberg, E. C. *et al. DNA Repair and Mutagenesis* 2nd edn, 463–555 (ASM Press, 2006).
4. Pham, P., Bertram, J. G., O'Donnell, M., Woodgate, R. & Goodman, M. F. A model for SOS-lesion-targeted mutations in *Escherichia coli*. *Nature* **409,** 366–370 (2001).
5. Becherel, O. J., Fuchs, R. P. & Wagner, J. Pivotal role of the β-clamp in translesion DNA synthesis and mutagenesis in *E. coli* cells. *DNA Repair* **1,** 703–708 (2002).
6. Lenne-Samuel, N., Wagner, J., Etienne, H. & Fuchs, R. P. The processivity factor β controls DNA polymerase IV traffic during spontaneous mutagenesis and translesion synthesis *in vivo*. *EMBO Rep.* **3,** 45–49 (2002).
7. Tang, M. *et al.* Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature* **404,** 1014–1018 (2000).

8. Kato, T. & Shinoura, Y. Isolation and characterization of mutants of *Escherichia coli* deficient in induction of mutations by ultraviolet light. *Mol. Gen. Genet.* **156,** 121–131 (1977).
9. Steinborn, G. *Uvm* mutants of *Escherichia coli* K12 deficient in UV mutagenesis. I. Isolation of *uvm* mutants and their phenotypical characterization in DNA repair and mutagenesis. *Mol. Gen. Genet.* **165,** 87–93 (1978).
10. Bruck, I., Woodgate, R., McEntee, K. & Goodman, M. F. Purification of a soluble UmuD′C complex from *Escherichia coli*: cooperative binding of UmuD′C to single-stranded DNA. *J. Biol. Chem.* **271,** 10767–10774 (1996).
11. Woodgate, R., Rajagopalan, M., Lu, C. & Echols, H. UmuC mutagenesis protein of *Escherichia coli*: purification and interaction with UmuD and UmuD′. *Proc. Natl Acad. Sci. USA* **86,** 7301–7305 (1989).
12. Tang, M. *et al.* Biochemical basis of SOS-induced mutagenesis in *Escherichia coli*: reconstitution of *in vitro* lesion bypass dependent on the UmuD′$_2$C mutagenic complex and RecA protein. *Proc. Natl Acad. Sci. USA* **95,** 9755–9760 (1998).
13. Fujii, S., Gasser, V. & Fuchs, R. P. The biochemical requirements of DNA polymerase V-mediated translesion synthesis revisited. *J. Mol. Biol.* **341,** 405–417 (2004).
14. Nohmi, T., Battista, J. R., Dodson, L. A. & Walker, G. C. RecA-mediated cleavage activates UmuD for mutagenesis: mechanistic relationship between transcriptional derepression and posttranslational activation. *Proc. Natl Acad. Sci. USA* **85,** 1816–1820 (1988).
15. Dutreix, M. *et al.* New *recA* mutations that dissociate the various RecA protein activities in *Escherichia coli* provide evidence for an additional role for RecA protein in UV mutagenesis. *J. Bacteriol.* **171,** 2415–2423 (1989).
16. Sweasy, J. B., Witkin, E. M., Sinha, N. & Roegner-Maniscalco, V. RecA protein of *Escherichia coli* has a third essential role in SOS mutator activity. *J. Bacteriol.* **172,** 3030–3036 (1990).
17. Kuzminov, A. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage λ. *Microbiol. Mol. Biol. Rev.* **63,** 751–813 (1999).
18. Bridges, B. A. & Woodgate, R. Mutagenic repair in *Escherichia coli*. X. The *umuC* gene product may be required for replication past pyrimidine dimers but not for the coding error in UV-mutagenesis. *Mol. Gen. Genet.* **196,** 364–366 (1984).
19. Echols, H. & Goodman, M. F. Mutation induced by DNA damage: a many protein affair. *Mutat. Res.* **236,** 301–311 (1990).
20. Sommer, S., Boudsocq, F., Devoret, R. & Bailone, A. Specific RecA amino acid changes affect RecA–UmuD′C interaction. *Mol. Microbiol.* **28,** 281–291 (1998).
21. Schlacher, K., Cox, M. M., Woodgate, R. & Goodman, M. F. RecA acts in *trans* to allow replication of damaged DNA by DNA polymerase V. *Nature* **442,** 883–887 (2006).
22. Eggler, A. L., Lusetti, S. L. & Cox, M. M. The C terminus of the *Escherichia coli* RecA protein modulates the DNA binding competition with single-stranded DNA-binding protein. *J. Biol. Chem.* **278,** 16389–16396 (2003).
23. Dutreix, M., Burnett, B., Bailone, A., Radding, C. M. & Devoret, R. A partially deficient mutant, *recA1730*, that fails to form normal nucleoprotein filaments. *Mol. Gen. Genet.* **232,** 489–497 (1992).
24. Wyatt, P. J. Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta* **272,** 1–40 (1993).
25. Schlacher, K. *et al.* DNA polymerase V and RecA protein, a minimal mutasome. *Mol. Cell* **17,** 561–572 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by National Institutes of Health grants to M.F.G. (ES12259; R37GM21422) and M.M.C. (GM32335), and funds from the NICHD/NIH Intramural Research Program to K.K. and R.W. The authors thank R. Britt for preparation of the RecA E38KΔC17 protein used in this study. MALS data were collected using the USC NanoBiophysics Core Facility, with the aid and cooperation of N. Chelyapov. We thank J. Bertram for his help in performing the MALS experiment and analysing the data. We thank J. Petruska for his comments.

Author Contributions Q.J. performed the experiments and experimental analyses; K.K. performed the construction and purification of His-tagged pol V; M.M.C. provided the purified RecA E38KΔC17 protein; Q.J. and M.F.G. designed the experiments with input from M.M.C. and R.W.; M.F.G., Q.J., M.M.C. and R.W. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.F.G. (mgoodman@usc.edu).

Vol 460|16 July 2009|doi:10.1038/nature08094

LETTERS

Contamination of the asteroid belt by primordial trans-Neptunian objects

Harold F. Levison[1,2], William F. Bottke[1,2], Matthieu Gounelle[3], Alessandro Morbidelli[4], David Nesvorný[1,2] & Kleomenis Tsiganis[5]

The main asteroid belt, which inhabits a relatively narrow annulus ~2.1–3.3 AU from the Sun, contains a surprising diversity of objects ranging from primitive ice–rock mixtures to igneous rocks. The standard model used to explain this assumes that most asteroids formed *in situ* from a primordial disk that experienced radical chemical changes within this zone[1]. Here we show that the violent dynamical evolution of the giant-planet orbits required by the so-called Nice model[2–4] leads to the insertion of primitive trans-Neptunian objects into the outer belt. This result implies that the observed diversity of the asteroid belt is not a direct reflection of the intrinsic compositional variation of the proto-planetary disk. The dark captured bodies, composed of organic-rich materials, would have been more susceptible to collisional evolution than typical main-belt asteroids. Their weak nature makes them a prodigious source of micrometeorites—sufficient to explain why most are primitive in composition and are isotopically different from most macroscopic meteorites[5,6].

Our result that a significant fraction of objects in the main asteroid belt are captured primordial trans-Neptunian objects (hereafter termed 'comets' or 'comet-like' to reflect their primitive, icy natures) comes from a series of numerical calculations of the trajectories of small bodies during dynamical events proposed in the Nice model[2,4]. In this model, the giant planets are assumed to have formed in a compact configuration (all were located between 5 and 15 AU from the sun) and to be surrounded by a $\sim 35 M_{\oplus}$ disk of comet-like planetesimals stretching between ~16 and ~30 AU from the sun ($M_{\oplus}$ denoting the mass of the Earth). After roughly 600 Myr, the orbits of the giant planets became unstable. Uranus and Neptune were gravitationally scattered outwards, thereby penetrating the disk and scattering its constituents throughout the Solar System. The planets evolved into their current orbits as a result of the gravitational effects of the 'comets'. This model is unique because it can quantitatively explain many of the observed characteristics of the Solar System (for example giant-planet orbits[2], Trojan asteroids[2,7], the Kuiper belt[8], the irregular satellites[9] and the origin of the so-called late heavy bombardment[4]).

Our current simulations follow the dynamical evolution of objects that originally formed in the Nice model's primordial comet disk as the giant planets' orbits evolved. We are primarily interested in the fate of objects that potentially could have entered the inner Solar System. Thus, we first integrated the orbits of a large number of massless planetesimals initially on Saturn-crossing orbits under the gravitational influence of the Sun, Jupiter and Saturn. The planets were forced to migrate by including a suitably chosen acceleration in the planets' equations of motion, so that they reproduced the evolution of the 'fast migration' run in ref. 3 (see Supplementary Information, section S1, for more details). During the integration, we supplied a steady flux of planetesimals through the Jupiter–Saturn system. These objects represent the planetesimals that originally formed in the trans-planetary disk, but were destabilized and fed inwards by Uranus and Neptune. The integrations covered 10 Myr. In all, we followed the evolution of ~31,000,000 disk particles. We find that a significant number of objects are captured either in Trojan orbits (as in ref. 3) or in orbits inside that of Jupiter. A Jupiter-crossing object can evolve into a stable orbit in the main asteroid belt through a combination of dynamical processes. Initially, Jupiter will scatter objects and occasionally one will fall in a mean-motion resonance, which can lower eccentricity. This would be temporary if there were no migration. With migration, however, the resonances move, and low-eccentricity objects can be left behind.

Next we tracked the long-term evolution of this captured comet-like population to compare it with observations. This population is affected by two processes: dynamical erosion and collisional grinding. To address the first, we integrated the orbits of the implanted comets under the gravitational effects of the Sun and gas giants for 4 Gyr. We removed an object if it either collided with a planet or reached a heliocentric distance greater than 15 AU or less than 1.5 AU. Figure 1 shows a comparison between the surviving objects (red dots) and a complete sample of known large asteroids (green pluses). We find that a substantial population of Trojans (32 of the 230 captured objects, or 13%) and Hilda asteroids (8%) are produced.

In addition, we captured comets in orbits in the main asteroid belt with semi-major axes as low as 2.68 AU. To compare these objects with the rest of the main belt, we used classification based on spectroscopy. According to our model, objects captured in the main belt should physically be similar to the resonant Trojans and Hildas. The vast majority of these resonant objects are classified as D- or P-type (hereafter D/P-type), which probably are organic-rich. Indeed, they are a good match to spectral characteristics of the observed dormant comets[10]. This is consistent with the Nice model's prediction that they formed beyond 15 AU.

The black dots in Fig. 1 show the D-type asteroids taken from the databases of refs 11, 12. There is reasonable agreement between the orbital element distribution of our captured objects and the known primitive asteroids (Trojans, Hildas and D-type main-belt objects). This result naturally leads to the controversial idea that most primitive asteroids formed beyond ~15 AU.

If the above model is correct, we need not only to be able to reproduce the orbits of the primitive, comet-like asteroids, but their total number and size distribution as well. To accomplish this, we need to account for their collisional evolution, which we do with a self-consistent code, CoDDEM, capable of following the collisional

[1]Southwest Research Institute, 1050 Walnut Street, Suite 300, [2]Center for Lunar Origin & Evolution, NASA Lunar Science Institute, Boulder, Colorado 80302, USA. [3]Laboratoire de Minéralogie et de Cosmochimie du Muséum, CNRS & Muséum National d'Histoire Naturelle, 75005 Paris, France. [4]Observatoire de la Côte d'Azur, Nice, Cedex 4, F-06304, France. [5]Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, 54006 Hellas, Greece.

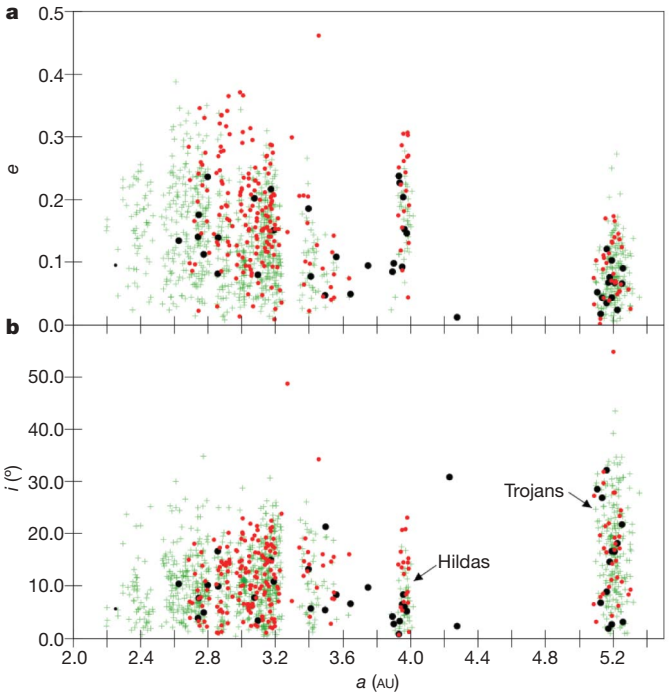


Figure 1 | The orbital element distributions of real and modelled asteroids. The distributions of eccentricity, e (**a**), and inclination, i (**b**), as functions of semi-major axis, a, of asteroids in Hilda (objects which orbit in the 2:3 mean-motion resonance with Jupiter near 3.9 AU), Trojan and main-belt populations. The small green pluses show all the numbered objects with diameters greater than 40 km (the asteroid belt is probably complete at these sizes) in the International Astronomical Union's Minor Planet Center database. We calculated the diameters from published absolute magnitudes, assuming the albedos (Supplementary Information, section S2.1). The black symbols show the D-type asteroids as catalogued in refs 11 and 12. It is important to note that asteroid (336) Lacadiera, at $a = 2.25$ AU, $e = 0.1$ and $i = 5.6°$, which is classified as a D-type in ref. 11, has an unusual spectrum (H. Campins, personal communication), and thus is probably a different type of object. We therefore use a smaller symbol to plot the location of this object and did not include it in our analysis. We only include D-type asteroids here because it is difficult to distinguish P-types from other, more processed, asteroids, with the result that the catalogue of P-types probably suffers from significant contamination[30]. The red dots show the location of objects captured during our simulations. The main result of our simulations is that, in addition to the resonant asteroids, a significant number of objects are trapped in the outer main belt (OMB). Unfortunately, it is not possible to perform a direct comparison between the orbital element distribution of our trapped objects and that of the D-types because the observations are biased owing to selection criteria, asteroid families and the like. However, we find noteworthy the fact that the inner edge of each populations is at $a \approx 2.6$ AU.

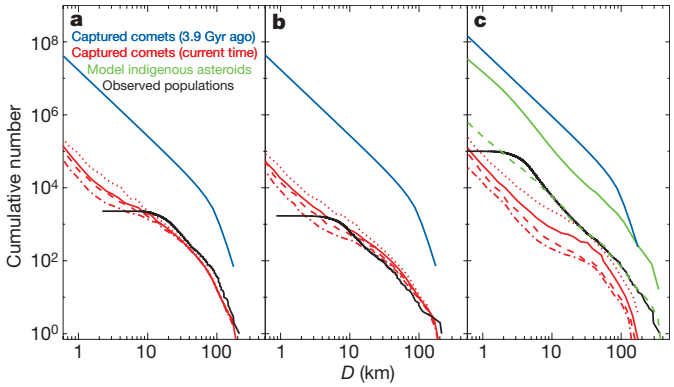


Figure 2 | The beginning and end states of the SFDs in the three regions studied. **a**, Trojans; **b**, Hildas; **c**, OMB objects. The blue curves show the initial conditions for the captured comets in our CoDDEM simulations. The shapes of the initial SFDs for these populations were taken from the currently observed Trojan SFD. We chose the Trojans because these objects have been least affected by collisional grinding. In particular, using the cumulative number $N(>D) \propto D^{-q}$, we assigned values of $q = 5.5$ for $D > 105$ km and $q = 1.8$ for $D < 105$ km, where D is diameter. The overall scales of the SFDs for the captured populations are derived from a combination of the observed Trojans and the results of the 10-Myr migration simulations (Supplementary Information, section S2.2). For each value of f_Q, we performed ~10 simulations using different random number generator seeds. The red curves show an average of these simulations and thus are our predictions for the present-day SFDs of the captured objects. In particular, the dotted, solid, dashed, and dash–dot red curves correspond to $f_Q = 1$, 3, 5 and 8, respectively. For the Hildas and Trojans, we expect the model SFD to match observations (black curves; note that the roll-off at small sizes is due to observational incompleteness). The agreement is quite good given the uncertainties in our models (Supplementary Information). Of particular note, the Trojans are known to have a significantly steeper SFD at $D \gtrsim 100$ km than the Hildas. Our model reproduces this because the Hildas, which cross the main belt, undergo more collisional grinding. For the main belt, the model also predicts that the SFD of the primitive bodies is steeper than that of the indigenous population for $D \gtrsim 20$ km, so the fraction of D/P-types should increase as diameter decreases. This appears to match observations[21]. The solid and dotted green curves in **c** refer to the indigenous OMB objects at the beginning and end of the simulation, respectively.

evolution and dynamical depletion of multiple interacting size-frequency distributions (SFDs)[13,14]. This code was modified to use an improved algorithm for handling the outcome of individual collisions, taken from ref. 15. We tracked five populations: (1) indigenous main-belt objects ($a < 2.82$ AU), (2) indigenous OMB objects, (3) captured OMB objects, (4) Hildas and (5) Trojans. The collision probabilities and impact velocities with both themselves and each other were computed from the observed objects or, in the case of the Trojans, were taken from refs 16–18. Populations 1 and 2, which represent native asteroids, had a bulk density of 2.7 g cm^{-3} and were assumed to follow the disruption scaling law for rocky objects[13]. We also assumed that the indigenous main-belt SFDs had approximately the same shape 3.9 Gyr ago as they do today[13,14].

The captured populations (3, 4 and 5) have a bulk density of 1 g cm^{-3}. We assumed that the initial shape of their SFDs was the same as that of the currently observed Trojans (Fig. 2 and Supplementary Information, section S2.2). Because the cometary disruption scaling law is not well understood, we performed four series of simulations in which this parameter was varied. In particular, following ref. 19, we assumed that the amount of energy required to catastrophically disrupt a comet is the same as that for solid ice[20], divided by a factor f_Q (Supplementary Information, section S2.2). We studied $f_Q = 1$, 3, 5 and 8 (the limit found in ref. 19). The required dynamical depletion rates were taken from the dynamical simulations described above, whereas those of the indigenous asteroids came from existing calculations of asteroid-belt depletion[4] in the Nice model (see the discussion about population 3 in Supplementary Information, section S1).

The results of our collisional calculations are shown in Fig. 2. Our model reproduces the SFDs of both the Trojan and Hilda populations for diameter $D > 40$ km fairly well. This agreement is important because each population has unique characteristics that test our model's assumptions. For example, comminution in the Trojan population comes mainly from members hitting one other, whereas that in the Hilda population is driven by impactors from across the asteroid belt. Matching the SFDs of these populations simultaneously in the same model, therefore, increases our confidence that our assumptions are reasonable.

The final state of the captured cometary OMB is strongly dependent on the disruption scaling laws used. For $f_Q = 1$, the model predicts that all of the OMB objects with $D > 40$ km are captured. For $f_Q = 3$, 5 and 8 we find that captured objects represent 48%, 20% and 16% of the OMB, respectively. This decrease is due to a much larger fraction of the captured objects being destroyed by impacts than their stronger indigenous counterparts. The best available bias-corrected

estimates of the fraction of D/P-types within the OMB over the same size range is ~20% (refs 21–23), suggesting that $f_Q \gtrsim 5$. Because the proportion of D/P-types is highly uncertain, this limit for f_Q is probably lower. The requirement that the primordial cometary disk survives the ~600 Myr before the late heavy bombardment requires that $f_Q < 5$ (Supplementary Information, section S4), which is consistent with our upper range. We conclude, therefore, that captured comets are weaker than native asteroids.

If the implanted comets are weaker than the indigenous main-belt asteroids, we can potentially resolve another long-standing problem in planetary science. The micrometeorites collected on the Earth are dominated by material reminiscent of carbonaceous chondrites. Indeed, the ratio of ordinary chondrite-type material to carbonaceous chondrite-type material (hereafter O/C) in micrometeorites is $\lesssim 0.16$ (ref. 24). Yet the asteroid belt, which has traditionally been considered the dominant source of this material[25] (see Supplementary Information, section S3, for alternatives), is roughly an equal mix of ordinary chondrite-like asteroids (for example S-types) and carbonaceous chondrite-like asteroids (for example C-types). The question has thus been that of why the micrometeorites we collect do not reflect this ratio.

In our successful collisional models (that is, those with $f_Q \geq 3$), we find that, at the current epoch, the implanted comets produce more than 15 times more dust per unit mass than the indigenous asteroids (the exact value depending on f_Q). This is because the implanted population is made up of weaker objects that are breaking up more easily. Furthermore, because the material created is mineralogically similar to that produced by primitive C-type asteroids, our model can explain the overabundance of carbonaceous chondrite micrometeorites. Indeed, collisional simulations of the current asteroid belt (Supplementary Information, section S3) show that an observed value of $O/C < 0.16$ is a common occurrence if $f_Q \geq 3$.

The idea that D/P-types are both weak and the source of micrometeorites also provides an explanation for the surprising fact that there are few micrometeorite analogues in our macroscopic meteorite collection. If we suppose that these micrometeorites are generated by the collisional cascade of embedded comets, we would not expect to see many macroscopic samples for two reasons. First, our implanted population resides in a region of the main belt from where it is very difficult to get meteorites[26,27]. In particular, macroscopic objects leaving this region of the asteroid belt are most likely to be ejected from the Solar System by Jupiter, whereas microscopic bodies, which are more susceptible to radiation forces, are more likely to be delivered to the Earth. In addition, the presumed weak nature of the captured objects probably makes their macroscopic fragments unlikely to survive the passage through the Earth's atmosphere or, if they do, survive for long on the Earth's surface.

Our conclusion that most primitive asteroids formed at large heliocentric distances and then were trapped in their current orbits, coupled with the idea that at least some asteroids were scattered into the main belt from the terrestrial-planet region[28], changes our view of the asteroid belt. The traditional interpretation of the diversity of the asteroid belt is that it represents the original condensation sequence in the protoplanetary disk[1]. Indeed, their orbital distribution has been used to constrain both the thermal structure of the nebula and the effectiveness of various heating mechanisms as a function of heliocentric distance (for example the decay of short-lived radionuclides such as ^{26}Al)[29]. If most D/P-types in the inner solar system were captured from farther out, however, our models of the protoplanetary disk would have to be significantly revised. Indeed, the diversity in the asteroid belt may be telling us more about the dynamical processes that controlled planet formation than about the physical nature of the protoplanetary nebula, with the main asteroid belt a collection point for rogue planetesimals from across the Solar System.

Received 3 February; accepted 24 April 2009.

1. Bell, J. F., Davis, D. R., Hartmann, W. K. & Gaffey, M. J. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 921–945 (Univ. Arizona Press, 1989).
2. Tsiganis, K., Gomes, R. S., Morbidelli, A. & Levison, H. F. Origin of the orbital architecture of the giant planets of the Solar System. *Nature* **435,** 459–461 (2005).
3. Morbidelli, A., Levison, H. F., Tsiganis, K. & Gomes, R. Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature* **435,** 462–465 (2005).
4. Gomes, R. S., Levison, H. F., Morbidelli, A. & Tsiganis, K. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435,** 466–469 (2005).
5. Matrajt, G. *et al.* Oxygen isotope measurements of individual unmelted Antarctic micrometeorites. *Geochim. Cosmochim. Acta* **70,** 4007–4018 (2006).
6. Clayton, R. N. Oxygen isotopes in meteorites. *Annu. Rev. Earth Planet. Sci.* **21,** 115–149 (1993).
7. Nesvorný, D. & Vokrouhlický, D. Chaotic capture of Neptune Trojans. *Astron. J.* **137,** 5003–5011 (2009).
8. Levison, H. F., Morbidelli, A., Van Laerhoven, C., Gomes, R. & Tsiganis, K. Origin of the structure of the Kuiper belt during a dynamical instability in the orbits of Uranus and Neptune. *Icarus* **196,** 258–273 (2008).
9. Nesvorný, D., Vokrouhlický, D. & Morbidelli, A. Capture of irregular satellites during planetary encounters. *Astron. J.* **133,** 1962–1976 (2007).
10. Licandro, J., de León, J., Pinilla, N. & Serra-Ricart, M. Multi-wavelength spectral study of asteroids in cometary orbits. *Adv. Space Res.* **38,** 1991–1994 (2006).
11. Tholen, D. J. & Barucci, M. A. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 298–315 (Univ. Arizona Press, 1989).
12. Bus, S. J. & Binzel, R. P. Phase II of the Small Main-Belt Asteroid Spectroscopic Survey. A feature-based taxonomy. *Icarus* **158,** 146–177 (2002).
13. Bottke, W. F. *et al.* The fossilized size distribution of the main asteroid belt. *Icarus* **175,** 111–140 (2005).
14. Bottke, W. F. *et al.* Linking the collisional history of the main asteroid belt to its dynamical excitation and depletion. *Icarus* **179,** 63–94 (2005).
15. Morbidelli, A., Bottke, W. F., Nesvorný, D. & Levison, H. F. Asteroids were born big. *Icarus*. (submitted).
16. Jewitt, D. C., Trujillo, C. A. & Luu, J. X. Population and size distribution of small Jovian Trojan asteroids. *Astron. J.* **120,** 1140–1147 (2000).
17. Yoshida, F. & Nakamura, T. Size distribution of faint Jovian L4 Trojan asteroids. *Astron. J.* **130,** 2900–2911 (2005).
18. Szabó, G. M., Ivezić, Ž., Jurić, M. & Lupton, R. The properties of Jovian Trojan asteroids listed in SDSS Moving Object Catalogue 3. *Mon. Not. R. Astron. Soc.* **377,** 1393–1406 (2007).
19. Leinhardt, Z. M. & Stewart, S. T. Full numerical simulations of catastrophic small body collisions. *Icarus* **199,** 542–559 (2009).
20. Benz, W. & Asphaug, E. Catastrophic disruptions revisited. *Icarus* **142,** 5–20 (1999).
21. Mothé-Diniz, T., Carvano, J. M. & Lazzaro, D. Distribution of taxonomic classes in the main belt of asteroids. *Icarus* **162,** 10–21 (2003).
22. Carvano, J. M., Mothé-Diniz, T. & Lazzaro, D. Search for relations among a sample of 460 asteroids with featureless spectra. *Icarus* **161,** 356–382 (2003).
23. Burbine, T. H. *et al.* Oxygen and asteroids. *Rev. Mineral. Geochem.* **68,** 273–343 (2008).
24. Genge, M. J. Ordinary chondrite micrometeorites from the Koronis asteroids. *Proc. Lunar Planet. Sci. Conf.* **37,** abstr. 1759 (2006).
25. Dermott, S. F., Durda, D. D., Grogan, K. & Kehoe, T. J. J. in *Asteroids III* (eds Bottke, W. F., Paolicchi, P., Binzel, R. P. & Cellino, A.) 423–442 (Univ. Arizona Press, 2002).
26. Morbidelli, A. & Gladman, B. Orbital and temporal distributions of meteorites originating in the asteroid belt. *Meteorit. Planet. Sci.* **33,** 999–1016 (1998).
27. Bottke, W. F. Jr, Rubincam, D. P. & Burns, J. A. Dynamical evolution of main belt meteoroids: numerical simulations incorporating planetary perturbations and Yarkovsky thermal forces. *Icarus* **145,** 301–331 (2000).
28. Bottke, W. F., Nesvorný, D., Grimm, R. E., Morbidelli, A. & O'Brien, D. P. Iron meteorites as remnants of planetesimals formed in the terrestrial planet region. *Nature* **439,** 821–824 (2006).
29. Grimm, R. E. & McSween, H. Y. Heliocentric zoning of the asteroid belt by aluminum-26 heating. *Science* **259,** 653–655 (1993).
30. Clark, B. E., Rivkin, A. S., Bus, S. J., Sanders, J. & X. E, M, and P-type asteroid spectral observations. *Bull. Am. Astron. Soc.* **35,** 955 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements H.F.L., W.F.B., A.M. and D.N. are grateful to NASA's Origins of Solar Systems and Outer Planet Research programmes. D.N. is also grateful to the US National Science Foundation's Astronomy & Astrophysics Grants programme for funding. M.G. wishes to thank Centre National de la Recherche Scientifique, Programme National de Planétologie and the European Community for funding. We thank S. Stewart and Z. Leinhardt for discussions and S. Weidenschilling and A. Harris, who acted as referees.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.F.L. (hal@boulder.swri.edu).

LETTERS

Manipulation of photons at the surface of three-dimensional photonic crystals

Kenji Ishizaki[1,2] & Susumu Noda[1,2]

In three-dimensional (3D) photonic crystals[1–12], refractive-index variations with a periodicity comparable to the wavelength of the light passing through the crystal give rise to so-called photonic bandgaps, which are analogous to electronic bandgaps for electrons moving in the periodic electrostatic potential of a material's crystal structure. Such 3D photonic bandgap crystals are envisioned to become fundamental building blocks for the control and manipulation of photons in optical circuits. So far, such schemes have been pursued by embedding artificial defects[3–5,8–12] and light emitters[4–9] inside the crystals, making use of 3D bandgap directional effects. Here we show experimentally that photons can be controlled and manipulated even at the 'surface' of 3D photonic crystals, where 3D periodicity is terminated, establishing a new and versatile route for photon manipulation. By making use of an evanescent-mode coupling technique, we demonstrate that 3D photonic crystals possess two-dimensional surface states, and we map their band structure. We show that photons can be confined and propagate through these two-dimensional surface states, and we realize their localization at arbitrary surface points by designing artificial surface-defect structures through the formation of a surface-mode gap. Surprisingly, the quality factors of the surface-defect mode are the largest reported for 3D photonic crystal nanocavities (*Q* up to ~9,000). In addition to providing a new approach for photon manipulation by photonic crystals, our findings are relevant for the generation and control of plasmon-polaritons in metals and the related surface photon physics. The absorption-free nature of the 3D photonic crystal surface may enable new sensing applications and provide routes for the realization of efficient light–matter interactions.

Surface states in 3D photonic crystals were initially discussed theoretically with a view to preventing them from disturbing the optical properties of embedded artificial defects[13]. If we could instead make use of the surface states positively to manipulate photons, it would open up a new route for the manipulation of photons by 3D photonic crystals, through which various types of access from outside would become straightforward compared with those formed inside the 3D photonic crystals. Such studies would simultaneously lead to the establishment of the surface science of 3D photonic crystals, and have great relevance to the surface plasmon-polariton effect of metals and the related surface photonic physics. However, experimental investigations of the manipulation of photons at the surface of 3D photonic crystals had not yet been performed. Here we demonstrate the presence of surface states at the surface of 3D photonic crystals, and show that the confinement and propagation of photons via the surface states is indeed possible. We also control the surface states, and demonstrate that the photons can be localized at desired and arbitrary surface positions by the formation of the surface-mode gap and the introduction of artificial localized surface-defect states.

The 3D photonic crystals we investigated have a stacked-stripe (or woodpile) structure (Fig. 1a) with a full 3D photonic bandgap[3,14,15]. The in-plane band structures of the 3D photonic crystals without and with the 'surface' are shown in Fig. 1b and c. The width and thickness of the individual dielectric rods were both set at 0.4*a*, where *a* is the centre-to-centre separation of the parallel dielectric rods. It is apparent from Fig. 1c (see red lines) that the surface states are formed in the frequency region where the bandgap would otherwise exist (Fig. 1b). Owing to the structural anisotropy for the *x*-direction and the *y*-direction at the surface (see inset in Fig. 1a), the dispersion relation of the surface states are different for the Γ–X_1 direction and the Γ–X_2 direction, respectively. The inset in Fig. 1c shows an example of the electromagnetic-field distribution of a particular surface state at the blue point. This confirmed that the photon mode is indeed formed at the surface of the 3D photonic crystal.

To demonstrate experimentally the presence of the surface states, we fabricated 3D photonic crystals with gallium arsenide (GaAs)-based eight-stacked-stripe layers[3,16,17], in which *a* was set at 500 nm, and the width and thickness of the individual dielectric rods were both set at 200 nm (=0.4*a*). Figure 2a (left-hand side) shows a scanning

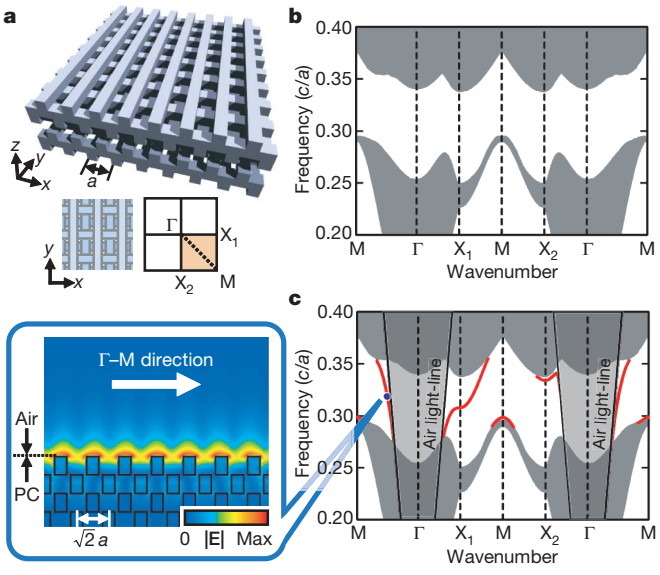


Figure 1 | 3D photonic crystal and band structures. a, Schematic structure of the 3D photonic crystals, and representative directions in real and reciprocal spaces. Owing to the anisotropy of the surface structure for the *x*-direction and the *y*-direction, we need to distinguish the Γ–X_1 direction and the Γ–X_2 direction. **b**, **c**, In-plane band structures for the 3D photonic crystal without (**b**) and with (**c**) 'surfaces'. The width and thickness of component dielectric rods are both set at 0.4*a*. The inset in **c** shows an example of the electric-field distribution **E** of a surface state with an in-plane wavenumber $k_{//} = 0.25\sqrt{2}\times(2\pi/a)$ for the Γ–M direction and a frequency of $f = 0.318c/a$ (blue circle). PC, photonic crystal.

[1]Department of Electronic Science and Engineering, Kyoto University, [2]Japan Science and Technology Agency, Kyoto 615-8510, Japan.

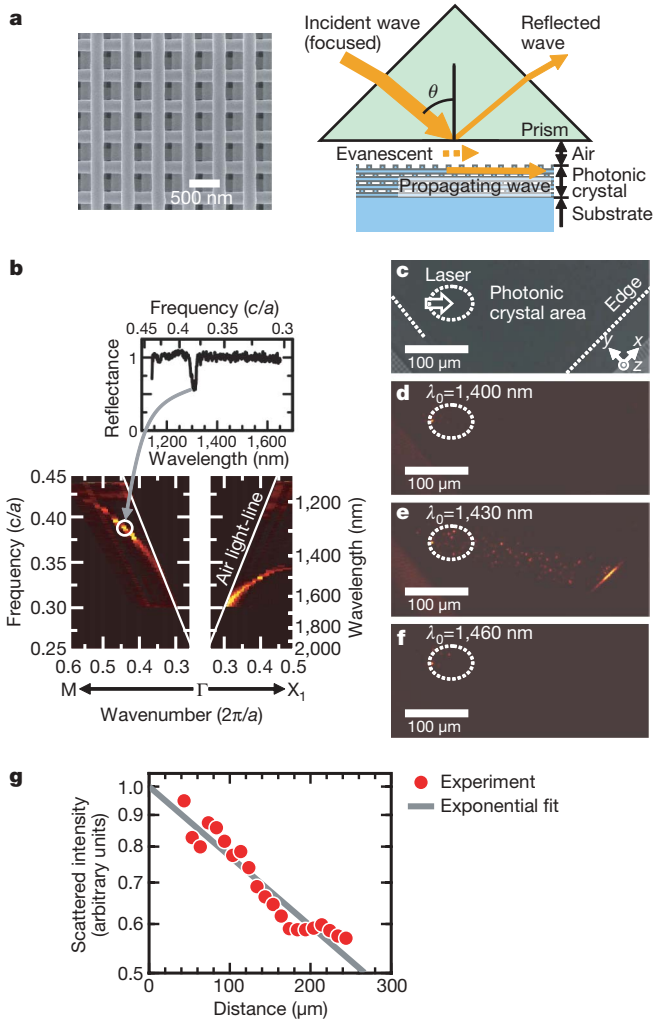


Figure 2 | Experimental demonstrations of the surface states of 3D photonic crystals. **a**, Scanning electron microscope image of the fabricated 3D photonic crystal (left panel) and the experimental set-up used to couple light into the surface modes (right panel). **b**, Measured dispersion characteristics of the surface modes for two representative directions, Γ–M and Γ–X_1. The inset shows one example of the reflection spectrum at $\theta = 52.3^\circ$ for the Γ–M direction, in which the minimum value of the reflection spectrum at $f = 0.382c/a$ corresponds to the white circle in the obtained dispersion curve. **c**, Optical microscope image of the surface of the 3D photonic crystal, showing the excitation point of the incident light and the edge of the photonic crystal. **d**–**f**, Experimental results for light propagation through the surface mode, in which the irradiating angle θ was fixed at 45.7°, and the wavelength was changed to 1,400 nm (**d**), 1,430 nm (**e**) or 1,460 nm (**f**). **g**, Scattered intensity at the edge of the 3D photonic crystal as a function of the distance between the excitation point and the crystal edge.

electron microscope image of the fabricated 3D photonic crystal. We initially measured the transmission spectra of the sample for various incident angles, and confirmed that the fabricated photonic crystals had full 3D bandgaps at wavelengths of 1,300–1,600 nm (see Supplementary Information). We then measured the band structure of the surface states in the 3D photonic crystal. When photonic bands are formed above the air light-line, direct irradiation of light to the photonic crystals[18–20] can be used. However, in the present case, as the surface states are formed below the air light-line, the coupling of light to the surface states is difficult using such a direct irradiation method. Thus, we used an evanescent-coupling method with a prism[21–24] to excite the surface states (Fig. 2a, right-hand side).

Light with wavelength λ_0 in vacuum was introduced to the bottom face of the prism with a refractive index (n_{prism}) of 1.5 at an angle of θ, which was larger than the critical angle of the total internal reflection. In this situation, the light is completely reflected but the evanescent wave can penetrate the bottom face of the prism. We then placed the photonic crystal below the bottom surface of the prism at a distance of ~500 nm, so that evanescent coupling of the light from the prism to the surface modes of the photonic crystal was achieved. The in-plane wavenumber $k_{//}$ of the coupled evanescent wave can be expressed as $n_{prism}k_0\sin\theta$, where k_0 is the wavenumber in the vacuum ($=2\pi/\lambda_0$). When the evanescent wave is coupled to the surface mode of the 3D photonic crystal, the optical power reflected at the bottom of the prism is reduced owing to the energy transfer to the surface mode of the 3D photonic crystal. By measuring the reflection spectra for various incident angles of θ and finding the frequency at which the minimum reflection is realized, we can derive the dispersion characteristics of the surface states for various photonic-crystal directions. Figure 2b shows the measured dispersion characteristics of the surface modes for two representative directions: Γ–M and Γ–X_1. The inset in Fig. 2b shows one example of the reflection spectrum at $\theta = 52.3^\circ$ for the Γ–M direction, where the minimum value at frequency $f = 0.382c/a$, where c is the velocity of light in vacuum, corresponded to the white circle in the obtained dispersion curve. The measured results agreed with the theoretical dispersion curve (Fig. 1c). These results clearly indicate that surface states are indeed formed in the 3D photonic crystal.

Next, we investigated whether photons could be confined and propagate through the surface modes of the 3D photonic crystals. We directly observed the propagation of photons via the surface state of the 3D photonic crystal. The surface state was excited by the evanescent-coupling method described above except that we placed the prism below the substrate of the photonic crystal and the light was irradiated from the bottom side. Figure 2c shows an optical microscope image of the photonic-crystal surface observed from the top side using an infrared camera. The incident angle θ was fixed at 45.7°, and the wavelength was changed. The measured dispersion curve (Fig. 2b) suggested that light with a specific wavelength of ~1,430 nm could be coupled to the surface mode and propagate. Figure 2d–f shows the experimental results. Light with a wavelength of 1,430 nm propagated through the surface of the 3D photonic crystal, and was scattered at the end and at the surface (the surface scattering being caused by minor, unintentional irregularities of the surface of the fabricated crystal) (Fig. 2e). In contrast, light with wavelengths of 1,400 and 1,460 nm could not propagate because of the lack of surface modes for the incident angle $\theta = 45.7^\circ$ (Fig. 2d and f). When the incident angle θ changed, the wavelength of light that could propagate also changed.

To estimate the propagation length via the surface states, the intensity of scattered light at the end of the photonic crystal was measured by changing the distance between the excitation point and the crystal edge. The intensity of scattered light exponentially decreased as the distance between the excitation point and the crystal edge increased (Fig. 2g). By evaluating the length at which the intensity of scattered light became $1/e$, where e is the base of the natural logarithm, the propagation length was estimated to be ~385 μm. This is reasonable, given that the stacked 3D photonic crystal had eight layers.

These experiments clearly indicate that photons can indeed be confined and propagate through surface states. However, the surface states were not spatially controlled, so photons inevitably spread in various in-plane directions. To manipulate photons arbitrarily at the surface of 3D photonic crystals, it is important to realize the spatial control and localization of photons at desired surface positions. To achieve this, we formed a surface-mode gap, by which the in-plane propagation of photons can be inhibited. We could then form the desired localized state at a specific position by introducing artificial surface defects. The theoretical calculation revealed that changing the dielectric arrangement of the surface layer of the 3D photonic crystal from a striped to a crossed geometry could achieve a surface-mode gap (Fig. 3a). We therefore fabricated 3D photonic crystals with modified surface structures in which the additional rod widths were

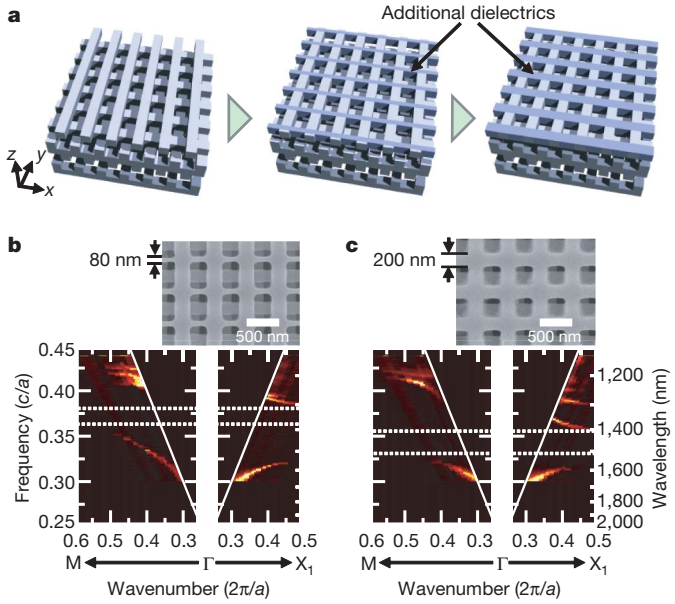


Figure 3 | Formation of a surface-mode gap by modification of the surface structure. **a**, Schematic showing the modification of the surface structure from a striped geometry to a crossed geometry by adding additional dielectrics. **b**, **c**, The measured dispersion characteristics for the photonic crystals with a crossed geometric surface. The widths of the additional rods were 80 nm (=0.16*a*) (**b**) and 200 nm (=0.4*a*) (**c**).

set at 80 and 200 nm, respectively (Fig. 3b and c). The measured dispersion characteristics are shown in Fig. 3b and c, respectively. It is clear that adding dielectric rods successfully opened the surface-mode gap (Fig. 3b and c, white dashed lines). The experimental results agreed well with the theoretical calculation. It is also noteworthy that the surface-mode gap occurred for both polarizations of the transverse electric and the transverse magnetic modes for any in-plane direction.

Our next step was the introduction of an artificial surface-defect structure. For this purpose, we partially increased the width of the dielectric rod, as shown in Fig. 4a. The calculated electric-field distribution of the defect mode revealed that a localized photon mode can be formed at the surface of the 3D photonic crystal (Fig. 4b). Hence, we fabricated the surface-defect structure shown in the inset in Fig. 4c, and measured the optical properties of the defect mode using the method shown, in which the excitation light was incident to the surface defect with an angle of 45° from the normal direction using a 50× objective lens (numerical aperture = 0.42), and the radiation from the defect was detected by the same lens. An example of the micro-mapping of the response of the artificial defect mode is shown in Fig. 4d. The spatially localized photon state indeed formed at the surface. Figure 4e shows the series of resonant spectra from the artificial surface defects with various lengths L_d. The surface defect had a clear resonance curve, and possessed fundamental and higher modes. The resonant wavelengths shifted to the longer wavelength side as the defect length L_d increased. These results agreed well with the theoretical calculation. Figure 4f shows the *Q* factors of individual defect modes. Surprisingly, the highest *Q* factor reached ~9,000, which is the highest value reported for 3D photonic crystals. Theoretically, increasing the stacking number of the 3D photonic crystals above eight layers could produce *Q* factors exceeding 100,000.

This work represents an important step in realizing a new route for the manipulation of photons by 3D photonic crystals, as well as establishing the surface science of 3D photonic crystals. The experimental demonstration of a mode gap in the surface states has other important implications, such as the protection of interior localized states and the efficient coupling of light from an external medium to a waveguide within the 3D photonic crystal without loss to undesirable

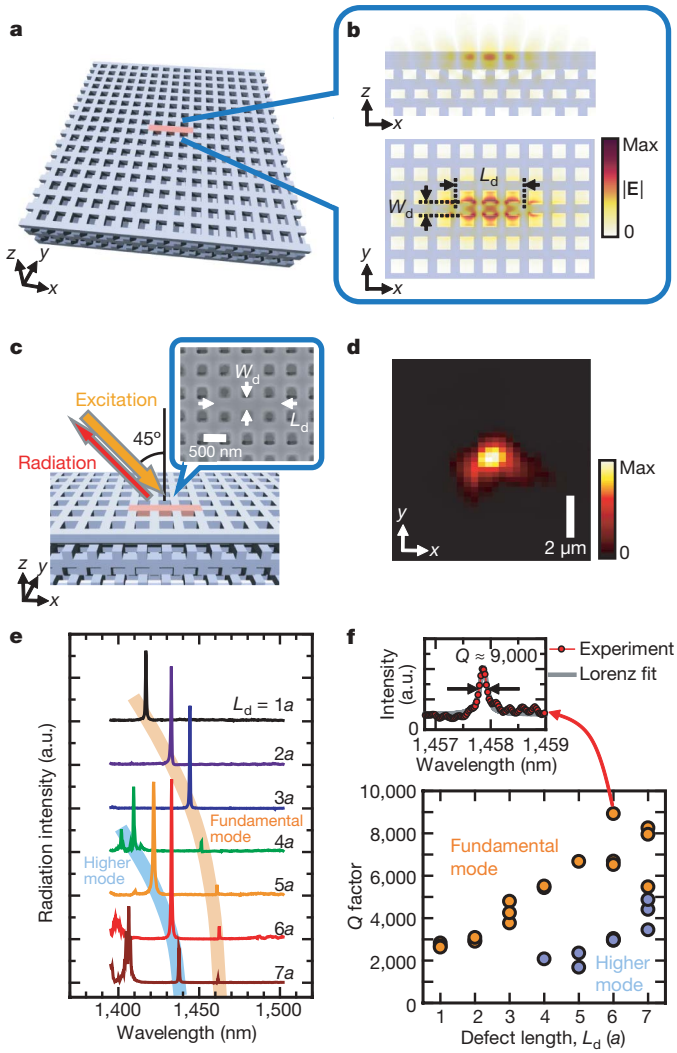


Figure 4 | Artificial surface-defect structure and its characteristics. **a**, Schematic example of a surface-defect structure, in which the width of the dielectric rod was partially widened. **b**, Electric-field distribution **E** of the surface-defect mode, in which the width W_d and the length L_d of the widened dielectric rod were 0.52*a* and 3*a*, respectively. **c**, Set-up for measuring radiation spectra from surface-defect modes, in which an excitation light was incident to the surface defect with an angle of 45° from the normal direction using a 50× objective lens (numerical aperture = 0.42), and the radiation from the defect was detected by the same lens. The inset shows an example of the fabricated surface defect structure, in which W_d and L_d were 275 nm (=0.55*a*) and 3*a*, respectively. **d**, An example of the micro-mapping of the response of the surface defect of L_d = 3*a* at the resonant wavelength of 1,446 nm. **e**, Series of resonant spectra from the artificial surface defects with various lengths L_d between 1*a* and 7*a*. (a.u., arbitrary units.) **f**, *Q* factors of individual defect modes. The inset shows an example of a resonant spectrum for the surface-defect mode with the highest *Q* factor (~9,000).

surface states. This work also has interesting analogies to the surface plasmon-polariton effect of metals and the related surface photon physics. The absorption-free nature of the 3D photonic-crystal surface means that various new sensing applications can be expected. Furthermore, various light–matter interactions could be achieved at the surface of 3D photonic crystals.

METHODS SUMMARY

The numerical analysis of the surface properties was performed by electromagnetic calculations using the 3D finite-difference time-domain (FDTD) method[25] with absorbing boundaries[26,27]. Photonic crystals with eight stacked layers were fabricated by stacking GaAs stripes[3,16]: initially, striped patterns were formed by electron-beam lithography and reactive ion etching in a single crystalline GaAs layer on an aluminium (Al)GaAs etch-stop layer and a GaAs substrate. A pair of

patterned wafers was stacked and bonded with a crossed configuration using a very accurate alignment system[17], and one of the substrates was removed. These processes were repeated until the desired structure was constructed. The in-plane area of the photonic crystal was set at 400 μm × 400 μm. The resulting structural misalignment was <80 nm (<0.16*a*). For the measurements of the band structures (Figs 2b and 3b, c), white light from a halogen lamp with a focal spot of ~60 μm by a 5× objective lens was used, and the reflection spectra were measured by a spectrometer with an indium (In)GaAs multi-channel detector. For the measurement of the light propagation via the surface state (Fig. 2d–f), monochromatic light from a tuneable continuous-wave semiconductor laser was used. The images were observed from the top side using a 20× objective lens and an InGaAs charge-coupled device (CCD) camera (320 × 256 pixels). For the measurements of the optical properties of the surface-defect modes, white light from a super-continuum source with a focal spot of ~4 μm by a 50× objective lens was used, and the radiation spectra were measured using a high-resolution spectrometer. The *Q* factors were estimated using the formula $Q = \lambda_0/\Delta\lambda_0$, where λ_0 and $\Delta\lambda_0$ are the measured central wavelength and the full-width at half-maximum (FWHM) of the radiation peaks, respectively.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 March; accepted 27 May 2009.

1. Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58,** 2059–2062 (1987).
2. John, S. Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett.* **58,** 2486–2489 (1987).
3. Noda, S., Tomoda, K., Yamamoto, N. & Chutinan, A. Full three-dimensional photonic bandgap crystals at near-infrared wavelengths. *Science* **289,** 604–606 (2000).
4. Noda, S., Fujita, M. & Asano, T. Spontaneous-emission control by photonic crystals and nanocavities. *Nature Photon.* **1,** 449–458 (2007).
5. Ogawa, S., Imada, M., Yoshimoto, S., Okano, M. & Noda, S. Control of light emission by 3D photonic crystals. *Science* **305,** 227–229 (2004).
6. Lodahl, P. *et al.* Controlling the dynamics of spontaneous emission from quantum dots by photonic crystals. *Nature* **430,** 654–657 (2004).
7. Li, J., Jia, B., Zhou, G. & Gu, M. Direction-dependent spontaneous emission from near-infrared quantum dots at the angular band edges of a three-dimensional photonic crystal. *Appl. Phys. Lett.* **91,** 254101 (2007).
8. Ogawa, S., Ishizaki, K., Furukawa, T. & Noda, S. Spontaneous emission control by 17 layers of three-dimensional photonic crystals. *Electron. Lett.* **44,** 377–378 (2008).
9. Aoki, K. *et al.* Coupling of quantum-dot light emission with a three-dimensional photonic-crystal nanocavity. *Nature Photon.* **2,** 688–692 (2008).
10. Qi, M. *et al.* A three-dimensional optical photonic crystal with designed point defects. *Nature* **429,** 538–542 (2004).
11. Rinne, S. A., García-Santamaría, F. & Braun, P. V. Embedded cavities and waveguides in three-dimensional silicon photonic crystals. *Nature Photon.* **2,** 52–56 (2008).
12. Imada, M., Lee, L. H., Okano, M., Kawashima, S. & Noda, S. Development of three-dimensional photonic-crystal waveguides at optical-communication wavelengths. *Appl. Phys. Lett.* **88,** 171107 (2006).
13. Meade, R. D., Brommer, K. D., Rappe, A. M. & Joannopoulos, J. D. Electromagnetic Bloch waves at the surface of a photonic crystal. *Phys. Rev. B* **44,** 10961–10964 (1991).
14. Ho, K. M., Chan, C. T., Soukoulis, C. M., Biswas, R. & Sigalas, M. Photonic band gaps in three dimensions: new layer-by-layer periodic structures. *Solid State Commun.* **89,** 413–416 (1994).
15. Sözüer, H. S. & Dowling, J. P. Photonic band calculations for woodpile structures. *J. Mod. Opt.* **41,** 231–239 (1994).
16. Yamamoto, N., Noda, S. & Chutinan, A. Development of one period of a three-dimensional photonic crystal in the 5–10 μm wavelength region by wafer fusion and laser beam diffraction pattern observation techniques. *Jpn. J. Appl. Phys.* **37,** L1052–L1054 (1998).
17. Kawashima, S., Imada, M., Ishizaki, K. & Noda, S. High-precision alignment and bonding system for the fabrication of 3-D nanostructures. *J. Microelectromech. Syst.* **16,** 1140–1144 (2007).
18. Astratov, V. N. *et al.* Photonic band-structure effects in the reflectivity of periodically patterned waveguides. *Phys. Rev. B* **60,** R16255–R16258 (1999).
19. Coquillat, D. *et al.* Equifrequency surfaces in a two-dimensional GaN-based photonic crstal. *Opt. Express* **12,** 1097–1108 (2004).
20. Paraire, N. & Banachour, Y. Investigation of planar photonic crystal band diagrams under the light cone using surface coupling techniques. *Appl. Phys. B* **89,** 245–250 (2007).
21. Otto, A. Excitation of nonradiative surface plasma waves in silver by the method of frustrated total reflection. *Z. Phys.* **216,** 398–410 (1968).
22. Kretschmann, E. & Raether, H. Radiative decay of nonradiative surface plasmons excited by light. *Z. Naturforsch.* **23a,** 2135–2136 (1968).
23. Tien, P. K., Ulrich, R. & Martin, R. J. Modes of propagating light waves in thin deposited semiconductor films. *Appl. Phys. Lett.* **14,** 291–294 (1969).
24. Galli, M. *et al.* Excitation of radiative and evanescent defect modes in linear photonic crystal waveguides. *Phys. Rev. B* **70,** 081307 (2004).
25. Yee, K. S. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Trans. Antenn. Propag.* **14,** 302–307 (1966).
26. Mur, G. Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic-field equations. *IEEE Trans. Electromagn. Compat.* **23,** 377–382 (1981).
27. Berenger, J. P. A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.* **114,** 185–200 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Q. Sun for assistance with the experiment and numerical analysis, and S. Kawashima, S. Takahashi, Y. Tanaka, T. Asano and M. Fujita for discussions and advice. This work was supported in part by the Global Centre of Excellence for Education and Research on Photonics and Electronics Science and Engineering of Kyoto University, Japan, and by Grants-in-Aid for Scientific Research of the Ministry of Education, Culture, Sports, Science and Technology of Japan, and the Japan Society for the Promotion of Science (JSPS). K.I. acknowledges support from a Research Fellowship of the JSPS.

Author Contributions S.N. planned and organized the project; K.I. conducted the calculations, fabrications and measurements; S.N. and K.I. discussed the results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.N. (snoda@kuee.kyoto-u.ac.jp).

METHODS

Numerical calculations of band structures and surface-defect modes. The numerical calculations were performed using the 3D FDTD method[25]. For the structure without the surface (Fig. 1b), the Bloch's periodic boundary was used for all directions. A unit cell of the calculation was set to be $a \times a \times 4h$, where a is the centre-to-centre separation of the parallel dielectric rods and h is the thickness of the rod. The refractive index n of the dielectric rods was set to be 3.375, based on GaAs in the optical communications wavelength range. The mesh size for the FDTD calculations was set to be $a/16$ for both the x-direction and the y-direction, and $a/15$ for the z-direction. The shadowed band areas shown in Fig. 1b were derived by projecting the modal frequency of the 3D wavevector $\mathbf{k}$ to their in-plane wavevector $\mathbf{k}_{//}$. For the structure with the surface (Fig. 1c), the Bloch's periodic boundary was used for the in-plane directions and the Mur's second-order absorbing boundary[26] was used for the vertical direction. The mesh size was set to be $a/16$ for all directions. The whole structure was constructed with an air region ($n = 1$), a 3D photonic crystal and a dielectric material as a substrate ($n = 3.375$), forming the surface modes on only one side of the photonic crystal edge, which is the interface between the photonic crystal and the air region. The electromagnetic field distribution (inset in Fig. 1c) was also calculated using this model, with the in-plane wavenumber $k_{//}$ fixed at a specific value of $0.25\sqrt{2} \times (2\pi/a)$ for the Γ–M direction. When calculating the properties of the surface-defect mode, a structural model with a photonic crystal having an in-plane size of $16a \times 16a$ and a dielectric substrate was considered. The absorbing boundary of the Berenger's perfectly matched layer[27] was used in three dimensions, and the mesh size used for the calculation was set at $a/16$.

Measurements of band structures and light propagation via the surface state. The evanescent-coupling method with a prism[21–24] was used to excite the surface modes. For the measurements of the band structure of the surface modes, a prism with a refractive index n_{prism} of 1.5 was positioned above the surface with an air gap of ~500 nm (Fig. 2a). A 5× objective lens (numerical aperture = 0.14) was used to focus the incident light onto an area with a diameter of ~60 μm, which was sufficiently smaller than the area of the fabricated photonic crystal (400 μm × 400 μm). In the measurement using the prism of $n_{\mathrm{prism}} = 1.5$, the critical angle θ_c of the total reflection at the bottom face of the prism was 41.8°. We used various values of the incident angle θ (43.0°, 44.3°, 45.7°, 47.0°, 48.3°, 50.3°, 52.3°, 54.9°, 57.5°, 60.1° and 64.5°) to change the in-plane wavenumber $k_{//}$. For each angle, the reflection spectra were measured using a spectrometer in the wavelength range of 1,140 to 1,650 nm. The inset in Fig. 2b illustrates an example of spectrum obtained in this way. The band structures shown in Fig. 2b and Fig. 3b and c were obtained from the reflectance spectra of all values of angle θ, where the results for two polarizations in which the vectors of the electric field were parallel and perpendicular to the plane of incidence, namely the P and S polarizations, were superposed.

For the measurement of the light propagation via the surface mode (Fig. 2c–f), the position of the prism was changed and replaced at the bottom of the GaAs substrate. The irradiated light was from a semiconductor laser with a wavelength that could be tuned from 1,270 to 1,630 nm, and the images were detected by an infrared camera. For the estimation of the propagation length, the light intensity scattered at the edge of the photonic crystal was measured and the background intensity was subtracted. The scattered intensity values were plotted for various distances between the edge and the excitation point (Fig. 2g), and the relation between the distance and the intensity was fitted using an exponential function by least-squares approximation.

Measurements of the characteristics of surface-defect modes. These measurements were performed using the method shown in Fig. 4c, in which the surface-defect modes were excited by white light generated by a super-continuum source. To eliminate the influence of the light reflected directly from the surface of the 3D photonic crystal, the irradiating angle was set to be 45°. Furthermore, two sets of polarizers with polarization directions that differed by 90° were inserted in front of the light source and the detector, respectively, to eliminate the background noise as much as possible.

LETTERS

Photoconductance and inverse photoconductance in films of functionalized metal nanoparticles

Hideyuki Nakanishi[1,2], Kyle J. M. Bishop[2], Bartlomiej Kowalczyk[1,2], Abraham Nitzan[1,3], Emily A. Weiss[1], Konstantin V. Tretiakov[2], Mario M. Apodaca[1], Rafal Klajn[2], J. Fraser Stoddart[1] & Bartosz A. Grzybowski[1,2]

In traditional photoconductors[1–3], the impinging light generates mobile charge carriers in the valence and/or conduction bands, causing the material's conductivity to increase[4]. Such positive photoconductance is observed in both bulk and nanostructured[5,6] photoconductors. Here we describe a class of nanoparticle-based materials whose conductivity can either increase or decrease on irradiation with visible light of wavelengths close to the particles' surface plasmon resonance. The remarkable feature of these plasmonic materials is that the sign of the conductivity change and the nature of the electron transport between the nanoparticles depend on the molecules comprising the self-assembled monolayers (SAMs)[7,8] stabilizing the nanoparticles. For SAMs made of electrically neutral (polar and non-polar) molecules, conductivity increases on irradiation. If, however, the SAMs contain electrically charged (either negatively or positively) groups, conductivity decreases. The optical and electrical characteristics of these previously undescribed inverse photoconductors can be engineered flexibly by adjusting the material properties of the nanoparticles and of the coating SAMs. In particular, in films comprising mixtures of different nanoparticles or nanoparticles coated with mixed SAMs, the overall photoconductance is a weighted average of the changes induced by the individual components. These and other observations can be rationalized in terms of light-induced creation of mobile charge carriers whose transport through the charged SAMs is inhibited by carrier trapping in transient polaron-like states[9,10]. The nanoparticle-based photoconductors we describe could have uses in chemical sensors and/or in conjunction with flexible substrates.

We used gold and silver nanoparticles (diameter 5.6 ± 0.8 nm, with surface plasmon resonance (SPR) maxima in solution of respectively $\lambda_{max}^{Au} \approx 520$ nm and $\lambda_{max}^{Ag} \approx 420$ nm) synthesized as described before[11–13]; these were stabilized by monocomponent or mixed SAMs of alkane thiols (Fig. 1a) abbreviated as C_3OH, C_6OH, $C_{11}OH$, $C_{10}COOH$, $C_{11}NMe_3^+$, C_6, PhOH and PhCOOH. The suspensions of the nanoparticles were drop-cast onto a glass or Teflon substrate, and the solvent was evaporated under vacuum to give nanoparticle films of thickness (h) 120–300 nm (determined by profilometry for each sample), and characterized by SPR bands red-shifted (by ~50–60 nm) and broadened owing to nanoparticle aggregation[13,14] (Fig. 1c). Next, Pd/Au electrodes (width $w = 5$ mm, separated by a gap of $L = 50$ μm; Fig. 1b) were sputtered onto the films, which were then dried under high vacuum for at least one week to remove traces of solvent and water. The samples were placed in a custom-made, hermetically sealed Faraday cage under inert atmosphere and in the presence of phosphorus pentoxide, a dehydrating agent. The electrodes were connected to a Keithley 6517 electrometer, a white-light LED (light-emitting diode) was placed above the film, and optical bandpass filters were used to select the desired wavelength of light. Unless otherwise noted, the temperature during all experiments was kept constant to within 0.02 °C (for further experimental details, see Supplementary Information section 1).

In the absence of irradiation, both Au and Ag films exhibit ohmic current density–applied field (j–E) characteristics in the field intensity range from 100 V m^{-1} (corresponding to 0.005 V across the system and sub-picoamp currents at the sensitivity limit of the electrometer) to 20 kV m^{-1} (higher fields cause irreversible changes in the nanoparticle films and coalescence of the neighbouring nanoparticles).

Irradiation with visible light alters the currents through the nanostructured films (Fig. 2). We note that the signs of the observed

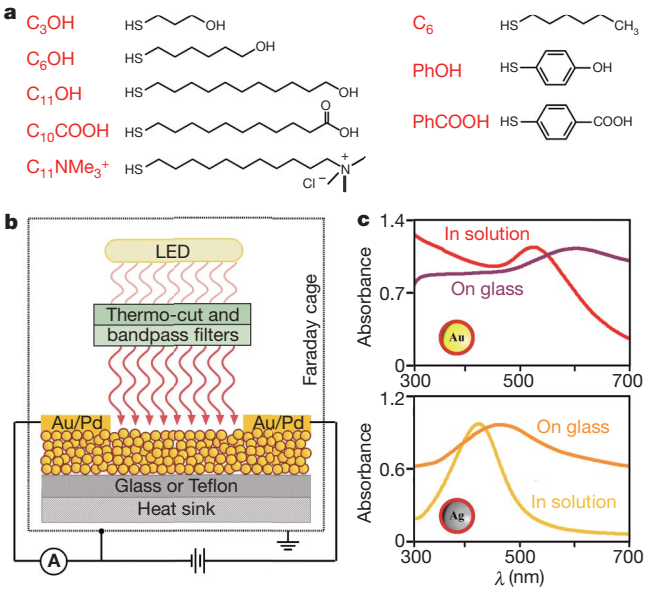


Figure 1 | Functionalized nanoparticles for photoconductive films. **a**, Structures of alkane thiols used to form SAMs on gold and/or silver nanoparticles. **b**, Diagram of the experimental set-up (not to scale). The thickness of the nanoparticle layer is ~120–300 nm. Thinner layers give currents of magnitude comparable to the external/instrument noise. **c**, Ultraviolet/visible spectra of Au/C_6OH (top) and Ag/C_6OH (bottom) nanoparticles in solution and when cast onto glass films. Note that owing to aggregation of the nanoparticles[12,13], the spectra of the particles in the films are broadened and red-shifted compared to those of free particles in solution. The degree of this spectral shift increases with decreasing thickness of SAMs separating the metal cores of the nanoparticles (for example, $\lambda_{max}^{Au} \approx 580$ nm for Au/C_6OH films but $\lambda_{max}^{Au} \approx 770$ nm for Au/C_3OH films).

[1]Department of Chemistry, [2]Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, USA. [3]School of Chemistry, Tel Aviv University, Tel Aviv, Israel.

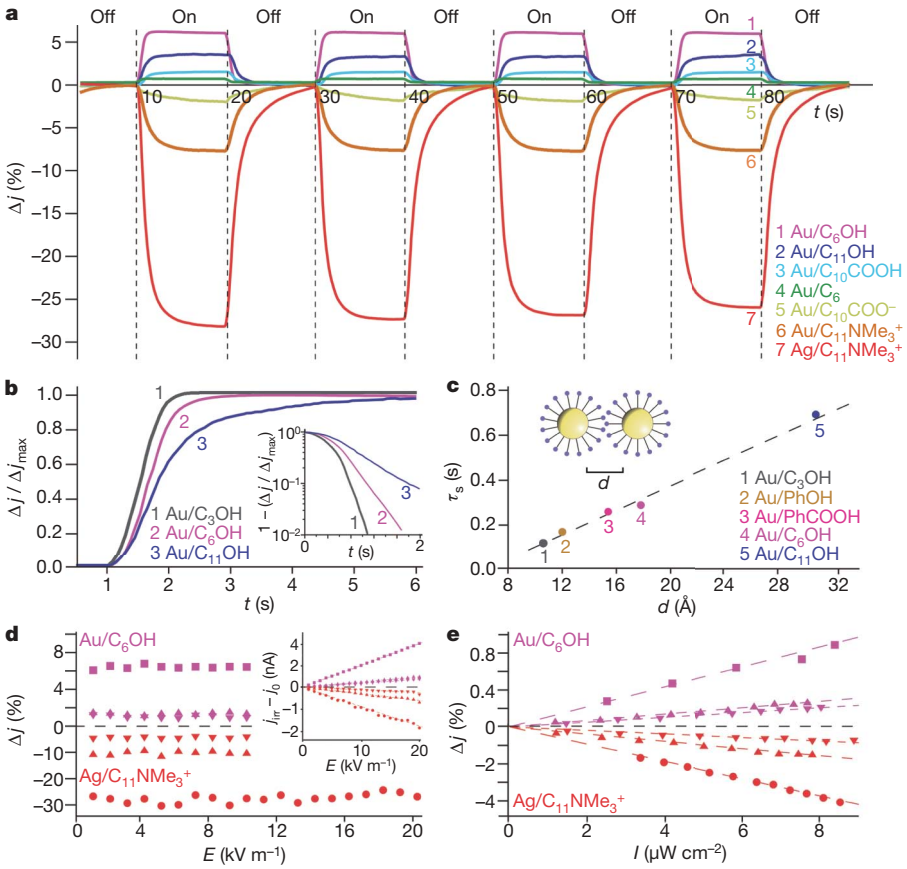


Figure 2 | Photoconductance and inverse photoconductance of irradiated nanoparticle films. **a**, Change in current density on irradiation, Δj, for various types of nanoparticle films irradiated periodically with white LED light. Key specifies metal cores and coating thiols (Fig. 1a). Control experiments with thin films of thermally evaporated Au or Ag show no conductivity changes on irradiation. **b**, Dependence of Δj on time t (immediately after the irradiation commences) for a series of uncharged thiols differing only in chain length. Inset, linear regions in the semi-log plot correspond to an exponential $\Delta j(t)$ dependence; slopes equal the inverse of τ_s. **c**, τ_s decreases approximately linearly with decreasing thickness, d, of SAMs separating nanoparticle metal cores; d was estimated from molecular modelling and, independently, from TEM images of the films (methods agreed to within 15%). Errors in the times are of the order of a few per cent (~7%) based on measurements of five cycles of irradiation for at least three independent samples. **d**, Irrespective of E, and for a given value of I (here, 60 μW cm^{-2}), the changes in photocurrent $|\Delta j(\lambda)|_I$ are maximal when λ is close to the SPR maximum of the nanoparticles in the film. Violet symbols, Au/C_6OH films (maximal $|\Delta j(\lambda)|_I$ at around $\lambda^{Au}_{max} = 580$ nm); red symbols, Ag/$C_{11}NMe_3^+$ films (maximal $|\Delta j(\lambda)|_I$ at $\lambda^{Ag}_{max} = 470$ nm); triangles, 370–440 nm; circles, 410–500 nm; squares, 490–590 nm; inverted triangles, 640–780 nm. Inset, absolute values of current density illustrating linear dependence of $j_{irr} - j_0$ on E. **e**, For constant E (here, 20 kV m^{-1}) and for different wavelengths, λ, relative current densities scale linearly with the intensity of irradiating light, $\Delta j(I) \propto I$. Symbols and wavelength ranges as in **d**. For different material samples (>5 for each condition/material tested), the values of τ_s. and Δj in **c**–**e** agree to within less than 6%.

changes depend on the chemical nature of the thiols coating the nanoparticles but not on the nature of the particles' metal cores. Figure 2a shows typical variations in the relative current density, $\Delta j \equiv (j_{irr} - j_0)/j_0$ (j_0 and j_{irr} indicate current density in the respective absence and presence of irradiation), recorded for different types of nanoparticle films (all at $E = 20$ kV m^{-1} and for nanoparticles coated with monocomponent SAMs) periodically exposed to white LED light of intensity $I_{white} = 60$ μW cm^{-2}. When the thiols are electrically neutral, current density increases on irradiation, $\Delta j > 0$. When, however, the thiols contain electrically charged groups, $\Delta j < 0$. Decreasing current densities are observed in films composed of either positively charged nanoparticles (coating thiol $C_{11}NMe_3^+$) or negatively charged nanoparticles (deprotonated thiol $C_{10}COO^-$). Interestingly, films comprising nanoparticles stabilized with protonated thiols $C_{10}COOH$ exhibit positive Δj changes (compare curves 3 and 5 in Fig. 2a).

When light is switched on or off, $\Delta j(t)$ depends on time t approximately exponentially, $\Delta j(t) \propto (-t/\tau_s)$, where τ_s denotes a characteristic 'switching' time, which is the same for 'on' and 'off' exponents, does not depend on the intensity of light, and is of the order of hundreds of milliseconds. For a given terminal functionality of the thiol molecules comprising the SAMs, the switching times increase approximately linearly with increasing chain length (Fig. 2b and c). For identical SAMs, the switching times increase with decreasing nanoparticle size (for example, $\tau_s \approx 0.29$ s for Au/C_6OH with metal core diameter 5.6 nm versus $\tau_s \approx 0.18$ s for 10.0-nm Au/C_6OH particles). We note that in all cases the values of τ_s are not related to the intrinsic current 'response' times of the measuring device (<1 ms) accompanying the step-wise change in the applied field.

A second set of observations relates the magnitudes of the relative current changes, $|\Delta j|$, to the field strength, E, the wavelength, λ, and the intensity, I, of the irradiating light. For all films studied and all values of λ, we find that both the 'dark' current and the light-induced change are linear in E, such that the relative light-induced change $|\Delta j(E)|_{I,\lambda}$ is independent of E (Fig. 2d). Furthermore, at constant λ, $|\Delta j(I)|_\lambda$ is linear in the light intensity I (up to about tens of μW cm^{-2}; for higher intensities, photocurrent effects are masked by increasing temperature of the film). Figure 2e illustrates this effect for Au/C_6OH and Ag/$C_{11}NMe_3^+$ films irradiated with light of intensity up to $I = 8$ μW cm^{-2}. The corresponding slope depends on the nature of the nanoparticles/SAM and on the wavelength λ. Finally, for all values of E and I, $|\Delta j(\lambda)|_{E,I}$ is maximal when λ coincides with the SPR of the aggregated nanoparticles' metal cores. Thus, for Au- and Ag-based films, $|\Delta j|$ is maximal when the wavelength is close to $\lambda^{Au}_{max} = 580$ nm

and $\lambda_{max}^{Ag} = 470$ nm, respectively (Fig. 2d and e). For films comprising both Au and Ag particles, the superposition of the two SPR peaks results in maximal current changes over a wider range of wavelengths, $\lambda \approx 400$–650 nm. In all of the above experiments, the largest current density change for uncharged thiols is ~6% for Au/C_6OH films. For charged thiols, changes as large as −27% are observed for films composed of Ag/$C_{11}NMe_3^+$ nanoparticles (compare Fig. 2a).

Both positive and inverse photoconductance can be explained qualitatively by a model (Supplementary Information sections 2, 3) in which the nanoparticle array is characterized by a density N_t of discrete 'trap sites' of which n_t are filled at thermal equilibrium. Physically, we suggest that these traps are of two types: (1) sites on the nanoparticle 'cores', and (2) sites on the organic ligands. The former arise because the Fermi energy of the nanoparticles is several eV below[15] that of the alkyl chains separating the particles. Therefore, each nanoparticle can accommodate a certain number of electrons depending on its charging energy, which in turn depends on the nanoparticle radius and capacitance[16,17]. In reality, charge carriers 'trapped' on the cores are delocalized over the particles' surface; however, for simplicity we treat each core as a collection of discrete trap sites, each of which can accommodate a single charge carrier. Additionally, the organic ligands separating the nanoparticle cores may also act as traps if (as in the case of charged thiols discussed below) their conduction orbitals are of lower energy than the plasmon state of the nanoparticles. Within the framework of this simple but widely used trapping model[18,19], the effective equilibrium between free and trapped carriers (n and n_t, respectively) is given by:

$$\frac{n[N_t - n_t]}{n_t} = K \qquad (1)$$

where K is an effective equilibrium constant characterizing the density of free and trapped carriers at the steady state (Supplementary Information section 3.4). In this way, the density of traps determines the density of free charge carriers, n, and thereby the conduction current, $j = -e\mu nE$, where e is the fundamental charge, and μ is the mobility of the charge carrier in the nanoparticle medium.

For arrays of nanoparticles capped with uncharged ligands, the only trap sites are those on the nanoparticle cores (see Fig. 3a), and electrons tunnel from nanoparticle to nanoparticle through a large barrier imposed by the insulating organic SAM. This barrier is 3–5 eV (refs 15, 20) relative to the Fermi energy of gold (−5.1 eV; ref. 15) or silver (−4.3 eV; ref. 15). Starting from the steady-state current in the dark, photoexcitation at or near the plasmon resonance of the nanoparticles (Fig. 2d, e) promotes electrons from the valence band to the conduction band of the nanoparticles. Subsequently, the photoexcited electrons 'fill' available trap sites present on the nanoparticles, thereby reducing the effective density of these sites by $\Delta N_t = N_t^{irr} - N_t < 0$. Through the effective equilibrium relation (1), the decrease in N_t acts to increase the density of free carriers, n, and produces a positive Δj in the uncharged-SAM systems.

For arrays of nanoparticles capped with charged SAMs, the traps on the nanoparticle cores are supplemented by trap sites on the organic ligands (see Fig. 3b). These additional traps probably arise due to the formation of polaron-like states via the so-called bootstrap mechanism, whereby electron localization occurs through the reorganization of the nuclear environment of the ionic moieties in the presence of an electron[9,10,21]. The degree of energetic stabilization of these polaronic traps via nuclear reorganization was estimated experimentally at ~1.0 eV from the temperature dependence of the conduction current (Fig. 3b right, and Supplementary Information section 2). In the dark, the energy of these states is expected to lie above the Fermi energy of the nanoparticles, and the organic ligands do not act as trap sites but rather as part of the tunnelling barrier (Fig. 3b left). On irradiation, however, excitation at the plasmon resonance of the nanoparticles promotes charge carriers to an injection energy that is ~2 eV above their injection energy in the dark[9,10,21]. Relative to this excited state, the charged ligands become effective trap sites on irradiation, thereby increasing the density of traps $\Delta N_t > 0$ (see Supplementary Information section 3.4 for an alternative but ultimately equivalent kinetic interpretation). Consequently, the density of free carriers, n, decreases in accordance with equation (1), and the current decreases to give a negative Δj in the charged systems.

We note that the formation of polaron-like traps cannot be affected by the electric field produced by the oscillating electron density of the plasmon on the nanoparticles (Supplementary Information section 4). Conduction electrons on the nanoparticles excited at or near the plasmon resonance oscillate at optical frequencies (~10^{14} Hz) and cannot couple to molecular motions/structural rearrangement within the organic SAMs. Furthermore, the d.c. fields produced by

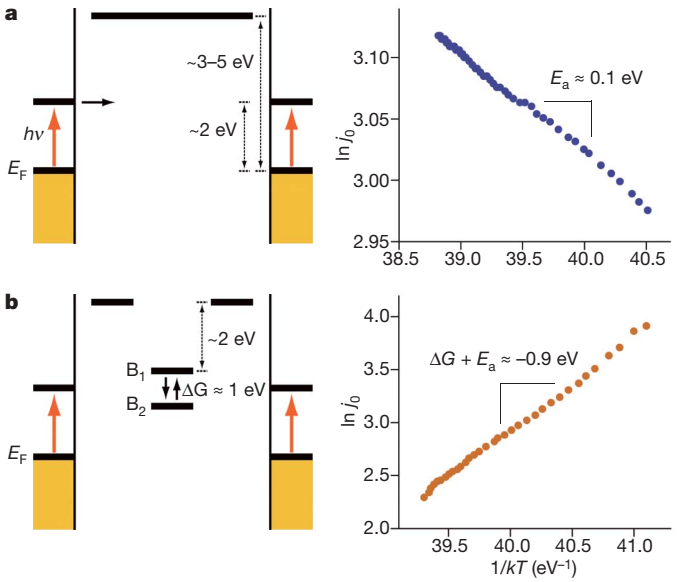


Figure 3 | Energy diagrams and temperature dependence of responses. **a**, Left panel: qualitative electronic structure diagram for nanoparticles capped with uncharged ligands. Yellow bars, nanoparticles' metal cores characterized by Fermi energies, E_F, which in the absence of irradiation define an approximate border between occupied and unoccupied states. On irradiation, electrons are promoted into excited metal states, ~2 eV above[9,10,21] E_F. Molecular levels of the non-polar alkane thiols lie ~3–5 eV above[15,20] E_F. Right panel: current though a film of Au/C_{11}OH nanoparticles as a function of inverse temperature (differences in slopes recorded for different material samples are within 4% of the average). Because $\ln(j_0) \approx -E_a/k_BT$ + constant (ref. 26), the slope (E_a) corresponds to the amount of energy that the environment contributes to the charge transport process (probably a mixture of thermally activated hopping and temperature-independent tunnelling). **b**, For nanoparticles capped with charged ligands, the scenario presented in **a** is augmented by the presence of polaron-like states within the organic layer separating the nanoparticles. The existence of these states is supported by temperature dependent current measurements (right; Au/$C_{11}NMe_3^+$, error in the slope <4%) showing that charge transport through arrays of charged nanoparticles is negatively activated (increasing the temperature decreases the current for a given applied voltage). Negative activation is common in charge transport systems[26] where an equilibrium exists between two conformational states of a donor-bridge-acceptor complex. Here, the molecular orbitals comprising the charged moieties (that is, the bridge) are characterized by two conformations (left), one of which (B_1) is more favourable for electron transfer than the other (B_2). Negative activation occurs when raising the temperature pushes the conformational equilibrium towards the unfavourable geometry (B_2) (Supplementary Information section 2). The observed activation barrier, $E_a^{obs} = E_a + \Delta G \approx -0.9$ eV, can be used to estimate the difference in free energy, ΔG, between the two bridge conformations—here, $\Delta G \approx -1$ eV. Also, the energy of bridge states corresponding to the polar groups before polaronic reorganization is ~2 eV below those of the alkyl chains (refs 27–29; Supplementary Information section 2). Thus, in the presence of light irradiation, the lower energy bridge state (B_2) can act as a trap for charge carriers, resulting in a negative photocurrent (main text).

these oscillations are, for the particle sizes used here, vanishingly small, and so cannot align dipoles within charged SAMs[22,23].

An interesting manifestation of the two trapping scenarios is seen when the films contain both uncharged and charged ligands. Figure 4a plots the results of a series of experiments with films comprising (1) binary mixtures of charged and uncharged nanoparticles (here, Au/$C_{11}NMe_3^+$ and Au/$C_{11}OH$; blue squares) and (2) nanoparticles coated with mixed SAMs[24] of charged and uncharged thiols ($C_{11}NMe_3^+$ and $C_{11}OH$; yellow circles). For both types of system, the overall relative current changes can be expressed as a linear combination of the current changes of monocomponent films, $\Delta j = \chi_{OH}\Delta j_{OH} + \chi_{NMe3+}\Delta j_{NMe3+}$, where χ_{OH} and $\chi_{NMe3+} = 1 - \chi_{OH}$ are the fractions of either nanoparticles of each type in the film or of different thiols on the nanoparticle surfaces. These linear dependences imply that the effects of polaronic-like centres on carrier trapping are additive, such that the number of trapping sites introduced on irradiation increases linearly with the number of charged moieties in the material (Supplementary Information section 3.3.vi). Interestingly, at a particular ratio $\chi_{OH} \approx 0.6$, the increase in the number of mobile carriers created by photoexcitation is exactly offset by the increase in the number of traps due to charged thiols, such that the effective number of trapping sites remains unchanged on irradiation and the photocurrent is identical to the dark current, $\Delta j = 0$.

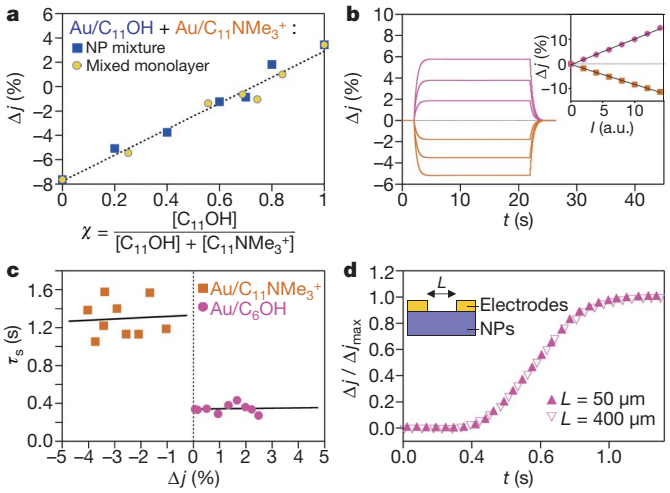


Figure 4 | Conductance of mixed films and theoretical predictions. **a**, Typical linear dependence of Δj on the composition χ of the nanoparticle films. Blue squares, films composed of Au/$C_{11}OH$ and Au/$C_{11}NMe_3^+$ nanoparticles (NPs). Yellow circles, films comprising Au nanoparticles coated with mixed SAMs of $C_{11}OH$ and $C_{11}NMe_3^+$ thiols. Correlation coefficients: $r^2 = 0.982$ for nanoparticle mixtures, $r^2 = 0.976$ for mixed-SAM nanoparticles, $r^2 = 0.962$ for all data points; errors based on the analysis of at least seven samples per condition are less than ~10% of the averages shown. Similar linear trends are observed for other thiol pairs studied. Dashed line (fit) indicates that this linear trend is also predicted by the theoretical model (see main text). **b**, Positive and negative photocurrent response as calculated by charge transport simulations (Supplementary Information section 3). The current change on irradiation (or ceasing of irradiation) is well approximated by an exponential saturation to the new steady-state current with a switching time characteristic of the trapping kinetics (here, $\tau_s = 0.5$ s). Inset, the magnitude of the photoresponse is linearly proportional to the light intensity. **c**, Experimentally recorded τ_s as a function of Δj for Au/$C_{11}NMe_3^+$ and Au/C_6OH nanoparticle arrays using light irradiation of different intensities. Solid lines, theoretical fits from the charge transport model, illustrating that τ_s is largely independent of the magnitude of Δj but depends on the nature of the ligands (here $\tau_s \approx 1.3$ s for Au/$C_{11}NMe_3^+$ and $\tau_s \approx 0.34$ s for Au/C_6OH). **d**, Photocurrent response for Au/C_6OH nanoparticles for two different electrode separations, $L = 50$ μm and 400 μm. The switching time does not depend on L (in agreement with theoretical prediction) but only on the nature of the 'trap sites' within the nanoparticles and, in the case of charge ligands, within the organic layer.

To gain further insight into the response of the material to light irradiation, it is necessary to account for the kinetic processes of trap-filling or trap-emptying. Starting from the charge transport equations (Supplementary Information section 3), the positive and negative changes in the current density on photoexcitation can be related to the changes in the total number of trap sites (from N_t in the dark to N'_t on irradiation) and the concomitant changes in the number of free charge carriers (from n to n') subject to equation (1). It can be shown that in the absence of significant space charge injection[18,19], the transport equations can be simplified to give the following equation relating changes in n to the trapping kinetics:

$$\frac{dn}{dt} = -\frac{dn_t}{dt} = k[-n(N'_t - n_t) + Kn_t] \quad (2)$$

where k is a rate constant for trapping. In the limit of $N_t \gg n_t$ this equation is linear and predicts the exponential saturation of n and n_t to their new steady-state values: importantly, as $j \propto n$, the model reproduces the exponential response of the system when light is switched on or off and also the linear dependence of j on the light intensity (Fig. 4b). Furthermore, $\tau_s \approx 1/kN_t$ (Supplementary Information section 3), so that the switching timescale observed in experiments (0.1–1 s) is equivalent to that associated with charge trapping. This timescale is commensurate with but slightly faster than the 10–100 s trapping times previously found for charge retention within gold nanoparticle transistors[25]. The model also correctly predicts that for a given type of SAM, the switching time depends neither on Δj (Fig. 4c) nor on the width of the gap between the electrodes (Fig. 4d). Finally, the model reproduces the linear dependence of τ_s on the SAM thickness (compare Fig. 2c) and its decrease

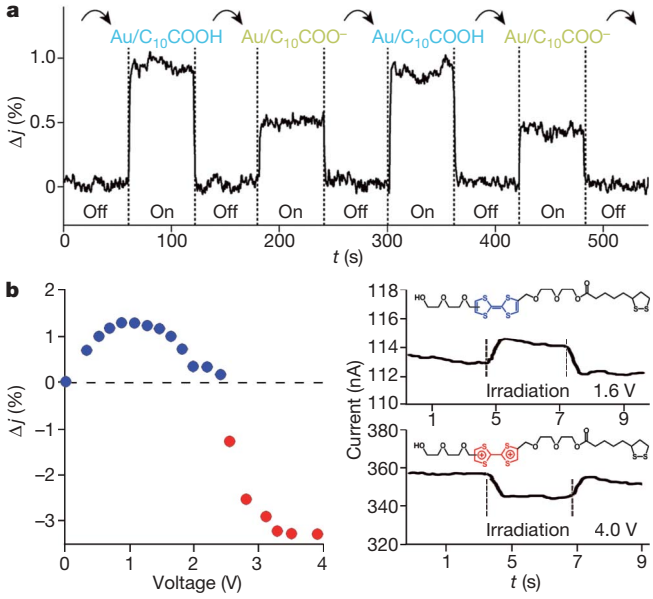


Figure 5 | Modulation and 'switching' of photoconductance. **a**, Exposure of a film of Au nanoparticles coated with COOH thiols to the vapours of ammonium hydroxide causes partial deprotonation of carboxylic acid groups and decreases the photocurrent; exposure to vapours of hydrochloric acid re-protonates these groups causing the photocurrent to increase. As the current does not switch to negative values observed for nanoparticles coated with fully deprotonated thiols (Fig. 2a), the deprotonation is probably only partial and/or limited by penetration of the film by the vapours' molecules. Note that this experiment is performed in the presence of humidity, and the photoconductance effect is still observed. In fact, we verified that exposure of the film to neutral water vapour (relative humidity RH < 30%) does not alter photoconductance. **b**, Switching from positive to negative (inverse) photoconductance in a film of Au nanoparticles coated with ligands terminated in tetrathiafulvalene (TTF; structural formulae shown). The effect can be attributed to the TTF → TTF^{2+} oxidation (blue, neutral TTF; red, TTF dication) previously reported to occur at ~2 V (ref. 30).

with increasing nanoparticle size (because as the volume fraction of gold within the film decreases, the density of traps within the array decreases).

We have described a class of nanostructured materials that exhibit photoconductance as well as previously undescribed inverse photoconductance. From a fundamental perspective, the present work provides new insights into electron transport in heterogeneous nanostructured media in which photoexcitation and trapping of charge carriers can be spatially separated between metallic nanoparticle and insulating SAM domains. Incorporation of molecules that can be interconverted between charged and uncharged forms could enable external control of the degree of photoconductance (for example, by chemical means, Fig. 5a) or even switching between normal and inverse photoconductance modes (for example, by the applied bias, Fig. 5b).

Received 9 December 2008; accepted 11 May 2009.

1. Adam, D. *et al.* Fast photoconduction in the highly ordered columnar phase of discotic liquid-crystal. *Nature* **371,** 141–143 (1994).
2. Sze, S. M. *Physics of Semiconductor Devices* (Wiley & Sons, 1981).
3. Yu, G., Gao, J., Hummelen, J. C., Wudl, F. & Heeger, A. J. Polymer photovoltaic cells — enhanced efficiencies via a network of internal donor-acceptor heterojunctions. *Science* **270,** 1789–1791 (1995).
4. Joshi, N. V. *Photoconductivity: Art, Science and Technology* (Marcel Dekker, 1990).
5. Han, S. *et al.* Photoconduction studies on GaN nanowire transistors under UV and polarized UV illumination. *Chem. Phys. Lett.* **389,** 176–180 (2004).
6. Hayden, O., Agarwal, R. & Lieber, C. M. Nanoscale avalanche photodiodes for highly sensitive and spatially resolved photon detection. *Nature Mater.* **5,** 352–356 (2006).
7. Love, J. C., Estroff, L. A., Kriebel, J. K., Nuzzo, R. G. & Whitesides, G. M. Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chem. Rev.* **105,** 1103–1169 (2005).
8. Witt, D., Klajn, R., Barski, P. & Grzybowski, B. A. Applications properties and synthesis of omega-functionalized n-alkanethiols and disulfides — the building blocks of self-assembled monolayers. *Curr. Org. Chem.* **8,** 1763–1797 (2004).
9. Galperin, M., Ratner, M. A. & Nitzan, A. Hysteresis, switching, and negative differential resistance in molecular junctions: a polaron model. *Nano Lett.* **5,** 125–130 (2005).
10. Kuznetsov, A. M. Negative differential resistance and switching behavior of redox-mediated tunnel contact. *J. Chem. Phys.* **127,** 084710 (2007).
11. Kalsin, A. M. *et al.* Electrostatic self-assembly of binary nanoparticle crystals with a diamond-like lattice. *Science* **312,** 420–424 (2006).
12. Kalsin, A. M., Kowalczyk, B., Smoukov, S. K., Klajn, R. & Grzybowski, B. A. Ionic-like behavior of oppositely charged nanoparticles. *J. Am. Chem. Soc.* **128,** 15046–15047 (2006).
13. Kalsin, A. M. *et al.* Electrostatic aggregation and formation of core-shell suprastructures in binary mixtures of charged metal nanoparticles. *Nano Lett.* **6,** 1896–1903 (2006).
14. Pinchuk, A. O., Kalsin, A. M., Kowalczyk, B., Schatz, G. C. & Grzybowski, B. A. Modeling of electrodynamic interactions between metal nanoparticles aggregated by electrostatic interactions into closely-packed clusters. *J. Phys. Chem. C* **111,** 11816–11822 (2007).
15. Engelkes, V. B., Beebe, J. M. & Frisbie, C. D. Length-dependent transport in molecular junctions based on SAMs of alkanethiols and alkanedithiols: effect of metal work function and applied bias on tunneling efficiency and contact resistance. *J. Am. Chem. Soc.* **126,** 14287–14296 (2004).
16. Bozano, L. D. *et al.* Organic materials and thin-film structures for cross-point memory cells based on trapping in metallic nanoparticles. *Adv. Funct. Mater.* **15,** 1933–1939 (2005).
17. Lee, J. S. *et al.* Layer-by-layer assembled charge-trap memory devices with adjustable electronic properties. *Nature Nanotechnol.* **2,** 790–795 (2007).
18. Lampert, M. A. Simplified theory of space-charge-limited currents in an insulator with traps. *Phys. Rev.* **103,** 1648–1656 (1956).
19. Many, A. & Rakavy, G. Theory of transient space-charge-limited currents in solids in presence of trapping. *Phys. Rev.* **126,** 1980–1988 (1962).
20. Lee, T. *et al.* Comparison of electronic transport characterization methods for alkanethiol self-assembled monolayers. *J. Phys. Chem. B* **108,** 8742–8750 (2004).
21. Nitzan, A. *Chemical Dynamics in Condensed Phases* (Oxford Univ. Press, 2006).
22. Willets, K. A. & Van Duyne, R. P. Localized surface plasmon resonance spectroscopy and sensing. *Annu. Rev. Phys. Chem.* **58,** 267–297 (2007).
23. Zou, S. L. & Schatz, G. C. in *Surface-Enhanced Raman Scattering: Physics and Applications* (eds Kneipp, K., Moskovits, M. & Kneipp, H.) 67–85 (Topics in Applied Physics, Vol. 103, Springer, 2006).
24. Kalsin, A. M., Kowalczyk, B., Wesson, P., Paszewski, M. & Grzybowski, B. A. Studying the thermodynamics of surface reactions on nanoparticles by electrostatic titrations. *J. Am. Chem. Soc.* **129,** 6664–6665 (2007).
25. Novembre, C., Guerin, D., Lmimouni, K., Gamrat, C. & Vuillaume, D. Gold nanoparticle-pentacene memory transistors. *Appl. Phys. Lett.* **92,** 103314 (2008).
26. Weiss, E. A. *et al.* Conformationally gated switching between superexchange and hopping within oligo-p-phenylene-based molecular wires. *J. Am. Chem. Soc.* **127,** 11842–11850 (2005).
27. Jortner, J. & Noyes, R. M. Some thermodynamic properties of the hydrated electron. *J. Phys. Chem.* **70,** 770–774 (1966).
28. Schnitker, J. & Rossky, P. J. Quantum simulation study of the hydrated electron. *J. Chem. Phys.* **86,** 3471–3485 (1987).
29. Zhan, C. G. & Dixon, D. A. The nature and absolute hydration free energy of the solvated electron in water. *J. Phys. Chem. B* **107,** 4403–4417 (2003).
30. Nguyen, T. D. *et al.* A reversible molecular valve. *Proc. Natl Acad. Sci. USA* **102,** 10029–10034 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Ratner, G. C. Schatz and R. van Duyne for discussions and advice. This work was supported by the Alfred P. Sloan Fellowship and the Dreyfus Teacher-Scholar Award (to B.A.G.).

Author Contributions H.N. performed the experiments, and collected and analysed the data; K.J.M.B., A.N., E.A.W. and B.A.G. developed the theoretical model; B.K. and R.K. synthesized nanoparticles and thiols; K.V.T. and M.M.A. helped with the construction of the Faraday cage and with data analysis; J.F.S. planned synthesis and helped with the interpretation of results; and B.A.G. conceived the experiments, analysed results, and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.A.G. (grzybor@northwestern.edu).

Vol 460|16 July 2009|doi:10.1038/nature08163

LETTERS

Evidence for middle Eocene Arctic sea ice from diatoms and ice-rafted debris

Catherine E. Stickley[1,3], Kristen St John[2], Nalân Koç[1,3], Richard W. Jordan[4], Sandra Passchier[5], Richard B. Pearce[6] & Lance E. Kearns[2]

Oceanic sediments from long cores drilled on the Lomonosov ridge, in the central Arctic[1], contain ice-rafted debris (IRD) back to the middle Eocene epoch, prompting recent suggestions that ice appeared in the Arctic about 46 million years (Myr) ago[2,3]. However, because IRD can be transported by icebergs (derived from land-based ice) and also by sea ice[4], IRD records[2,3] are restricted to providing a history of general ice-rafting only. It is critical to differentiate sea ice from glacial (land-based) ice as climate feedback mechanisms vary and global impacts differ between these systems: sea ice directly affects ocean–atmosphere exchanges[5], whereas land-based ice affects sea level and consequently ocean acidity[6]. An earlier report[3] assumed that sea ice was prevalent in the middle Eocene Arctic on the basis of IRD, and although somewhat preliminary supportive evidence exists[2], these data are neither comprehensive nor quantified. Here we show the presence of middle Eocene Arctic sea ice from an extraordinary abundance of a group of sea-ice-dependent fossil diatoms (*Synedropsis* spp.). Analysis of quartz grain textural characteristics further supports sea ice as the dominant transporter of IRD at this time. Together with new information on cosmopolitan diatoms and existing IRD records[2], our data strongly suggest a two-phase establishment of sea ice: initial episodic formation in marginal shelf areas ~47.5 Myr ago, followed ~0.5 Myr later by the onset of seasonally paced sea-ice formation in offshore areas of the central Arctic. Our data establish a 2-Myr record of sea ice, documenting the transition from a warm, ice-free[3] environment to one dominated by winter sea ice at the start of the middle Eocene climatic cooling phase[7].

Current thinking suggests that the Cenozoic cryosphere evolved in phases, starting in the late–middle Eocene with the presence of small, isolated ice caps in both hemispheres[8–10], and interrupted by at least one short warming event[11] before major ice-sheet expansion on Antarctica and concomitant deep-sea cooling in the late Eocene and early Oligocene[7–9]. Much of the focus has been on land-based ice, but sea ice is an integral part of past and present cryospheric systems as it is involved in important climate feedbacks, such as changes in the global energy budget through the albedo effect[5,8]. Yet relatively little is known of the temporal and geographic distribution of sea ice before the Quaternary, largely because its effects are expressed more subtly and, arguably, less widely in the geological record than those of land-based ice. Part of the problem is that transient or localized sea ice may leave no obvious sedimentary record, or that its effects are easily overlooked. Application of one of the conventional methods for tracing past sea ice, using sea-ice-associated diatoms in marine sediments[12], is less obvious for the pre-Quaternary and is certainly challenging for the current patchy Eocene diatom record, particularly in Antarctica (Supplementary Information).

In the Arctic, pre-Quaternary sea-ice patterns are beginning to be understood through the Integrated Ocean Drilling Program Expedition 302 ('ACEX'), which recovered the first long Cenozoic palaeoceanographic record from the Lomonosov ridge in the central Arctic[1,3,13] (Fig. 1). It is now known, for example, that episodic perennial sea ice first appeared in the middle Miocene[14,15]. Here we present evidence for middle Eocene seasonal sea ice from a ~2-Myr-long record of fossil sea-ice diatoms and sea-ice-dominated IRD; this record is contained within a ~36-m interval of finely laminated, organic-rich, biosiliceous sediments[1,16] that represent the start of Cenozoic cooling following the Early Eocene Climatic Optimum[7] (Supplementary Fig. 1). These sediments contain abundant marine and freshwater siliceous and organic-walled microfossils[1,16–18], including diverse and largely endemic diatom assemblages characterized by heavily silicified marine Palaeogene taxa such as *Hemiaulus* spp.[1,16] and *Anaulus arcticus*[18].

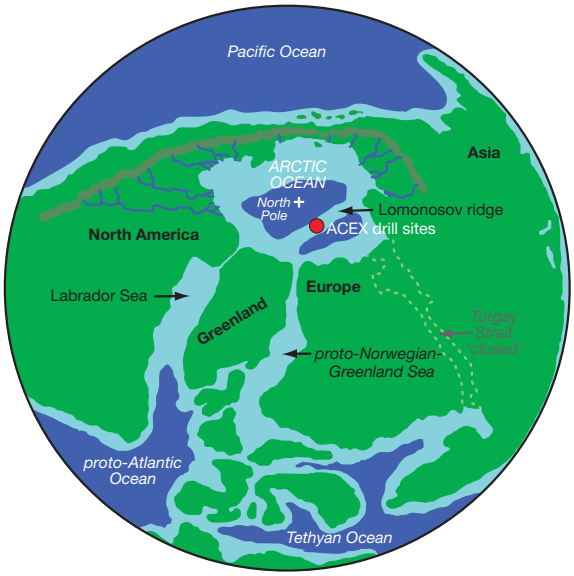


Figure 1 | Idealized palaeogeography of the Arctic region for the early middle Eocene during the phase of biosilica production and preservation at the Lomonosov ridge (~50–45 Myr ago). Redrawn from ref. 17 with the difference that the Turgay Strait (outlined) is essentially closed to any major exchange at this time, based on siliceous microfossil data (Supplementary Information).

[1]Department of Geology, University of Tromsø, N-9037 Tromsø, Norway. [2]Department of Geology and Environmental Science, James Madison University, Harrisonburg, Virginia 22807, USA. [3]Norwegian Polar Institute, Polar Environmental Centre, N-9296 Tromsø, Norway. [4]Department of Earth and Environmental Sciences, Yamagata University, Yamagata 990-8560, Japan. [5]Department of Earth and Environmental Studies, Montclair State University, Montclair, New Jersey 07043, USA. [6]National Oceanography Centre, Southampton, University of Southampton, Southampton SO14 3ZH, UK.

Above 260.30 m composite depth (m.c.d.), weakly silicified, araphid, needle-shaped pennate diatoms *Synedropsis* spp. (Figs 2 and 3; Supplementary Figs 2–4) are preserved in abundances typically exceeding 25%, but reaching as high as ~61% of the diatom assemblage. Previously overlooked in initial investigations[1], these are some of the most delicately silicified diatoms discovered in the fossil record, with preservation so remarkable that micro-aggregate and colony integrity is commonly still intact (Supplementary Information). Sediment trap studies show that most diatom valves do not reach the sea floor because normal sea water is under-saturated in silica, making opaline (amorphous hydrous) silica—that impregnates the diatom cell-wall—highly susceptible to dissolution and biasing diatom records towards relatively large or heavily silicified types[19]. Weakly silicified diatoms are preserved only under exceptional circumstances, and are particularly unusual for Palaeogene sediments. In this respect, the ACEX diatom record is unprecedented and reflects a most unusual environmental setting. We attribute the preservation of *Synedropsis* spp. in these sediments to the presence of sea ice and silica-enriched waters.

Today, colonies of extant *Synedropsis* spp. are peculiar to polar sea ice[20] (Supplementary Information). The ACEX fossil *Synedropsis* spp. are uniquely associated with the middle Eocene IRD record from the same cores (Fig. 3), and here we demonstrate that this IRD record is predominantly derived from sea ice. On the basis of this association and by analogy with living *Synedropsis* spp., including the manner in which they are preserved (Supplementary Information), we propose that the ACEX *Synedropsis* spp. were, like their modern counterparts, sea-ice-dependent and, as such, highly specialized extremophiles, uniquely adapted for surviving the lengthy polar darkness and freezing temperatures. These diatoms provide the most compelling evidence for ancient sea ice, as they rely on this medium for their survival. Furthermore, the ACEX *Synedropsis* spp. represent the earliest known fossil record of sea-ice diatoms (Supplementary Information). Also we found no reports in the literature of *Synedropsis* spp. (or of any synonyms[20] or misidentifications) in pre-Holocene sediments from elsewhere in the Arctic and subarctic regions, suggesting their geographic and stratigraphic limitation.

The presence of *Synedropsis* spp. in the ACEX sediments shows that non-permanent (for example, seasonal) sea ice existed over the

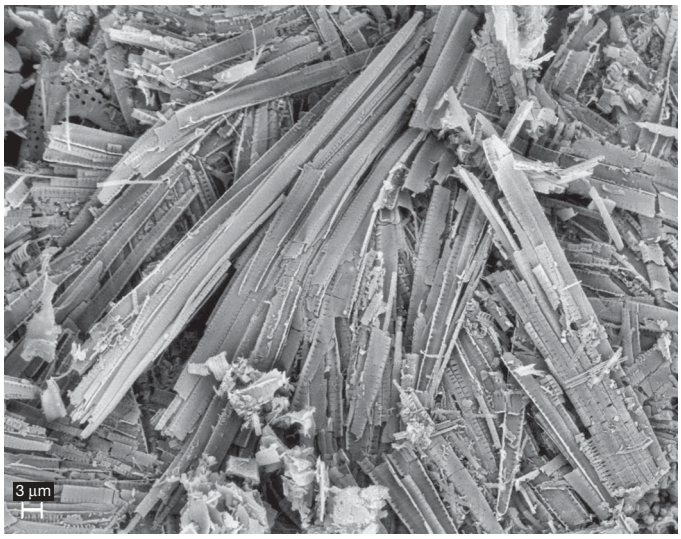


Figure 2 | SEM image of part of an aggregate of near-whole needle-shaped *Synedropsis* sp. valves. Imaged along a lamina-parallel fracture (topographic) surface of unprocessed core material at 240.90 m.c.d. (302-2A-55X5, 19.5 cm; 46.16 Myr ago). The main part of the image shows most valves are missing just the apices. Valves that have been fractured in several places are visible in the lower left and upper right corners of this image. Scale bar, 3 µm. Images of colonies, aggregates, and details of the apices are shown in Supplementary Figs 3 and 4.

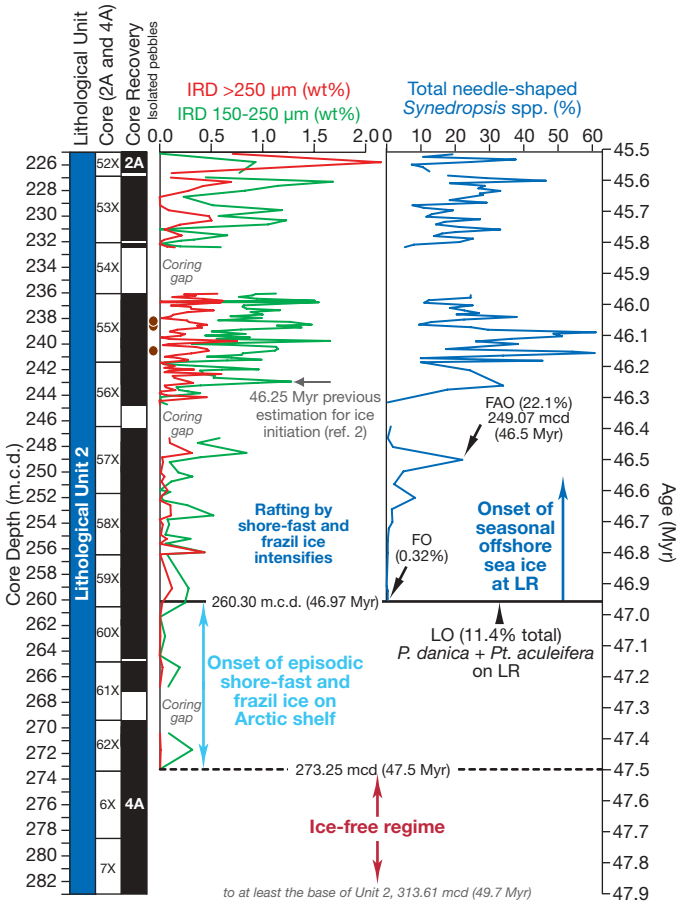


Figure 3 | IRD and sea-ice diatom abundance in the ACEX cores. Shown are data for the >250 µm (red) and 150–250 µm (green) quartz fractions, and the relative abundance (percentage of total diatoms) of needle-shaped *Synedropsis* spp. (blue). Over the age range shown, shallow exchange between the Lomonosov ridge and the proto-Norwegian-Greenland sea remains possible year-round until 46.97 Myr ago, but becomes seasonally restricted at younger ages. FO, first occurrence; FAO, first abundant occurrence; LO, last occurrence; LR, Lomonosov ridge. IRD data are from ref. 2. Age model is from ref. 13.

Lomonosov ridge, and that regular (for example, annual) melting episodes, into silica-enriched waters, permitted their preservation. It is likely that sea ice formed in autumn and winter and melted in spring and summer, as seasonal sea ice does today. We suggest that *Synedropsis* spp. over-wintered within the sea ice and later bloomed there after irradiance increased in the following spring. Following sea-ice melt, they were released into stratified surface waters where aggregation was undoubtedly intensified by virtue of their stickiness[21] and needle shape. As aggregation enhances sinking rates, these were unlikely candidates for seeding the melt zone[21]. We therefore suggest they did not constitute a significant part of the summer growth alongside other diatom taxa. Instead, they underwent rapid flux to the sea floor where they remained largely undisturbed under an anoxic benthic environment[16,22]. Such sea-ice-driven seasonality is evident from our preliminary investigations of the laminations that indicate *Synedropsis* spp. form regularly occurring, near-monogeneric laminations, commonly associated with IRD. This indicates that production and flux of sea-ice diatoms occurred at a different stage in the annual cycle to that of the other ACEX diatoms, and that *Synedropsis* spp. laminations probably represent episodes of spring or summer sea-ice melt and flux, a hypothesis necessitating examination of the lamination sequences through the core.

The current estimate for the initiation of ice in the Arctic is ~46.25 Myr ago[2] (Fig. 3), based on IRD evidence only, in particular on the doubling of sand-sized IRD flux to 0.02 g cm^{-2} kyr^{-1} and

establishment of a quasi-cyclicity[2]. However, IRD exists in the ACEX cores as far back as ~47.5 Myr ago[2] (Fig. 3), albeit in low abundance. *Synedropsis* spp. appear in the record 46.97 Myr ago, thereafter co-varying with IRD throughout the biosilica-rich core sections (Supplementary Fig. 2). The appearance of *Synedropsis* spp. in the ACEX cores coincides with the disappearance of a cosmopolitan (warmer water) diatom component characterized by *Porotheca danica* and *Pterotheca aculeifera* (Fig. 3), whereas other endemic taxa persist throughout the cores. Considering that the IRD record is largely derived from sea-ice transportation (as we show here), the timing of this floral turnover strongly suggests a two-phase establishment of middle Eocene Arctic sea ice (Fig. 3).

We propose that the onset of IRD in the ACEX cores ~47.5 Myr ago marks the onset of shore-fast ice and/or frazil ice, forming up-current in one or more shallow areas of the broad marginal Arctic shelf, which entrained and (via wind-driven surface currents) transported sand-sized terrigenous material to the central Arctic. Marginal sea-ice formation was likely to have been episodic, based on the low abundance of IRD (Fig. 3) and the absence of sea-ice diatoms in this phase (Supplementary Information). We further propose that the appearance of well-preserved, lamina-forming sea-ice diatoms and the concurrent disappearance of cosmopolitan diatoms ~47 Myr ago marks the onset of seasonally paced offshore sea-ice formation—for example, pack ice. It is conceivable that—under generally declining middle Eocene temperatures and atmospheric carbon dioxide (CO_2) levels (refs 23, 24)—sea ice would have formed initially at the margins, later advancing into offshore areas of the Arctic Ocean, to include at least the surface-waters above, or close to, the palaeo-location of the ACEX drillsite (Fig. 1). A parallel increase in IRD (Fig. 3) indicates that ice-rafting of shallow shelf sediments intensified at the same time.

Before ~47 Myr ago, year-round shallow oceanic exchange between the Lomonosov ridge and the proto-Norwegian-Greenland Sea was possible, indicated by the presence of *P. danica* and *Pt. aculeifera* in coeval sediments from this area. Once an offshore seasonal sea-ice regime was established, the northern limit of ice-intolerant cosmopolitan diatoms was driven southwards, and the potential for shallow oceanic exchange became restricted to the summer melt season.

Our data provide compelling evidence for the establishment of an offshore winter sea-ice regime ~47 Myr ago that was stable enough to support an ecology and persistent enough to cause the southward migration of cosmopolitan taxa. Sea-ice-driven positive climate feedbacks, such as reduced oceanic heat loss to the atmosphere and enhanced albedo—by sea ice[5] (in spring) and cloud formation[25] (in summer)—would therefore have become particularly important at this time. A stable sea-ice regime also suggests the possibility of concomitant glacial ice. If glaciers were also present, then an understanding of the relative contribution of both sea ice-IRD and iceberg-IRD is essential.

A possible discriminator in the Arctic is quantitative analysis of the surface textures of ice-rafted quartz grains[26], using a statistical approach[27]. IRD grains exhibiting both sea-ice and iceberg textures are present in a short study interval (~236–242 m.c.d.; Fig. 4). Discriminant scores are significantly different for the two groups with 93% correctly classified, although some overlap exists between the populations (Supplementary Fig. 5). Our data indicate that sea ice was the dominant mode of ice transport between ~46.2 Myr and ~46.0 Myr ago. Before ~46.15 Myr ago (240.53 m.c.d.), 80–100% of the grains exhibit sea-ice textures (0–20% iceberg-IRD textures). The highest input of iceberg-IRD occurs ~46.15 Myr ago, after which peaks in iceberg-IRD broadly correspond to peaks in total IRD abundance (Fig. 4). This suggests the presence of small isolated glaciers that pre-date estimates for those on Greenland (~38 Myr or ~44 Myr ago, refs 10 and 28 respectively), in keeping with results of a global climate/ice-sheet model[9]; this model shows the potential for small isolated ice caps on higher Arctic elevations in the Eocene

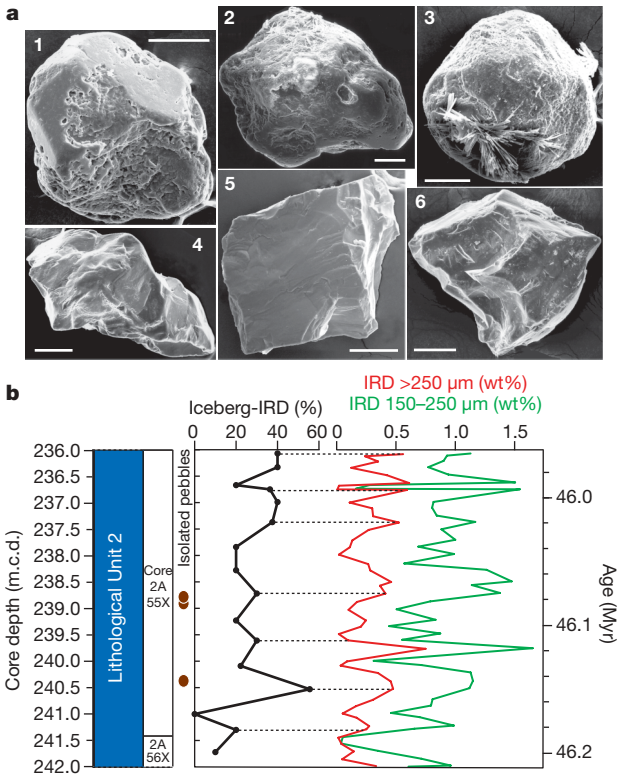


Figure 4 | Results of IRD analysis in a short core interval. a, SEM images of IRD with textural characteristics of sea-ice transport (upper row) and iceberg transport (lower row). **b**, Percentage of iceberg-rafted versus sea-ice-rafted quartz sand (>250 μm) and correlation with IRD abundance. Details are provided in Supplementary Information.

and Oligocene, and no major glaciation until ~23 Myr ago or later[9]. Nevertheless, even where total IRD input is high, 60–80% of grains are classified as sea ice-IRD, confirming the importance of sea ice in the middle Eocene Arctic.

The ACEX middle Eocene sea-ice record pre-dates known major Cenozoic glacial and sea-ice records, which has two intriguing consequences: (1) the current 'icehouse' cryosphere started to evolve earlier than previously thought; and (2) significant sea ice formed in the central Arctic some ~24 Myr before major ice-sheet expansion in the region[9]. Therefore, sea ice–ice sheet linkages, such as those surmised for Cenozoic Antarctica[5], could not have existed in the middle Eocene Arctic. Furthermore, together with results of a numerical model for Cenozoic Antarctic climate evolution[5], and known Cenozoic Antarctic sea-ice records (Supplementary Information), our data indicate that sea ice formed in the Arctic before it did in Antarctica. Under generally declining middle Eocene atmospheric CO_2 (refs 23, 24), this implies that the threshold for sea-ice formation was crossed in the Arctic first, a hypothesis opposite to that modelled for glacial ice, whereby Antarctica is shown to glaciate much earlier (that is, at higher levels of CO_2) than circum-Arctic continents[9].

Arctic palaeogeography and palaeoenvironments may ultimately account for these consequences. Strong salinity stratification[16,22] prevailed within a basin essentially isolated from the world's oceans (Fig. 1), where low surface salinities[29,30] appreciably facilitated annual freezing. A notable influx of fresh water[30] and subsequent fall in salinity starting ~47.6 Myr (refs 29, 30) or ~47.2 Myr (ref. 16) ago strongly suggests salinity-controlled preconditioning just before and during sea-ice initiation.

METHODS SUMMARY

Quantitative diatom analysis. Dried sediment samples from the interval ~201–302 m.c.d. (cores 302 2A 47X to 4A 11X) were weighed to a value between

~0.1–0.2 g, processed for biogenic silica and analysed using the quantitative and light microscopy methodologies outlined elsewhere[16]. Sampling resolution was at an average of ~20 cm over critical intervals (otherwise at every ~30–100 cm excluding coring gaps). Light microscope images were taken using a Panasonic DMC-FS5 camera at ISO 1600 and at ×1,000 magnification. Further aliquots were dried onto aluminium scanning electron microscope (SEM) stubs, gold sputter coated and imaged using an LEO 1450VP SEM at 15–20 kV. Lamina-parallel fracture (topographic) surfaces of unprocessed core material were also imaged in the SEM.

Quantitative IRD abundance. We followed previously outlined methods[2].

Quantitative IRD surface texture analysis. We analysed 160 grains; 10 quartz grains from the >250 μm size range were randomly picked for SEM imaging from each of 16 samples between ~236 and 242 m.c.d. (cores 302 2A 55X-56X). This interval was selected for its good core recovery[1], good age control[13], robust IRD record[2], and for encompassing the oldest recorded dropstone[1–3] ('isolated pebbles' on Figs 3 and 4 and Supplementary Fig. 2). Samples were analysed using an LEO 1430VP SEM. Elemental (EDS) analysis verified that each grain was quartz. Statistical analysis included principal component analysis (PCA) and discriminant analysis[27].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 January; accepted 21 May 2009.

1. Backman, J., Moran, K., McInroy, D. B., Mayer, L. A. & the Expedition 302 Scientists. *Proc. IODP* **302,** doi:10.2204/iodp.proc.302.2006 (2006).
2. St John, K. Cenozoic ice-rafting history of the central Arctic Ocean: terrigenous sands on the Lomonosov Ridge. *Paleoceanography* **23,** PA1S05, doi:10.1029/2007PA001483 (2008).
3. Moran, K. *et al.* The Cenozoic palaeoenvironment of the Arctic Ocean. *Nature* **441,** 601–605 (2006).
4. Nurnberg, D. *et al.* Sediments in Arctic sea ice: implications for entrainment, transport, and release. *Mar. Geol.* **119,** 185–214 (1994).
5. DeConto, R., Pollard, D. & Harwood, D. Sea ice feedback and Cenozoic evolution of Antarctic climate and ice sheets. *Paleoceanography* **22,** PA3214, doi:10.1029/2006PA001350 (2007).
6. Merico, A., Tyrrell, T. & Wilson, P. A. Eocene/Oligocene ocean de-acidification linked to Antarctic glaciation by sea level fall. *Nature* **452,** 979–982 (2008).
7. Zachos, J. *et al.* Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292,** 686–693 (2001).
8. DeConto, R. M. & Pollard, D. Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO_2. *Nature* **421,** 245–249 (2003).
9. DeConto, R. M. *et al.* Thresholds for Cenozoic bipolar glaciation. *Nature* **455,** 652–656 (2008).
10. Eldrett, J. S. *et al.* Continental ice in Greenland during the Eocene and Oligocene. *Nature* **466,** 176–179 (2007).
11. Bohaty, S. M. *et al.* Coupled greenhouse warming and deep-sea acidification in the middle Eocene. *Paleoceanography* **24,** PA2207, doi:10.1029/2008PA001676 (2009).
12. Armand, L. K. & Leventer, A. in *Sea Ice: An Introduction to its Physics, Chemistry, Biology and Geology* (eds Thomas, D. N. & Dieckmann, G. S.) 333–372 (Blackwell Science, 2003).
13. Backman, J. *et al.* Age model and core-seismic interpretation for the Cenozoic Arctic Coring Expedition sediments from the Lomonosov Ridge. *Paleoceanography* **23,** PA1S03, doi:10.1029/2007PA001476 (2008).
14. Darby, D. A. Arctic perennial ice cover over the last 14 million years. *Paleoceanography* **23,** PA1S07, doi:10.1029/2007PA001479 (2008).
15. Krylov, A. A. *et al.* A shift in heavy and clay mineral provenance indicates a middle Miocene onset of a perennial sea ice cover in the Arctic Ocean. *Paleoceanography* **23,** PA1S06, doi:10.1029/2007PA001497 (2008).
16. Stickley, C. E. *et al.* A siliceous microfossil view of middle Eocene Arctic environments: a window of biosilica production and preservation. *Paleoceanography* **23,** PA1S14, doi:10.1029/2007PA001485 (2008).
17. Brinkhuis, H. *et al.* Episodic fresh surface waters in the early Eocene Arctic Ocean. *Nature* **441,** 606–609 (2006).
18. Suto, I., Jordan, R. W. & Watanabe, M. Taxonomy of middle Eocene diatom resting spores and their allied taxa from IODP sites in the central Arctic Ocean (the Lomonosov Ridge). *Micropaleontology* **55,** 259–312 (2009).
19. Barron, J. A. & Baldauf, J. G. in *Siliceous Microfossils* (eds Blome, C. D. *et al.*) 107–118 (Short Courses in Paleontology, Vol. 8, Paleontological Society, 1995).
20. Hasle, G. R., Medlin, L. K. & Syvertsen, E. E. *Synedropsis* gen. nov., a genus of araphid diatoms associated with sea ice. *Phycologia* **33,** 248–270 (1994).
21. Riebesell, U., Schloss, I. & Smetacek, V. Aggregation of algae released from melting sea ice: implications for seeding and sedimentation. *Polar Biol.* **11,** 239–248 (1991).
22. Stein, R., Boucsein, B. & Meyer, H. Anoxia and high primary productivity in the Paleogene central Arctic Ocean: first detailed record from the Lomonosov Ridge. *Geophys. Res. Lett.* **33,** L18606, doi:10.1029/2006GL026776 (2006).
23. Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **406,** 695–699 (2000).
24. Lowenstein, T. K. & Demicco, R. V. Elevated Eocene atmospheric CO_2 and its subsequent decline. *Science* **313,** 1928 (2006).
25. Abbot, D. S., Huber, M., Bousquet, G. & Walker, C. C. High-CO_2 cloud radiative forcing feedback over both land and ocean in a global climate model. *Geophys. Res. Lett.* **36,** L05702, doi:10.1029/2008GL036703 (2009).
26. Dunhill, G. Comparison of sea-ice and glacial-ice rafted debris: grain size, surface features, and grain shape. *US Geol. Surv. Open File Rep.* **OF 98-0367,** 1–82 (1998).
27. Williams, A. T. & Morgan, P. Scanning electron microscope evidence for offshore-onshore sand transport at Fire Island, New York, USA. *Sedimentology* **40,** 63–77 (1993).
28. Tripati, A. *et al.* Evidence for glaciation in the Northern Hemisphere back to 44 Ma from ice-rafted debris in the Greenland Sea. *Earth Planet. Sci. Lett.* **265,** 112–122 (2008).
29. Waddell, L. M. & Moore, T. C. Jr. Salinity of the Eocene Arctic Ocean from isotope analysis of fish bone carbonate. *Paleoceanography* **23,** PA1S12, doi:10.1929/2007PA001451 (2008).
30. Gleason, J. D. *et al.* Early to middle Eocene history of the Arctic Ocean from Nd-Sr isotopes in fossil fish debris, Lomonosov Ridge. *Paleoceanography* **24,** PA2215, doi:10.1029/2008PA001685 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research used samples and data provided by the Integrated Ocean Drilling Program (IODP). C.E.S. and N.K. acknowledge funding by VISTA (Norwegian Academy of Science and Letters and StatoilHydro) and the Research Council of Norway. We thank R. M. DeConto, M. P. Olney, S. M. Bohaty, A. Davies, M. A. Pearce, F. Sangiorgi, H. Brinkhuis, P. K. Bijl, J. Backman, J. Pike and A. E. S. Kemp for discussions. C.E.S. thanks StatoilHydro, Bergen for access to facilities, J. A. Barron, E. Fourtanier, I. Suto and J. Onodera for preliminary talks, and W. Hale for facilitating sample collection.

Author Contributions C.E.S. wrote the manuscript except for the sections on IRD and surface textures, which were written by K.S.J. and S.P. The manuscript incorporates comments by all authors, who also advised on its contents, structure and remit. Diatom analysis and light microscope imaging was undertaken by C.E.S.; statistical work on the surface textures was carried out by S.P.; R.B.P. and C.E.S. imaged *Synedropsis* spp. in the SEM at the NOCS; L.E.K. and K.S.J. undertook SEM analysis of surface textures; and SEM observations on *Synedropsis* spp. were also performed by R.W.J. and N.K.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.E.S. (catherine.stickley@npolar.no).

METHODS

Quantitative IRD abundance. Bulk sediment samples were freeze-dried and weighed. Samples were wet-sieved and dried at 60 °C in an oven after which the sediment fractions were weighed. Visual estimates of the vol.% of the terrigenous abundance in the 150–250 and >250 μm fractions were made using a binocular microscope. The terrigenous grains in these sand-size fractions are considered to be IRD. The wt% IRD is calculated by multiplying the wt% abundance of the sand fraction (150–250 μm, or >250 μm) by the visual estimate of the abundance of terrigenous grains in the sand fraction.

Quantitative IRD surface texture analysis. A checklist approach[27,31] was used, involving photographing and recording the presence/absence of a series of features on every grain. Checklist categories are a hybrid from refs 26 and 27. The number of variables (textures) was reduced to result in a data set with as many sea-ice textures as iceberg textures and the values for all variables were normalized to between 0 and 1. Based on three principal components, accounting for ~50% of the variance in the data set, individual grains were classified as either sea-ice or iceberg sourced, based on the classification of ref. 26. For the discriminant analysis, grains were assigned to groups according to the PCA results.

31. Helland, P. E. & Diffendal, R. F. Jr. Probable glacial climatic conditions in source areas during deposition of parts of the Ash Hollow Formation, Ogallala Group (Late Tertiary), of Western Nebraska. *Am. J. Sci.* **293**, 744–757 (1993).

Vol 460|16 July 2009|doi:10.1038/nature08189

LETTERS

Migration of the subtropical front as a modulator of glacial climate

Edouard Bard[1] & Rosalind E. M. Rickaby[2]

Ice cores extracted from the Antarctic ice sheet suggest that glacial conditions, and the relationship between isotopically derived temperatures and atmospheric p_{CO_2} have been constant over the last 800,000 years of the Late Pleistocene epoch[1]. But independent lines of evidence, such as the extent of Northern Hemisphere ice sheets[2], sea level[3] and other temperature records[4], point towards a fluctuating severity of glacial periods, particularly during the more extreme glacial stadials centred around 340,000 and 420,000 years ago (marine isotope stages 10 and 12). Previously unidentified mechanisms therefore appear to have mediated the relationship between insolation, CO_2 and climate. Here we test whether northward migration of the subtropical front (STF) off the southeastern coast of South Africa acts as a gatekeeper for the Agulhas current[5,6], which controls the transport of heat and salt from the Indo-Pacific Ocean to the Atlantic Ocean. Using a new 800,000-year record of sea surface temperature and ocean productivity from ocean sediment core MD962077, we demonstrate that during cold stadials (particularly marine isotope stages 10 and 12), productivity peaked and sea surface temperature was up to 6 °C cooler than modern temperatures. This suggests that during these cooler stadials, the STF moved northward by up to 7° latitude, nearly shutting off the Agulhas current. Our results, combined with faunal assemblages from the south Atlantic[7,8] show that variable northwards migration of the Southern Hemisphere STF can modulate the severity of each glacial period by altering the strength of the Agulhas current carrying heat and salt to the Atlantic meridional overturning circulation. We show hence that the degree of northwards migration of the STF can partially decouple global climate from atmospheric partial pressure of carbon dioxide, p_{CO_2}, and help to resolve the long-standing puzzle of differing glacial amplitudes within a consistent range of atmospheric p_{CO_2}.

The fluctuating severity of glacial periods (see Supplementary Information) challenges our understanding of the sensitivity of global climate to forcing by astronomical parameters, and p_{CO_2}. We investigate the migration of the STF in the southern Indian Ocean over multiple glacial cycles to explore whether the latitude of the STF acts as an insolation-sensitive feedback in the climate system, and amplifies low-latitude temperature anomalies to global climate phenomena. The latitude of the STF is determined by the strength and location of atmospheric winds. The subtropical convergence, characterized by negative wind stress curl in the atmosphere and associated Ekman transport convergence in the ocean, denotes the region between the core of the trade winds and the maximum southwesterly winds. Embedded in the subtropical convergence and towards its poleward side is the STF, a zonal band of enhanced productivity and weak (2 °C) meridional sea surface temperature (SST) gradients. The STF defines the southern limit of the subtropical gyres and separates them from the broad eastward flow of the Antarctic circumpolar current (ACC) further south.

We reconstruct the latitude of the STF from hydrographic changes at core MD962077, currently situated under the Agulhas current of the subtropical gyre of the Indian Ocean over the last 800 thousand years (kyr) (see online-only Methods). We test whether northward migration of the STF relative to South Africa acts as a gatekeeper for the leakage of heat and salt between the Atlantic and Indo-Pacific oceans. The Agulhas valve comprises surface and intermediate water transport from the Indian Ocean into the Atlantic. This takes place through the detachment of rings, eddies and filaments from the Agulhas retroflection, consisting of warm and salty Indian Ocean waters and providing a key return route for the supply of waters to the Atlantic meridional overturning circulation (AMOC)[5,6]. Core MD962077 is situated ~7° latitude equatorwards of the modern STF (33.17 °S, 31.25 °E, 3,781 m, Fig. 1), and provides a sensitive monitor of the meridional movements of the STF, which marks the southern boundary of the Agulhas return current.

Our 800-kyr records of SST (U^K_{37}', and $\delta^{18}O$ of *Globigerina inflata* and coccolithophore fraction; Fig. 2a) and biological productivity (abundance of authigenic uranium U_{auth}, total organic carbon TOC and total C_{37} alkenones C_{37tot}; Fig. 2b, data available in Supplementary Information) indicate that each 100-kyr glacial cycle is marked by cooler temperatures and higher productivity, indicating a northerly migration of the STF more proximal to core MD962077, and displaying the same distinctive pattern of a secular trend in the glacial minima. The glacial northward shift of the STF agrees with records from foraminiferal and coccolithophore assemblages in this region[7–11], which highlight that the reappearance and abundance, in the south Atlantic, of the 'Agulhas leakage fauna' is due to a deglacial strengthening of the Agulhas connection.

The marine isotope stages MIS-12 and MIS-10 stand out as exceptionally cool and productive glacial periods. Alkenone-derived SSTs record a subtropical Indian Ocean 4 °C cooler than the Holocene during MIS-2, but MIS-12 and MIS-10 were an additional 2 °C colder. This secular trend in SST is matched by the *G. inflata* $\delta^{18}O$ value (the MIS-12 $\delta^{18}O$ is heavier than that of MIS-2 by 0.5‰, ~2 °C) and coccolithophore fraction $\delta^{18}O$ (MIS-12 is heavier than MIS-2 by 1.5 ‰ but natural effects amplify the coccolithophore fraction $\delta^{18}O$; see online-only Methods). All three productivity proxies imply enhanced productivity during MIS-12 and 10 compared to other glacials. Individually, these proxies cannot be scaled linearly to productivity but are highly sensitive to productivity changes (see online-only Methods). However, the coherence in signal between three SST and productivity proxies, recorded in different phytoplankton and different phases of the sediment susceptible to a range of biases, lends confidence that our secular trend in temperature and productivity is robust.

A picture emerges that the STF migrated northwards towards core MD962077 during each glacial period. During MIS-12 and MIS-10,

[1]CEREGE (UMR 6635), Collège de France, University Paul-Cézanne Aix-Marseille, CNRS, IRD, Europole de l'Arbois BP 80, 13545 Aix-en-Provence Cedex 4, France. [2]Department of Earth Sciences, Oxford University, Parks Road, Oxford OX1 3PR, UK.

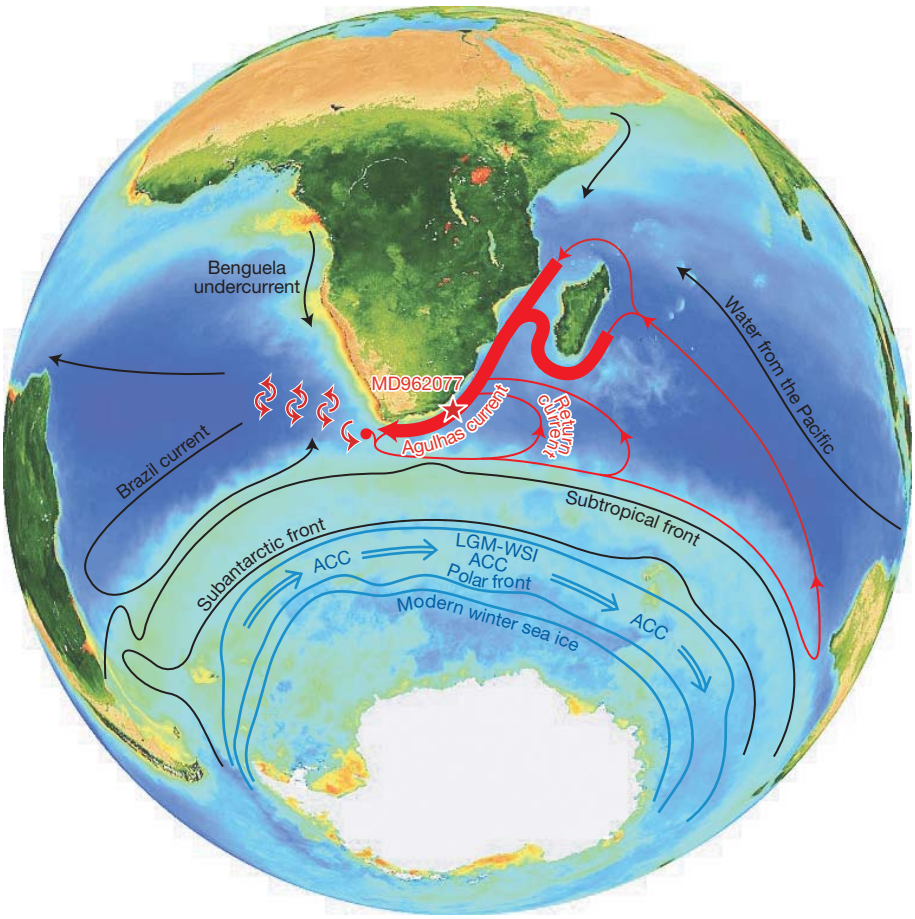


Figure 1 | SeaWiFS (Sea-viewing Wide Field-of-view Sensor) image of ocean colour during austral summer with the highly productive STF to the south of core MD962077 (red star). The core is at present in the path of the Agulhas current, and is influenced by the southward flow of warm and saline waters that are ultimately transported from the Indian to the Atlantic oceans today. The core site is highly sensitive to meridional movements of the fronts in this critical region. LGM-WSI, Last Glacial Maximum winter sea ice extent.

the STF advanced by at least 7° latitude to reach the lowest latitude (33 °S) of the past 800 kyr, as marked by the additional 2 °C cooling that is characteristic of the SST gradient of the modern STF. The STF impinged upon the southern African continent in the Indian Ocean during these extreme glacial periods and almost completely shut down the Agulhas leakage[8,10,11], with a tiny residual, perhaps intermittent, flow to account for the rare occurrence of the Agulhas leakage fauna in cores from the Atlantic sector[11,12]. An associated migration northwards of the southwesterly winds by 5–9° latitude during glacial periods is corroborated by a number of proxy records[13]. This shift and change in intensity of the winds is difficult to capture consistently in intermediate-complexity models[14] and coupled global climate models[15].

We propose that a northerly migration of the STF relative to South Africa acts as a choke point in the climate system. The latitude of the STF has the potential to influence both surface buoyancy fluxes of heat and salt that set the strength and shape of the AMOC[16] and also mixing and wind-driven upwellings that input energy to sustain the AMOC.

As a buoyancy driver of the AMOC, the transport flux of heat and salt in surface waters between the major ocean basins is crucial to return water, salt and heat to the northward flow of the AMOC[17,18]. The exchange of surface water occurs along a 'warm' route and a 'cold' route; the balance between these routes may alter as oceanographic fronts migrate relative to the tips of continents in the Southern Ocean[19]. Important topographic valves restrict the water flow in the Indonesian archipelago, Drake passage, Bering strait and the Agulhas valve of the southern tip of Africa[5]. These valves were drastically different during the 100–150 m sea-level drop of glaciations and could control ocean currents through threshold effects. Our inferred most northerly migration of the STF to 33 °S during MIS-12 and MIS-10 severely reduced the warm saline influence of Indian Ocean thermocline water and the Atlantic salinity budget owing to the Agulhas leakage[5,18–20]. This effect causes cooling and freshening of the Atlantic and implies maximal reduction in North Atlantic deep water formation during MIS-12 and MIS-10.

Energy input to maintain the AMOC derives, in part, from the coupling between strong circumpolar southwesterly winds, which induce a vigorous ACC, upwelling of dense waters from the ocean interior, and northward Ekman transport of waters near the ocean surface[21] as well as southward eddies. Today, the frontal system and associated southwesterly winds cross the continent of South America, which diverts the ACC flow southwards through the Drake passage (56° S). At this point, the ACC is decoupled from the winds and frontal systems that transfer energy to the current[21]. For the rest of the ACC pathway in the modern ocean, the westerly winds align with the ACC. We infer that either the core or meridional extent of the wind field of the southwesterly winds spread to their most northerly position during MIS-12 and MIS-10, and encountered two further continental barriers (Australia at 39 °S, and South Africa at 35 °S), decoupling the topographically constrained ACC to the south from a larger fraction of the frontal-associated winds and reducing the energy transfer to the ocean[21]. These additional circuit breaks in energy transfer from the southwesterly winds to the ACC would amplify a decrease in deep ventilation, on top of a severely reduced Agulhas leakage, during MIS-12 and MIS-10.

To test whether northwards migration of the STF is communicated globally via an influence on AMOC, we consider our data in the context of a record of the $\delta^{13}C$ contrast between deep waters in the

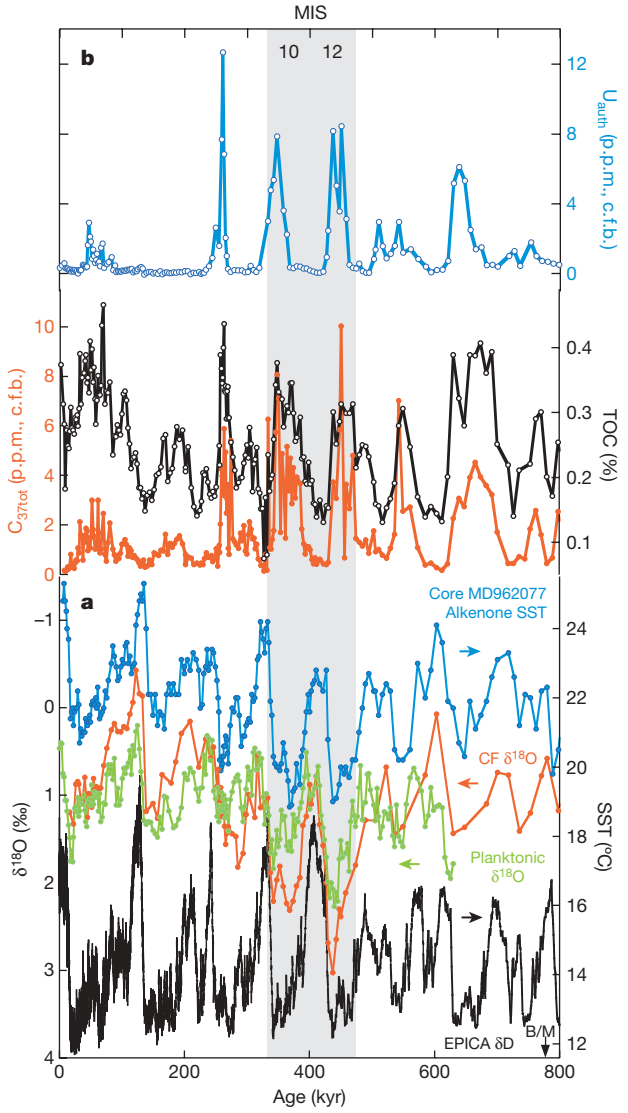


Figure 2 | Long-term trends in SST and productivity-sensitive parameters from core MD962077 compared to the stable baseline glacial temperatures recorded in EPICA Dome C. **a**, Coccolith fraction $\delta^{18}O$ (CF, red dots, left axis), *G. inflata* $\delta^{18}O$ (planktonic, green dots, left axis), and U^{K}_{37}' (alkenone, blue dots, right axis) highlight the evolution of subtropical region glacial SSTs, compared to the stable baseline temperature estimates (black line, the relative abundance of deuterium, δD, scaled according to $(T + 34)/2$ to fit the left axis) from glacial periods in Antarctica recorded in EPICA. The black arrow labelled 'B/M' shows the position of the Brunhes/Matuyama in the deeper section of core MD962077. **b**, Long-term trends of productivity proxies—U_{auth} in parts per million, p.p.m. (blue open circles), percentage TOC (black open circles) and C_{37tot} (red filled circles)—which show high productivity during each glacial period indicative of a northwards migration of the STF. This migration evolves, the STF being at its most northerly position during MIS-12 and MIS-10. (c.f.b., carbonate-free basis; see online-only Methods.)

Atlantic and Pacific oceans, an indicator of the efficiency of deep-ocean ventilation (Fig. 3, Supplementary Information). Our secular trends in temperature and frontal location are accompanied by a glacial decreased contrast between the Pacific and Atlantic deep water masses ($\Delta\delta^{13}C$), with near-zero $\Delta\delta^{13}C$ during MIS-12 and MIS-10; a picture also obtained from the $\delta^{13}C$ vertical gradient in the Southern Ocean[22]. This lack of $\delta^{13}C$ gradient could indicate drastically reduced ventilation of the Atlantic, or enhanced ventilation in the Pacific. Both changes could have occurred with a similar ventilation contribution from the Southern Ocean. Recent numerical simulations of a latitudinal shift of southwesterly winds support a scenario with a decrease of Agulhas leakage by a factor of seven, and a substantial

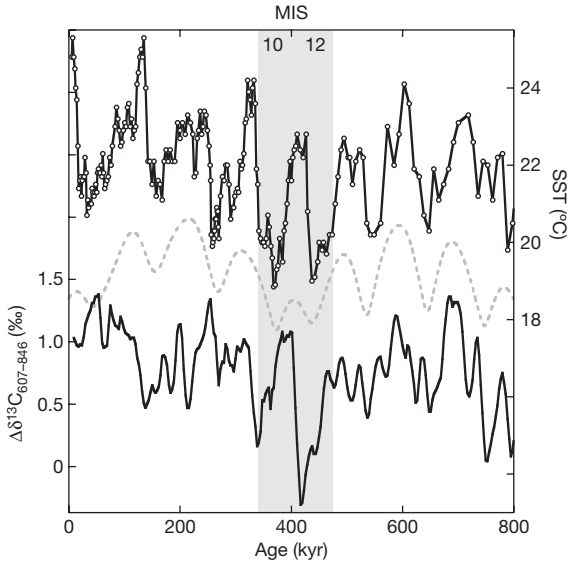


Figure 3 | Mechanistic link between the migration of the STF, orbital eccentricity and ocean overturning. The position of the STF, denoted by our record of SST (black, open circles) co-evolved with orbital eccentricity (grey dashed line) and a measure of overturning strength (black line) based on the $\delta^{13}C$ gradient between the Atlantic (Ocean Drilling Program ODP site 607) and the Pacific (ODP site 846, $\Delta\delta^{13}C$) oceans. This provides an independent measure of the ventilation strength[31] of the AMOC over the last 800 kyr. We derive the $\Delta\delta^{13}C$ value from the difference between the $\delta^{13}C$ values of ODP Site 607 and ODP Site 847.

decrease of North Atlantic deep water transport, with a concomitant increase and deepening of North Pacific convection[20]. These effects result from the wind shift and the enhanced build-up of Pacific surface salinity.

The decrease in ventilation by the AMOC leads to reduced polar heat transport and amplifies the growth of northern ice sheets, and also leads to lowest sea levels during MIS-12 and MIS-10. These frontal choke points of the STF impinging on Australia and South Africa transfer the tropical cooling of MIS-12 and MIS-10 to a global climate signal.

The position of the STF during MIS-12 and 10 could be a response to cooler global climates. Alternatively, a more regional forcing of anomalously cool equatorial temperatures could modulate the position of the STF.

Global cooling drives an equatorward migration of the STF. A global cooling of 3 °C achieves a 7° equatorward migration of the fronts and the peak westerly wind stress position, according to an idealized model[23]. The residual between the EPICA (European Project for Ice-Coring in Antarctica) Dome C temperature record and the temperature component that is linearly linked with greenhouse gas radiative forcing shows a long-term trend of decreasing temperatures towards 400 kyr and increasing towards modern times, in parallel with the longer-term trend of the STF position[24]. However, the minimum levels of p_{CO_2} during MIS-12 and MIS-10 are not exceptionally low compared to other glacial periods[1]. Although increased global coccolithophore production[25] implies a quasi-400-kyr change in the global carbon cycle, Pleistocene trends in p_{CO_2} are compensated biogeochemically[25] or masked by higher-frequency cycles in the record. Exceptional glacial conditions during MIS-12 and MIS-10 require an alternative trigger and amplifier.

Insolation could induce anomalous tropical SSTs that determine the strength of Hadley circulation and position of the STF. Our SST and productivity records show a strong signal at the 100-kyr and 400-kyr frequencies of eccentricity (Fig. 3). Eccentricity modulation of precession may lead to anomalous tropical SSTs owing to a truncated response[26], or to the double maximum of daily irradiation characteristic of the tropics[27,28] rather than the high latitudes.

Forcing at 100-kyr and 400-kyr frequencies is amplified, relative to half-precession, by increasing the fraction of the year over which the maximum and minimum equatorial insolation is time-averaged[28]. Such time-averaging may best approximate the thermal inertia of the tropical SST response to maximum equatorial insolation. Anomalously cool tropical SSTs, caused by low maximum equatorial insolation (modern and about 400 kyr ago), could reduce the Hadley cell (the tropical atmospheric convection cell driven by equatorial heating), draw the STF northwards[28] and act as a trigger for the proposed AMOC positive feedback for exceptional global cooling during MIS-12 and MIS-10.

An implication of the amplification of cooling during MIS-12 and MIS-10 is that there must be some decoupling between Antarctic climate, global climate and p_{CO_2} (ref. 1). The central Antarctic ice sheet may be decoupled from global climatic shifts either by the slow transmission of signals across a strengthened glacial ACC[29] or by the enlarged extent of sea ice during the glacial periods insulating the Antarctic continent from changes in the ocean, analogous to the buffering of Greenland during the extreme cold of Heinrich events by sea ice and the southward migration of the deep ventilation site[30].

It appears that p_{CO_2} can be decoupled from climate at glacial extremes such that insolation modulates the maxima and minima of climate. Modulation of equatorial insolation provides a way of amplifying eccentricity forcing into the climate cycles of Pleistocene glaciations. A proposed mechanism of links between topographic barriers, the STF and the AMOC may explain the puzzling fact that most palaeo-records indicate that glacial periods have different amplitudes. Understanding the effect of the Agulhas valve on ocean circulation is important for modern and future climate studies, as illustrated by recent work on decadal variations of the southwesterly winds[6].

METHODS SUMMARY

The chronology of core MD962077 relies on the last three geomagnetic reversals and tuning of the foraminiferal $\delta^{18}O$ to a global stack. Alkenones were analysed at CEREGE by gas chromatography using an automated Dionex accelerated solvent extractor (ASE-200) to perform lipid extraction. Core MD962077 was sampled every 10–20 cm in plastic vials to measure U, Th and other trace elements by inductively coupled plasma mass spectrometry (ICP-MS). We measured $\delta^{18}O$, in parts per thousand relative to the Vienna Pee Dee Belemnite standard (‰ VPDB) on the multispecies coccolithophore fraction, with minimal contamination from other carbonate components such as foraminiferal fragments. The foraminiferal $\delta^{18}O$ record was previously published[12]. Given the limited constraints on long-term whole-ocean changes in $\delta^{18}O$, and salinity at core MD962077, we have not deconvolved a record of SST from planktonic $\delta^{18}O$. Nonetheless, MIS-12 $\delta^{18}O$ is heavier than MIS-6 by 0.5‰, which is equivalent to a maximum estimate of additional cooling of $\sim 2\,^{\circ}C$. The coccolith fraction $\delta^{18}O$ records the largest isotopic change: MIS-12 is heavier than MIS-2 by 1.5‰, but changes in the coccolithophore assemblage occur on these timescales in core MD962077[25] and large interspecific fractionation factors of up to 4‰ can account for a fraction of this change.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 January; accepted 29 May 2009.

1. Luthi, D. *et al.* High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* **453,** 379–382 (2008).
2. Svendsen, J. I. *et al.* The Quaternary ice sheet history of northern Eurasia. *Quat. Sci. Rev.* **23,** 1229–1271 (2004).
3. Rohling, E. J. *et al.* Magnitude of sea-level lowstands of the past 500,000 years. *Nature* **394,** 162–165 (1998).
4. McManus, J. F. *et al.* A 0.5-million-year record of millennial-scale climate variability in the North Atlantic. *Science* **283,** 971–975 (1999).
5. Lutjeharms, J. R. E. in *The South Atlantic: Present and Past Circulation* (eds Wefer, G., Berger, W. H., Siedler, G. & Webb, D.) 125–162 (Springer, 1996).
6. Biastoch, A., Boning, C. W. & Lutjeharms, J. R. E. Agulhas leakage dynamics affects decadal variability in Atlantic overturning circulation. *Nature* **456,** 489–492 (2008).
7. Flores, J. A., Gersonde, R. & Sierro, F. J. Pleistocene fluctuations in the Agulhas Current Retroflection based on the calcareous plankton record. *Mar. Micropaleontol.* **37,** 1–22 (1999).
8. Peeters, F. J. C. *et al.* Vigorous exchange between the Indian and Atlantic Oceans at the end of the past five glacial periods. *Nature* **430,** 661–665 (2004).
9. Howard, W. R. & Prell, W. L. Late Quaternary Surface circulation of the Southern Indian Ocean and its relationship to orbital variations. *Paleoceanography* **7,** 79–117 (1992).
10. Berger, W. H. & Wefer, G. in *The South Atlantic: Present and Past Circulation* (eds Wefer, G., Berger, W. H., Siedler, G. & Webb, D. J.) 363–410 (Springer, 1996).
11. Bé, A. W. & Duplessy, J. C. Subtropical convergence fluctuations and quaternary climates in the middle latitudes of the Indian Ocean. *Science* **194,** 419–422 (1976).
12. Rau, A., Rogers, J. & Chen, M.-T. Quaternary palaeoceanographic record in giant piston cores off South Africa, possibly include evidence of neotectonism. *Quat. Int.* **148,** 65–77 (2006).
13. Ledru, M. P., Rousseau, D. D., Riccomini, F. W. C. Jr, Karmann, I. & Martin, L. Paleoclimate changes during the last 100,000 yr from a record in the Brazilian Atlantic rainforest region and interhemispheric comparison. *Quat. Res.* **64,** 444–450 (2005).
14. Menviel, L., Timmerman, A., Mouchet, A. & Timm, O. Climate and marine carbon cycle response to changes in the strength of the Southern Hemispheric westerlies. *Paleoceanography* **23,** doi:10.1029/2008PA001604 (2008).
15. Rojas, M. *et al.* The Southern Westerlies during the last glacial maximum in PMIP2 simulations. *Clim. Dyn.*. doi:10.1007/s00382–008–0421–7 (2008).
16. Kuhlbrodt, K. *et al.* On the driving processes of the Atlantic Meridional Overturning Circulation. *Rev. Geophys.* **45,** RG2001 (2004).
17. Gordon, A. L. Interocean exchange of thermocline water. *J. Geophys. Res.* **91,** 5037–5046 (1986).
18. Weijer, W., de Ruijter, W. P. M., Sterl, A. & Drujfhout, S. S. Response of the Atlantic overturning circulation to South Atlantic sources of buoyancy. *Global Planet. Change* **34,** 293–311 (2002).
19. Knorr, G. & Lohmann, G. Southern Ocean origin for the resumption of Atlantic thermohaline circulation during deglaciation. *Nature* **424,** 532–536 (2003).
20. Sijp, W. P. & England, M. H. Southern Hemisphere westerly wind control over the ocean's thermohaline circulation. *J. Clim.*. (in the press).
21. Toggweiler, J. R. & Russell, J. Ocean circulation in a warming climate. *Nature* **451,** 286–288 (2008).
22. Hodell, D. A., Venz, K. A., Charles, C. D. & Ninnemann, U. S. Pleistocene vertical carbon isotope and carbonate gradients in the South Atlantic sector of the Southern Ocean. *Geochem. Geophys. Geosyst.* **4,** doi:10.1029/2002GC000367 (2002).
23. Williams, G. P. & Bryan, K. Ice age winds: an aquaplanet model. *J. Clim.* **19,** 1706–1715 (2006).
24. Masson-Delmotte, V. *et al.* EPICA Dome C record of glacial and interglacial intensities. *Quat. Sci. Rev.* (in the press).
25. Rickaby, R. E. M. *et al.* Coccolith chemistry reveals secular variations in the global ocean carbon cycle? *Earth Planet. Sci. Lett.* **253,** 83–95 (2007).
26. Short, D. A., Mengel, J. G., Crowley, T. J., Hyde, W. T. & North, G. R. Filtering of Milankovitch cycles by Earths geography. *Quat. Res.* **35,** 157–173 (1991).
27. Berger, A., Loutre, M. F. & Melice, J. L. Equatorial insolation: from precession harmonics to eccentricity frequencies. *Clim. Past* **2,** 131–136 (2006).
28. Ashkenazy Y. & Gildor, H. Timing and significance of maximum and minimum equatorial insolation. *Paleoceanography* **23,** PA1206, doi:1029/2007001436 (2008).
29. Schmittner, A., Yoshimori, M. & Weaver, A. J. Instability of glacial climate in a model of the ocean-atmosphere-cryosphere system. *Science* **295,** 1489–1493 (2002).
30. Ganopolski, A. & Rahmstorf, S. Rapid changes of glacial climate simulated in a coupled climate model. *Nature* **409,** 153–158 (2001).
31. Raymo, M. E. *et al.* Stability of North Atlantic water masses in face of pronounced climate variability during the Pleistocene. *Paleoceanography* **19,** PA2008, doi:10.1029/2003PA000921 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. P. Schrag and E. Goddard for their help with analyses in the early stages of this work. We thank C. Sonzogni, F. Rostek, N. Thouveny and J. Carignan for help with analyses and age model, and D. Sansom for help with the figures. We also thank H. Gildor, Y. Ashkenazy, R. Toggweiler, M. Meredith, M.-F. Loutre, P. Huybers and N. Edwards for comments and discussion on an earlier version of this manuscript. R.E.M.R. is grateful to the Royal Society for the International Outgoing Visit award for the exchange visits to CEREGE, which facilitated the development of the ideas and the manuscript. Palaeoclimate work at CEREGE is supported by grants from the Gary Comer Foundation, the CNRS and the Collège de France. Core MD962077 was collected by the RV *Marion Dufresne* supported by the Institut Polaire Français (IPEV).

Author Contributions E.B. and R.E.M.R. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.B. (bard@cerege.fr) or R.E.M.R. (rosr@earth.ox.ac.uk).

METHODS

The chronology of core MD962077 relies on the last three geomagnetic reversals and tuning the foraminiferal $\delta^{18}O$ to the global stack[32]. Alkenones were analysed at CEREGE by gas chromatography following the method of ref. 33 except that an automated Dionex accelerated solvent extractor (ASE-200) was used to perform lipid extraction[34]. Proper identification and quantification of alkenones was verified by gas chromatography mass spectrometry. The analytical precision of the method is about 0.01 units for U^{K}_{37}' (roughly equivalent to 0.3 °C) based on repeated extraction of internal laboratory sediment standard during the course of this study. The accuracy of our procedure has been checked within the framework of the international alkenone intercomparison[35]. For simplicity and to allow comparison with other palaeo-data sets, the SST values were calculated using the equation of ref. 36 which is equivalent to that derived from the core-top compilation[37]. Culture studies show that there is no systematic offset between SST relationships of *Gephyrocapsa oceanica* and *Emiliania huxleyi*[38].

We measured $\delta^{18}O$ on the multispecies coccolithophore fraction, with minimal contamination from other carbonate components such as foraminiferal fragments. The methodology for separating the coccolithophore fraction was based on that of ref. 39. All plasticware was acid-cleaned before use, and deionized water used throughout. 0.5 g of sediment was sieved using 90% proof ethanol through a $<$20 μm sieve. The collected slurry was poured into 15 ml centrifuge tubes and allowed to settle for 10 min. The supernatant containing the suspended particles was pipetted into a second tube and allowed to settle for a further 24 h. The supernatant, considered to contain the major clay fraction, was pipetted away and filtered for recycling. The remaining $<$10 μm fraction was dried and a subsample taken and rinsed three times with deionized water to ensure elimination of any relict ethanol. A small fraction was weighed into boats for analysis by stable isotope mass spectrometry (PRISM at Harvard University). The foraminiferal $\delta^{18}O$ record was previously published[12]: for each analysis, 60 specimens of *Globorotalia inflata* were picked from the 250–350 μm fraction and analysed at the National Taiwan Ocean University in Keelung using standard procedures. Given the limited constraints on long-term whole-ocean changes in $\delta^{18}O$, and salinity at core MD962077, we have not deconvolved a record of SST from planktonic $\delta^{18}O$. Nonetheless, MIS-12 $\delta^{18}O$ is heavier than MIS-6 by 0.5‰, equivalent to a maximum estimate of additional cooling of ~2 °C. The coccolith fraction $\delta^{18}O$ records the largest isotopic change: MIS-12 is heavier than MIS-2 by 1.5‰, but changes in the coccolithophore assemblage occur on these timescales in this core[25] and large interspecific fractionation factors of up to 4‰ can account for a fraction of this change[40]. Measuring three thermometric proxies such as U^{K}_{37}' of alkenones and $\delta^{18}O$ of both foraminifera and coccoliths is important because these independent proxies provide confirmation of temperature trends. Indeed, it is crucial to be able to rule out possible artefacts related to species, seasonality or habitat depth of signal markers and of post-deposition processes such as sediment mixing. Differential smoothing or lagging can occur in sediments between different size fractions of the same core[41]. In core MD962077, the $\delta^{18}O$ of coccoliths is measured in the same fine fraction as alkenones, while the $\delta^{18}O$ of foraminifera belong to the coarse fraction. When similar trends are observed for the three proxies (as is the case), it is possible to rule out such an artefact as an alternative explanation to temperature changes.

Core MD962077 was sampled every 10–20 cm in plastic vials to measure U, Th and other trace elements by ICP-MS. Procedures are similar to those used in our previous work[42] except that ICP-MS measurements were performed with a Perkin-Elmer Elan 6000 at Service d'Analyse des Roches et des Mineraux (SARM) of the CNRS in Nancy, France. For both elements, the uncertainty is $<$5% and the determination level is 0.01 p.p.m. U_{auth} concentrations were calculated by subtracting a crustal component determined from the measured Th concentration and a detrital Th/U weight ratio of 4.71 (ref. 43). The detrital U correction is relatively small (about 1 p.p.m. on average) in comparison with the observed U variations. Moreover, U and U_{auth} profiles are roughly parallel and the two records exhibit the same variability. Therefore, using either the U or U_{auth} profile would not change the interpretation of our results. All concentrations shown in Fig. 2b were converted to concentrations on a carbonate-free basis (c.f.b.), which is the conventional procedure to account for the effect of dilution by carbonate (see ref. 42 for example). The U_{auth}, TOC and C_{37tot} concentration profiles with and without this correction are roughly parallel and using raw concentrations would lead to the same interpretation.

TOC, C_{37tot} and U_{auth} concentrations have been used as proxies for the biological productivity in the Southern Ocean related to migrations of the STF and the subantarctic front[44,45]. Following the work of these authors and from our own studies in other regions[42,46] it is clear that these proxies cannot be scaled linearly to productivity. Indeed, the organic carbon raining from the surface is partly remineralized in the sediments, which creates more-reducing conditions, leading to U diffusion in pore waters and trapping in the sediments[44]. These reducing conditions also lead to preferential preservation of lipids such as alkenones, which always exhibit a much larger dynamic range than observed with TOC[46]. Trace metals can also be remobilized by additional processes such as oxic front diffusion and bioturbation[47]. All these natural processes are superimposed on each other, making these markers only qualitative, although very sensitive, proxies for the surface productivity. Despite these complications, the strength of our approach is that we consider all three proxies measured independently. When TOC, C_{37tot} and U_{auth} trends agree, indicating enhanced values, it is reasonable to assume that the observed maxima are associated with a significant productivity increase linked to the STF migration. MIS-12 and MIS-10 are indeed characterized by maxima of TOC, C_{37tot} and U_{auth}.

32. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography* **20**, PA1003 (2005).
33. Sonzogni, C. *et al.* Core-top calibration of the alkenone index vs sea surface temperature in the Indian Ocean. *Deep-Sea Res.* **44**, 1445–1460 (1997).
34. Pailler, D. & Bard, E. High frequency paleoceanographic changes during the past 140,000 years recorded by the organic matter in sediments off the Iberian Margin. *Palaeogeogr. Palaeoclim. Palaeoecol.* **181**, 431–452 (2002).
35. Rosell-Melé, A. *et al.* Precision of the current methods to measure the alkenone proxy UK37′ and absolute abundance in sediments: results of an interlaboratory comparison study. *Geochem. Geophys. Geosyst.* **2**, 2000GC000141, 1–28 (2001).
36. Prahl, F. G., Muehlhausen, L. A. & Zahnle, D. L. Further evaluation of long-chain alkenones as indicators of paleoceanographic conditions. *Geochim. Cosmochim. Acta* **52**, 2303–2310 (1988).
37. Müller, P. J., Kirst, G., Ruhland, G., von Storch, I. & Rosell-Mele, A. Calibration of the alkenone paleotemperature index U^{k}_{37}' based on core-tops from the eastern South Atlantic and the global ocean (60°N–60°S). *Geochim. Cosmochim. Acta* **62**, 1757–1772 (1998).
38. Conte, M., Thompson, A., Lesley, D. & Harris, R. P. Genetic and physiological influences on the alkenone/alkenoate versus growth temperature relationship in *Emiliania huxleyi* and *Gephyrocapsa oceanica*. *Geochim. Cosmochim. Acta* **62**, 51–68 (1998).
39. Stoll, H. M. & Schrag, D. P. Coccolith Sr/Ca as a new indicator of coccolithophorid calcification and growth rate. *Geochem. Geophys. Geosyst.* **1**, doi:10.1029/1999GC000015 (2000).
40. Ziveri, P. *et al.* Stable isotope ''vital effects'' in coccolith calcite. *Earth Planet. Sci. Lett.* **210**, 137–149 (2003).
41. Bard, E. Paleoceanographic implications of the difference in deep-sea sediment mixing between large and fine particles. *Paleoceanography* **16**, 235–239 (2001).
42. Pailler, D. *et al.* Burial of redox-sensitive metals and organic matter in the equatorial Indian Ocean linked to precession. *Geochim. Cosmochim. Acta* **66**, 849–865 (2002).
43. Taylor, S. R. & McLennan, S. M. *The Continental Crust: Its Composition and Evolution* (Blackwell, 1985).
44. Rosenthal, Y., Boyle, E. A., Labeyrie, L. & Oppo, D. Glacial enrichments of authigenic Cd and U in subantarctic sediments—a climatic control on the element's oceanic budget. *Paleoceanography* **10**, 395–413 (1995).
45. Sachs, J. P. & Anderson, R. F. Increased productivity in the Subantarctic ocean during Heinrich events. *Nature* **434**, 1118–1121 (2005).
46. Schulte, S. & Bard, E. Past changes in biologically mediated dissolution of calcite above the chemical lysocline recorded in Indian ocean sediments. *Quat. Sci. Rev.* **22**, 1757–1770 (2003).
47. Zheng, Y., Anderson, R. F., Van Geen, A. & Fleisher, M. Q. Remobilization of authigenic uranium in marine sediments by bioturbation. *Geochim. Cosmochim. Acta* **66**, 1759–1772 (2002).

LETTERS

Global patterns of speciation and diversity

M. A. M. de Aguiar[1,2], M. Baranger[1,3], E. M. Baptestini[2], L. Kaufman[1,4] & Y. Bar-Yam[1]

In recent years, strikingly consistent patterns of biodiversity have been identified over space, time, organism type and geographical region[1,2]. A neutral theory (assuming no environmental selection or organismal interactions) has been shown to predict many patterns of ecological biodiversity[2,3]. This theory is based on a mechanism by which new species arise similarly to point mutations in a population without sexual reproduction. Here we report the simulation of populations with sexual reproduction, mutation and dispersal. We found simulated time dependence of speciation rates, species–area relationships and species abundance distributions consistent with the behaviours found in nature[1–13]. From our results, we predict steady speciation rates, more species in one-dimensional environments than two-dimensional environments, three scaling regimes of species–area relationships and lognormal distributions of species abundance with an excess of rare species and a tail that may be approximated by Fisher's logarithmic series. These are consistent with dependences reported for, among others, global birds[4] and flowering plants[5], marine invertebrate fossils[6], ray-finned fishes[7], British birds[8,9] and moths[10], North American songbirds[11], mammal fossils from Kansas[12] and Panamanian shrubs[13]. Quantitative comparisons of specific cases are remarkably successful. Our biodiversity results provide additional evidence that species diversity arises without specific physical barriers[6,11,14]. This is similar to heavy traffic flows, where traffic jams can form even without accidents or barriers[15].

Speciation studies have identified conditions under which speciation events can occur. Allopatry is considered the dominant form of speciation. Studies of partial barriers (parapatric speciation) and debates about sympatric speciation focus on whether such types of speciation are possible and, if so, whether they are plausible in nature[16,17]. Simulations of homogeneous spatial environments have suggested the necessity of trait-based variation of sexual selection, competition or diffusion rate, or use outbreeding depression[18–20]. Studies exist indicating these may not be necessary even in sympatry, at least where there is no genetic linkage[21,22]. Significantly for the results we report here, no simulations of speciation that include sexual mating have yet been shown to yield the observed patterns of species diversity.

We simulated the evolution of a population whose members, at the beginning, are uniformly distributed in space and have identical genomes. The population evolves under the combined influences of sexual reproduction, mutations and dispersal. During reproduction, potential mates are identified from among those in a spatial region around an individual (specified by a spatial mating distance, S) whose genomes are sufficiently similar to that of the individual (specified by a genetic mating distance, G). This is a minimal form of sexual selection, essential (necessary but not sufficient) for speciation, called assortative mating[23] (postzygotic genetic incompatibilities may have a role but are not essential[24]). A mate is chosen from this set at random. Reproduction with crossover and mutation occurs. An offspring is then dispersed within a region around the originating and expiring parent. Genetic variation grows over time, due to mutation and recombination. We identify a species as a group of organisms reproductively separated from all others by the genetic restriction on mating and connected among themselves by the same condition, without requiring all members of the group to be able to mate with each other. An example of such a case is a ring species[14], in which progressive differences along a chain of individuals result in individuals at one end not being able to mate with those at the other end. We consider this a single species, owing to the ongoing gene flow. No condition on spatial proximity is imposed on the members of a species.

We simulated many variants of the model; the one we report here uses haploid and hermaphroditic individuals placed randomly in the space (for convenience annotated as a 128×128 or 256×256 lattice). Multiple individuals can exist at the same site but typically do not, owing to low densities. Unlike parapatric simulations that are otherwise similar[24–26], our density of organisms is less than one per site (there are no a-priori-defined groups whose genetic divergence is a possible mechanism of speciation), so our speciation may be termed 'topopatric'. Genomes consist of binary strings of length 125. Variation in the number of offspring of an individual was included by randomly choosing a neighbour to reproduce, instead of the expiring parent, in $Q = 30\%$ of cases. Other assumptions, including variations of parameters or more basic changes such as separating two sexes, give qualitatively similar results (Methods).

Depending on the model parameters, we find sexual isolation of subpopulations in genome space (speciation) and spatial inhomogeneity in spite of the absence of geographical barriers, resource gradients or natural selection. We show an example in Fig. 1. The first branching into two species occurs after about three hundred generations; additional speciation events take place later. Each of the species is coloured differently. Species extinctions also occur, owing to random variation in population size and resulting in a variable number of species over time. Although species are not homogeneously distributed and tend to be concentrated in different regions, there is considerable overlap among their domains. Newly branched species are genetically close to their parent species, whereas species formed in the far past reach the maximum average genetic distance in the genome space. The species that form are completely separated by genetic difference according to the mating condition. The spatial and genetic clustering occurs even though the underlying mating behaviour couples organisms in both physical and genetic space.

In Fig. 2, we show the effect on the pattern of speciation of varying the spatial, S, and genetic, G, mating distances. Speciation becomes less likely as the critical mating distances increase. Neither spatial nor genetic restrictions on mating alone leads to speciation. The former is required for the distinction of spatial location, and the latter is required for genetic isolation. The largest number of species arises for small values of both parameters. However, for intermediate values of spatial distance (for example $S = 6$ in Fig. 2), intermediate values of genetic distance yield the most species, as small values of the

[1]New England Complex Systems Institute, Cambridge, Massachusetts 02138, USA. [2]Universidade Estadual de Campinas, Unicamp, Campinas, São Paulo 13083-970, Brazil. [3]University of Arizona, Tucson, Arizona 85719, USA. [4]Boston University, Boston, Massachusetts 02215, USA.

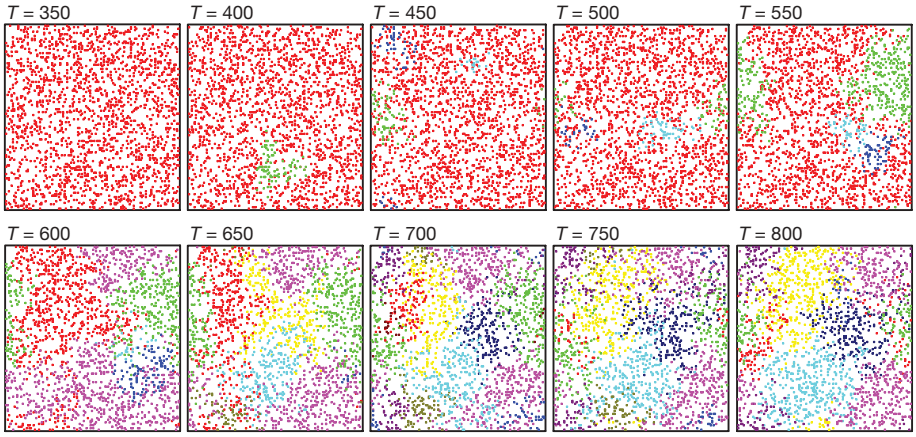


Figure 1 | Time evolution for 2,000 individuals on a 128 × 128 lattice. Maximum mating distance between two organisms is $S = 6$ lattice cells, and $G = 20$ genetic differences out of 125 total genes. Reproductively isolated species are shown in different colours.

genetic distance reduce the probability of genetic fluctuations arising from mutation and recombination that lead to speciation.

The average spatial density of a species over a population (Supplementary Fig. 2a, b) is consistently well described by a Gaussian

$$\rho(r) = \rho_0 e^{-r^2/R^2}$$

where r is the distance from the geographic centre, ρ_0 is approximately the average density, N/L^2 (where N is the number of organisms and L^2 is the area), and, in the regime of large numbers of species, R is given by

$$R = \alpha\sqrt{\frac{qG}{\mu}} + \beta S^2$$

with $\alpha = 0.18$ and $\beta = 0.26$. By integration, the number of species is $N_S = N/\pi\rho_0 R^2$.

The model describes the spatial dynamics of speciation, from which we can obtain the time dependence, the spatial dependence (species–area relationship) and the distribution of species sizes (species abundance distribution). In Fig. 3 (see also Supplementary Fig. 2c, d), we summarize the following four key predictions for species diversity consistent with observation.

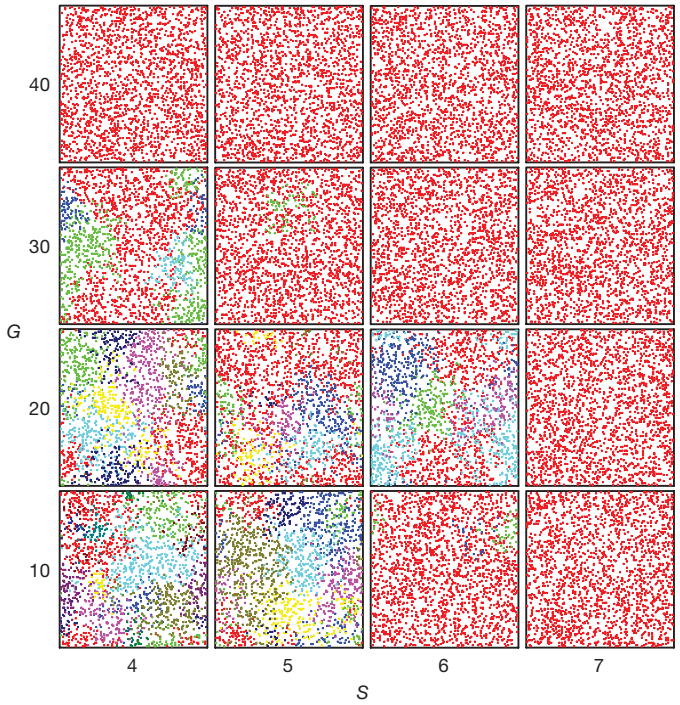


Figure 2 | Spatial snapshots after 1,000 generations for 2,000 individuals on a 128 × 128 lattice. S and G are as shown. Colours are for different species.

First, we predict constant rates of new species over long times, as captured by the fossil and genetic data. Figure 3a shows how the number of species varies over time. After a transient period, during which mutation and recombination increase the variation in the initially identical population, the number of species increases rapidly and reaches a steady state subject to statistical fluctuations. The existence of a steady state is sufficient to imply a constant rate of new species arising over long times. This result is consistent with the record of marine invertebrate fossils from the Phanerozoic eon[6]. Over shorter time frames, genetic differences among North American songbirds[11] and the fossils of mammals from the Meade basin of Kansas[12] show remarkably steady speciation rates (consistent with our results) and do not correlate with glaciations, disproving this once-dominant view, which was held because of the expectation that speciation is promoted by physical barriers. A constant speciation rate in the absence of landscape forcing is generally supported by the literature on whole-tree cladistic analysis[26]. This does not discount the potential importance of habitat expansion, such as occurred for North American organisms following glacial retreat.

Second, we predict a high diversity of freshwater versus marine ray-finned fishes when measured by area. Figure 3b shows that the number of species formed increases with the available area if the density of individuals (and corresponding values of other parameters) is kept constant. In one-dimensional environments, the number of species formed is much greater per unit area. Thus, the lower dimensionality of rivers relative to oceans has an enhanced effect on speciation even when specific barriers do not exist. We studied one-dimensional and two-dimensional cases, and river networks may be considered fractal with dimension between one and two, but the same considerations apply. This is consistent with the high diversity of freshwater ray-finned fishes when measured by area and similarly the preponderance of near-shore shallow-water marine species[7].

Third, we predict a species–area relationship that is largely independent of the type of species or geographical location, is nearly linear on large and small scales and has intermediate scaling behaviour with a low slope that varies among ecological communities[1,2]. The species–area relationship within a system is shown in Fig. 3c on a log–log plot. The simulated results correspond very well to those found, among others, in birds[4] and flowering plants[5] sampled on local to global scales, as well as exhaustive studies of trees and shrubs in Panama[13] (detailed fit, as shown). On intermediate scales, the slope can be varied with model parameters (inset, Fig. 3c).

Finally, we predict the distribution of species abundance to be well fit by a lognormal distribution with excess rare species. Simulated species abundance distributions (detailed correspondence shown in

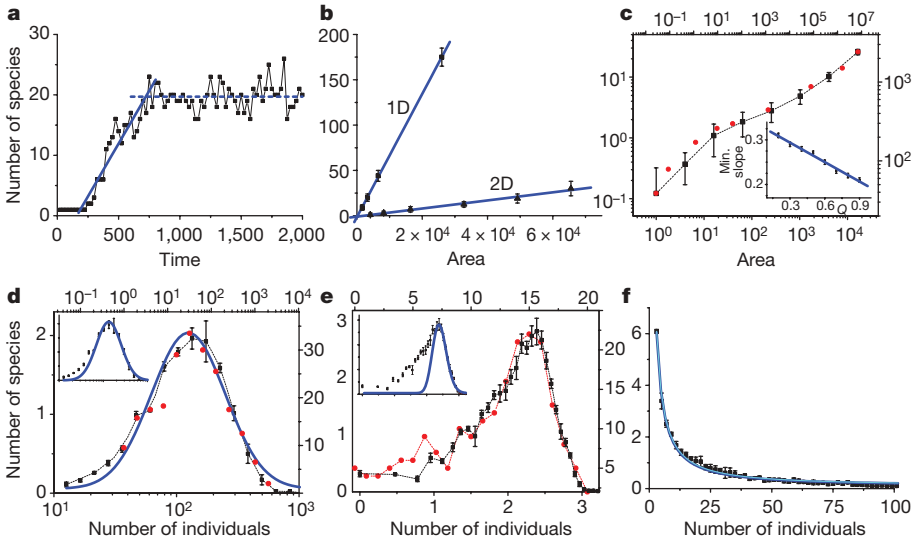


Figure 3 | Species abundance. **a**, Simulated number of species as a function of time ($S = 5$, $G = 20$, also used for **b**, **c**, **d**). Blue fits are a guide to the eye for the increase (solid) and the steady state (dashed). Time measured in generations. **b**, Simulated number of species for increasing system size in one (1D) and two (2D) dimensions ($P < 0.0001$). Area measured in lattice units. Error bars, s.e.m. **c**, Empirical species–area relationship from Panamanian trees[13] (red; upper and right-hand axes) compared with simulation (black; lower and left-hand axes) of samples with lattice sizes of up to 128×128, taken from a single simulation of 8,000 individuals on a 256×256 lattice (ranges shown are standard deviations across multiple simulations) fitted by anchoring lowest and highest points ($R^2 = 0.88$). Inset shows smallest slope of simulations as a function of Q (error bars, s.e.m.). **d**, Empirical species abundance distribution of Panamanian trees[3] (red; upper and right-hand axes) compared with simulation (black, with error bars; lower and left-hand axes; $\chi^2 = 3.1$, $P = 0.01$) and a lognormal distribution (blue; $\chi^2 = 3.9$, $P = 0.08$). Inset shows a lognormal distribution fitted only to highest and rightmost simulation points. Error bars, s.e.m. **e**, As in **d** (lognormal only in inset, and dashed red line added for clarity), but for British birds ($S = 6$, $G = 10$, $N = 8{,}000$, $Q = 0.5$) ($\chi^2 = 5.5$, $P < 0.0001$). The distribution is consistent with a larger effective mating distance, although other parameters may affect this conclusion. Even migratory birds satisfy the model assumptions of mating proximate to their locations of birth[30]. Error bars, s.e.m. **f**, Simulated species abundance for sampling over a 64×64 subset of a 256×256 lattice ($S = 6$, $G = 10$, $N = 8{,}000$; Supplementary Fig. 1), fitted to two essentially overlapping curves: lognormal (blue; $\chi^2 = 0.009$, $P < 0.0001$) and log-series (cyan; $\chi^2 = 0.005$, $P < 0.0001$). Error bars, s.e.m.

Fig. 3d, e) fit observations of Panamanian trees[3] and British birds[8,9], and are of the form found for moths[10] and various studies of trees[2], among others[2,3]. As the number of samples decreases relative to the number of species, the right-hand tail of the distribution, consisting of common species, dominates the sampling, leading to a distribution that can be fit both by the tail of the lognormal and by Fisher's logarithmic series[27] (Fig. 3f), as found among many observations[1,2] (Supplementary Fig. 1). Our species abundance rank distributions also display the observed S-shaped curve[2].

The key success of neutral theory with a point-mutation model of species formation[2,3] has been its ability to describe the forms of species–area relationships and species abundance distributions found among many species. Figure 3c–f shows that our model yields similar results. These results are found even though our mechanism includes not only the appearance of localized species, but also the splitting (fission) of a larger species into smaller populations. This does not correspond to allopatric speciation, owing to the absence of geographic barriers. Such splitting leads to key differences from the point-mutation model. For example, the rate at which the steady state arises is much faster because each new species need not start from a single individual. Moreover, the formation of species is not independent of genealogical history. Species fission might be expected to significantly change the behaviour of the species abundance distributions, but we nevertheless find a lognormal distribution with excess rare species and detailed agreement with observed species abundance and species–area relationships.

We note that there are also regimes of the model in which speciation is quite difficult. This accounts for the existence of some observations where geographic barriers and selection do have an important role; for example, observations of greenish warblers[14] have shown the direct impact of geographic barriers and genetic variation across continent-wide environmental gradients.

We have shown that our distributions of species reach a steady state rapidly, with extinctions balanced by speciation consistent with observations, and that species–area relationships and species abundance distributions are consistent with those frequently found in nature. The general forms of the distributions are as observed, and specific simulations are in remarkable correspondence with specific observations. Underlying the agreement is a simple model of a sexually reproducing population in a spatial context with mating restrictions depending on genetic and spatial distance—these are the essential properties for the dynamic biodiversity patterns. Our model dynamically subdivides spatial regions, which may be called topopatric speciation. The spontaneous creation of new spatial patterns is an example of the general process of spontaneous pattern formation and symmetry breaking found in many systems[28], including traffic jams[15]. Pre-existing geographic features and variation may influence patterns, but are not necessary for their formation. The relationship between this process and niche-ordering models[29] could be explored.

METHODS SUMMARY

We used agent-based simulations in which agents identified by geographical location and genotype undergo sexual reproduction in pairs limited by geographical and genetic proximity. Each offspring replaces a parent and differs in genotype according to genetic inheritance from both parents, with crossover and mutation, and in geographical location by dispersal. According to tests of multiple model variants, including parameter variation, sequential and synchronous mating, hermaphroditism and two sexes, single or multiple crossovers, direct assortative mating and separate fitness and sexual selection genetic components, our results apply quite generally, with the key properties needed being the limitations on spatial and genetic distances of mating.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 February; accepted 22 May 2009.

1. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
2. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).

3. Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Patterns of relative species abundance in rainforests and coral reefs. *Nature* **450,** 45–49 (2007).
4. Preston, F. W. Time and space and the variation of species. *Ecology* **41,** 611–627 (1960).
5. Schmida, A. & Wilson, M. V. Biological determinants of species diversity. *J. Biogeogr.* **12,** 1–20 (1985).
6. Alroy, J. *et al.* Phanerozoic trends in the global diversity of marine invertebrates. *Science* **321,** 97–100 (2008).
7. Moyle, P. B. & Cech, J. J. *Fishes: An Introduction to Ichthyology* 5th edn (Benjamin Cummings, 2003).
8. Gibbons, D. W., Reid, J. B. & Chapman, R. A. (eds) *The New Atlas of Breeding Birds in Britain and Ireland: 1988–1991* (Poyser, 1993).
9. Gregory, R. D. Species abundance patterns of British birds. *Proc. R. Soc. Lond. B* **257,** 299–301 (1994).
10. Williams, C. B. *Patterns in the Balance of Nature and Related Problems of Quantitative Biology* (Academic, London, 1964).
11. Klicka, J. & Zink, R. M. The importance of recent ice ages in speciation: a failed paradigm. *Science* **277,** 1666–1669 (1997).
12. Rosenzweig, M. L. Tempo and mode of speciation. *Science* **277,** 1622–1623 (1997).
13. D'Arcy, W. E. *Flora of Panama: Checklist and Index* Vols 1–4 (Missouri Botanical Garden, 1987).
14. Irwin, D. E., Bensch, S., Irwin, J. H. & Price, T. D. Speciation by distance in a ring species. *Science* **307,** 414–416 (2005).
15. Nagel, K. & Schreckenberg, M. A cellular automaton model for freeway traffic. *J. Phys. I* **2,** 2221–2229 (1992).
16. Doebelli, M., Dieckmann, U., Metz, J. A. J. & Tautz, D. What we have also learned: adaptive speciation is theoretically plausible. *Evolution* **59,** 691–695 (2005).
17. Gavrilets, S. "Adaptive speciation"—it is not that easy: a reply to Doebeli et al. *Evolution* **59,** 696–699 (2005).
18. Payne, R. J. H. & Krakauer, D. C. Sexual selection, space, and speciation. *Evolution* **51,** 1–9 (1997).
19. Doebeli, M. & Dieckmann, U. Speciation along environmental gradients. *Nature* **421,** 259–264 (2003).
20. Hoelzer, G. A., Drewes, R., Meier, J. & Doursat, R. Isolation-by-distance and outbreeding depression are sufficient to drive parapatric speciation in the absence of environmental influences. *PLoS Comput. Biol.* **4,** e1000126 (2008).
21. Higgs, P. G. & Derrida, B. Stochastic models for species formation in evolving populations. *J. Phys. A* **24,** L985–L991 (1991).
22. Kondrashov, A. S. & Shpak, M. On the origin of species by means of assortative mating. *Proc. R. Soc. Lond. B* **265,** 2273–2278 (1998).
23. Lewontin, R., Kirk, D. & Crow, J. Selective mating, assortative mating and inbreeding: definitions and implications. *Eugen. Q.* **15,** 141–143 (1968).
24. Dobzhansky, T. *Genetics and the Origin of Species* (Columbia Univ. Press, 1937).
25. Gavrilets, S., Li, H. & Vose, M. D. Patterns of parapatric speciation. *Evolution* **54,** 1126–1134 (2000).
26. Ricklefs, R. E. Estimating diversification rates from phylogenetic information. *Trends Ecol. Evol.* **22,** 601–610 (2007).
27. Fisher, R. A., Corbet, A. S. & Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12,** 42–58 (1943).
28. Bar-Yam, Y. *Dynamics of Complex Systems* (Perseus, 1997).
29. Sugihara, G., Bersier, L.-F., Southwood, T. R. E., Pimm, S. L. & May, R. M. Predicted correspondence between species abundances and dendrograms of niche similarities. *Proc. Natl Acad. Sci. USA* **100,** 5246–5251 (2003).
30. Greenwood, P. J. & Harvey, P. H. The natal and breeding dispersal of birds. *Annu. Rev. Ecol. Syst.* **13,** 1–21 (1982).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge internal support by the New England Complex Systems Institute. M.A.M.d.A. and E.M.B. acknowledge financial support from the Fundação de Amparo à Pesquisa do Estado de São Paulo and the Conselho Nacional de Desenvolvimento Científico e Tecnológico, and L.K. from the Marine Management Area Science Programof Conservation International and the Gordon and Betty Moore Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Y.B.-Y. (yaneer@necsi.edu).

METHODS

For the simulations reported in the paper, we considered N haploid and hermaphroditic individuals placed randomly on a lattice (128×128 or 256×256) where multiple individuals can exist at the same site. Genomes consist of binary strings of length B, and the genetic distance between two individuals is the number of differences along the genome. Reproduction starts with an individual looking for a mate inside its spatial mating radius, S. To qualify, an individual must have a genetic distance no greater than G from the seeker, reflecting differences in traits or offspring viability. From all such potential mates, we select one at random. If there are fewer than a minimum number of possible mates (an additional model parameter, P), we weaken the spatial and genetic constraints as follows for that mating: $S \rightarrow S + 1$ and $G \rightarrow G + 1$. If the number of available mates is still smaller than P, the process is repeated. (We also investigated letting $S \rightarrow S + 1$ without changing G and, if there were not enough mates in the entire space, choosing a neighbour to reproduce instead, with similar results.)

Each individual is a seeker once in a generation, but there is a probability, Q, that it will not reproduce at all. When that happened, another individual is randomly selected from its spatial neighbourhood to reproduce in its place, with the offspring placed in the location of the original seeker. This maintains an approximately homogeneous demographic density, corresponding to a fixed local carrying capacity. Each reproduction occurs with one genetic crossover (the genome up to a certain point is from one parent and after that is from the other parent), reflecting genetic linkage. Mutation occurs with a probability μ. Offspring dispersal occurs with probability D; the offspring is placed at one of the 20 neighbouring sites, chosen randomly. For the results reported, $B = 125$, $\mu = 0.001$, $Q = 30\%$, $P = 8$ and $D = 0.01$. Parameter variation affected details, but not the overall behaviour. For example, robustness can be demonstrated by setting $Q = 0\%$, $D = 0$ and $P = 1$. The effects of varying S and G are described in the text. Results for the number of species and distributions are obtained by time-averaging over 10,000 generations, sampling every 25 generations after reaching a steady state.

The results reported here are for asynchronous mating, although many variations of the model were implemented, with no essential changes in the behaviour. Synchronous mating gave similar results, although speciation times were typically twice as long. We also considered models that indirectly impose the genetic mating distance through selection on genetic differences that affect offspring viability and genetic differences regulating behavioural traits that promote reproductive isolation, models with multiple crossovers, models with adaptive changes in spatial and genetic mating distances when few potential mates were available, and models with populations of two sexes. We also obtained speciation for significantly lower mutation rates. For $\mu = 0.00001$, speciation occurred for $S = 3$, $G = 7$ and $N = 2{,}500$ in a 128×128 lattice after 7,000 generations. Robustness suggests that improving the realism of the model by including other assumptions (details of mating and recombination, diploidy and dioecism) will not affect the basic results.

LETTERS

Evolution of a malaria resistance gene in wild primates

Jenny Tung[1,2], Alexander Primus[1,2], Andrew J. Bouley[1], Tonya F. Severson[2], Susan C. Alberts[1,2,3,4]* & Gregory A. Wray[1,2,3]*

The ecology, behaviour and genetics of our closest living relatives, the nonhuman primates, should help us to understand the evolution of our own lineage. Although a large amount of data has been amassed on primate ecology and behaviour, much less is known about the functional and evolutionary genetic aspects of primate biology, especially in wild primates. As a result, even in well-studied populations in which nongenetic factors that influence adaptively important characteristics have been identified, we have almost no understanding of the underlying genetic basis for such traits. Here, we report on the functional consequences of genetic variation at the malaria-related *FY* (*DARC*) gene in a well-studied population of yellow baboons (*Papio cynocephalus*) living in Amboseli National Park in Kenya. *FY* codes for a chemokine receptor normally expressed on the erythrocyte surface that is the known entry point for the malarial parasite *Plasmodium vivax*[1–3]. We identified variation in the *cis*-regulatory region of the baboon *FY* gene that was associated with phenotypic variation in susceptibility to *Hepatocystis*, a malaria-like pathogen that is common in baboons[4,5]. Genetic variation in this region also influenced gene expression *in vivo* in wild individuals, a result we confirmed using *in vitro* reporter gene assays. The patterns of genetic variation in and around this locus were also suggestive of non-neutral evolution, raising the possibility that the evolution of the *FY cis*-regulatory region in baboons has exhibited both mechanistic and selective parallels with the homologous region in humans[6–8]. Together, our results represent the first reported association and functional characterization linking genetic variation and a complex trait in a natural population of nonhuman primates.

In humans, a transition from the wild-type T variant to a C variant at a single polymorphic site in the *FY cis*-regulatory region causally abolishes all expression of this gene in erythrocytic precursors. As a result, C homozygotes at this site are strongly protected from infection by *P. vivax*[9], and a lower level of protection is also conferred on C/T heterozygotes[10,11]. The C variant has apparently arisen independently at least twice in geographically distinct human populations (in Africa and in Papua New Guinea[2,11]), and has been driven to high frequencies on at least two haplotypic backgrounds within Africa[6]. Additionally, the pattern of variation in the *cis*-regulatory region as a whole strongly indicates a historical pattern of natural selection in different populations around the world, probably as the product of directional selection in some populations (for example, local positive selection), and a complex mix of selection and demographic history in others[6–8]. The unusual evolutionary history of this locus led us to investigate the pattern of genetic variation in its baboon homologue, and to explore the possibility that it might also explain phenotypic variation in parasite infection in a wild primate population, the well-studied baboon population of the Amboseli basin in East Africa[12–14].

Baboons are not generally infected by *Plasmodium* in the wild, but are vulnerable to infection by several closely related haematoprotozoans[4,5] including *Hepatocystis kochi*, a blood parasite nested within the paraphyletic *Plasmodium* genus[15]. *Hepatocystis* parasites do not produce the cyclical fever spikes typical of malaria in humans, but do produce anaemia and visible merocyst formation, followed by scarring on the liver[4].

We tested for the presence of *Hepatocystis* parasites by screening DNA samples extracted from baboon blood for 190 individuals in the Amboseli baboon population. We found a high incidence of *Hepatocystis* infection in the Amboseli population (61.9%), although rates of infection varied substantially between different social groups and over time, possibly because of differences in home range and hence exposure to the vector, a biting midge[16].

In 174 of 190 baboons that we screened for *Hepatocystis*, we also sequenced the region of baboon DNA homologous to the annotated human *FY cis*-regulatory region. We identified six single nucleotide polymorphisms (SNPs) in the baboon *FY cis*-regulatory region (Fig. 1 and Supplementary Fig. 1; the malaria-associated SNP documented in humans was invariant in the baboons). *Hepatocystis* infection was significantly associated with an A/G variable site in the *FY cis*-regulatory region, in a model that took social group (a significant source of variance in infection) and genetic background into account. The risk of infection decreased as the number of G alleles an individual carried increased ($P < 0.012$, $n = 174$; Fig. 2).

We also investigated whether *FY cis*-regulatory variation in baboons causally influences gene expression, as the C/T variant does in humans. We collected 101 samples of RNA-preserved blood from adults in six baboon social groups between 2004 and 2008, and used

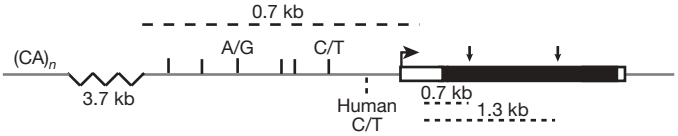


Figure 1 | Schematic of the baboon *FY* gene (not to scale). Boxed regions are regions of the gene present in mature messenger RNA (open boxes are untranslated regions; black boxes are protein coding sequence). Grey lines indicate untranscribed regions. The bent arrow indicates the start of transcription. The downward arrows show two baboon SNPs used as markers for the pyrosequencing assays. The vertical black bars mark *cis*-regulatory SNPs in baboons, with the *Hepatocystis*-associated SNP labelled A/G and the allelic-imbalance-associated SNP labelled C/T. The dashed vertical bar marks the location of the functional SNP known in humans. The dashed horizontal lines provide relative distances and/or sizes of features. The location of the $(CA)_n$ microsatellite used in the F_{st} analyses is also shown upstream.

[1]Department of Biology, Duke University, [2]Institute for Genome Sciences and Policy, Duke University, [3]Department of Evolutionary Anthropology, Duke University, Durham, North Carolina 27708, USA. [4]Institute of Primate Research, National Museums of Kenya, Nairobi, Kenya.
*These authors contributed equally to this work.

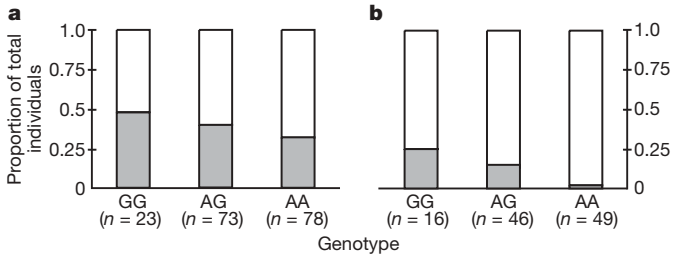


Figure 2 | Genotype at the *FY cis*-regulatory A/G SNP is associated with *Hepatocystis* infection. The proportion of uninfected individuals is shown in grey, and the proportion of infected individuals is shown in white. **a**, Results for the entire sample set ($n = 174$; $P < 0.012$); **b**, results only for members of the six groups with high prevalence (>75%) of *Hepatocystis* infection ($n = 111$; $P < 0.004$). Numbers below each genotype show the number of individuals for the given genotype.

these samples to measure allele-specific expression at the *FY* locus using pyrosequencing. Specifically, we investigated whether the level of *FY* expression driven by one *cis*-regulatory *FY* allele differed from the level of *FY* expression driven by the other *cis*-regulatory *FY* allele, within the same individual. Because allele-specific expression compares the relative amounts of gene expression within individuals, it controls for effects on gene expression operating in *trans*, such as those produced by genetic background or by environmental main effects[17,18]. If the two alleles within an individual drive expression differently (allelic imbalance), that individual is likely to harbour a functional *cis*-regulatory variant that influences gene expression.

We measured allele-specific expression in 38 individuals (all the individuals among the 101 RNA-sampled baboons that were heterozygous at a transcribed pyrosequencing assay SNP: see Fig. 1). Average $\log_2$ fold-change differences in expression between alleles within heterozygous individuals ranged from −0.002 (no difference between alleles) to 2.13 (substantial difference between alleles). This suggested that one or more common functional *cis*-regulatory variants influenced expression of the baboon *FY* gene in the Amboseli population. We predicted that if a *cis*-regulatory variant contributes to variation in gene expression, then individuals heterozygous at the *cis*-regulatory site would show significantly higher levels of allelic imbalance than individuals homozygous for the same variant.

Genotypes at four of the six SNPs in the baboon *FY cis*-regulatory region were sufficiently variable to test for an association with allelic imbalance. Genotype at the SNP closest to the start of transcription, a C/T transition, was significantly associated with allelic imbalance in the predicted direction: heterozygotes exhibited higher levels of allelic imbalance than homozygotes ($P < 0.002$, $n = 38$; Fig. 3). However, this site explained only 22.0% of the overall variance in the allelic imbalance samples, after taking into account year of sampling. These results suggested that this C/T SNP functionally influences gene expression of the *FY* gene within the baboon population, but that additional *cis*-regulatory variants probably also play a part.

We next investigated the A/G SNP in the *FY cis*-regulatory region that we had associated with *Hepatocystis* infection risk. Of the 38 individuals for whom we measured allele-specific expression, 37 were heterozygous for this A/G SNP. Hence, we could not compare allelic imbalance levels between heterozygotes and homozygotes at this site. We therefore tested this variant for possible functional effects using an *in vitro* approach in cell culture. We also used this framework to test further whether the C/T variant that was associated with allelic imbalance causally influenced *FY* expression. For each of these two regulatory SNPs (A/G and C/T), we built two plasmid constructs consisting of the *FY cis*-regulatory region linked to the firefly *luciferase* gene, such that the two constructs differed only at the variable site. We then tested the ability of these constructs to drive gene expression in a human erythroleukaemic (HEL) cell line. For the C/T SNP, the T allele construct drove significantly higher levels of expression than

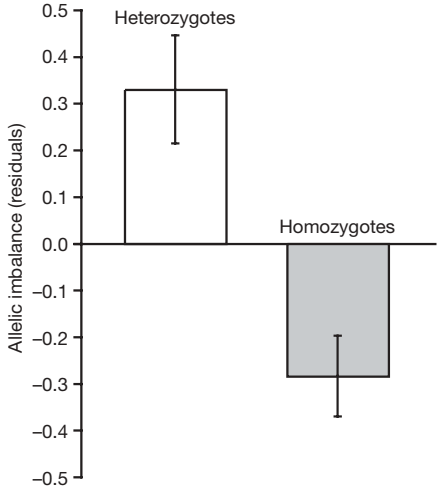


Figure 3 | Allelic imbalance is associated with *FY cis*-regulatory genotype. Height of the bars shows the mean for each genotypic class (heterozygotes: 0.329 ± 0.116 s.e.m.; homozygotes: −0.285 ± 0.086 s.e.m.). The *y* axis gives the residuals of $\log_2$-transformed allelic imbalance on year of sampling ($n = 38$). Heterozygotes at the C/T SNP exhibit high values of allelic imbalance relative to homozygotes at the C/T SNP after controlling for the effect of year of sampling. High values indicate that the two alleles of the *FY* gene are transcribed at different levels within individuals, and suggest a functional *cis*-regulatory role for this SNP.

the alternative C allele construct ($P < 0.0001$; Fig. 4a). Similarly, for the A/G SNP, the G allele construct drove significantly higher levels of expression than the alternative A allele construct ($P < 0.0001$; Fig. 4b). These results suggest that both SNPs have the capacity to drive differential expression of the *FY* gene. In the case of the C/T SNP, for which both *in vivo* and *in vitro* analyses were possible, the T allele was associated with higher levels of expression in both experiments. Unlike the human case, in which one regulatory SNP results in null expression, all the baboon haplotypes we tested drove robust expression of the gene.

Together, these data indicate that the *FY cis*-regulatory region is associated with parasite infection in a wild population of baboons, and that functional sequence variants within this region causally influence the level of expression of the *FY* gene. As in humans, variation in gene expression at the *FY* locus may therefore be important in parasite susceptibility, either through altering the direct access of *Hepatocystis* to baboon erythrocytes or, as has recently been demonstrated in humans, by altering a more general property of the immune system, such as relative white blood cell counts[19].

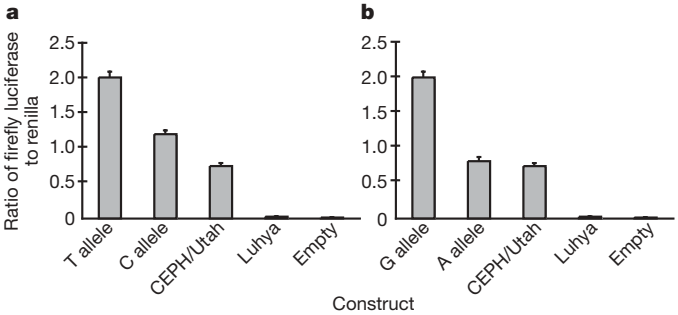


Figure 4 | *FY cis*-regulatory variation drives differential expression *in vitro*. **a**, The C/T SNP identified through the allelic imbalance measurements. **b**, The A/G SNP that associates with *Hepatocystis* infection drives differential gene expression compared to the alternative allele of the same SNP in cell culture. Values on the *y* axis give the relative ratio of firefly luciferase luminescence to a control renilla luciferase reporter. Human constructs from a normal-expressing individual (haplotype from the CEPH/Utah HapMap panel) and a null-expressing individual (Luhya) and an empty vector are shown for comparison. Error bars show s.e.m. for each construct.

These results suggest that the genetic basis of phenotypic variation in different primate species can exhibit a remarkable degree of parallelism. In this case, not only are these similarities present on the molecular level or on the level of trait association, as shown by previous work[20,21], but they also extend to the mechanism that links molecular and phenotypic variation (which is probably gene expression). Given these parallelisms, baboons, like humans, may also exhibit evidence of non-neutral evolution at the *FY cis*-regulatory region. We detected an increased level of population differentiation among East African baboon populations around *FY*, by comparing a *FY*-linked microsatellite with 35 neutral microsatellites ($F_{st} = 0.31$, $P < 0.029$; range of F_{st} for the neutral markers was 0.008–0.346; F_{st} is a metric describing genetic divergence between populations based on allele frequency differences at variable sites; Supplementary Fig. 3). We also detected a higher value for the Tajima's *D* statistic ($D = 1.26$) in this region relative to nine of nine other resequenced putative *cis*-regulatory regions in the Amboseli population and 11 of 12 resequenced transcribed regions (range of *D* for all other loci was −1.60 to 2.12). The only locus with a higher value of *D*, a transcribed portion of the gene *MSR1*, exhibited an even more extreme value than that identified for the *MHC DQA1* promoter in baboons[22], which is known to evolve under strong *trans*-specific balancing selection[20].

Interestingly, a sliding window analysis showed that the peak values of *D* corresponded well with the *Hepatocystis* and allelic-imbalance-associated SNPs (Supplementary Fig. 3). Given that rates of *Hepatocystis* infection appear to vary across different populations (30% in the Masai Mara Reserve, Kenya, $n = 10$; 90% in Mikumi National Park, Tanzania, $n = 20$; see Supplementary Tables 1 and 2), these results suggest that the baboon *FY* cis-regulatory region may be subject to a complex selective history similar to the case described in humans[7,23] in which differing levels of pathogen pressure across populations are associated with high levels of population differentiation around the *FY* gene, and varying signatures of selection within populations.

In spite of the parallels that we have documented for baboon and human *FY*, the functional variants we have identified in baboons are not homologous to the known functional variant in humans, which reveals that phenotypic variation in different primate species may show similar, but not precisely convergent, patterns of evolution. Indeed, while in humans the *FY*–malaria relationship is Mendelian, both *FY* expression and infection by *Hepatocystis* in baboons are clearly complex traits: even individuals homozygous for the *Hepatocystis* 'resistance' variant (the G allele at the A/G SNP) suffer from parasitism, albeit at a lower rate (52.2% of GG homozygotes were infected, versus 67.9% of AA homozygotes, across all study groups: see Fig. 2 and Supplementary Fig. 2). Additionally, the *in vitro* cell culture experiments suggest that the G allele of this variant actually drives higher expression of *FY* than the alternative A allele, even though the G allele is associated with a lower risk of *Hepatocystis* infection. The relationship between *FY* gene expression and *Hepatocystis* in baboons is therefore clearly different from that in humans, perhaps owing to balancing the cost of infection by other blood parasites, some of which are not known to co-occur with, and might be excluded by, *Hepatocystis*[24]. Alternatively, while the *in vitro* data on the A/G variant strongly suggest that this site has the capacity to influence *FY* gene expression, the direction and magnitude of its effects may differ in its natural cellular context.

This possibility is supported by the differences in magnitude of the effect of the C/T *cis*-regulatory variant in the *in vitro* transfection assays and the *in vivo* allelic imbalance measurements. *In vivo* gene expression measurements are complicated by variation in genetic background and in the environment, both of which can modify functional *cis*-regulatory effects[25,26]. Indeed, our results show that even baboons that are homozygotes at the C/T site sometimes exhibit allelic imbalance in *FY* expression, suggesting that other, unidentified functional *cis*-regulatory variants are also segregating in the population. In contrast, in the *in vitro* comparisons, only a single *cis*-regulatory site differed between the experimental constructs, thus controlling for both environment and genetic backgrounds. Using both approaches in tandem can be synergistic: while *in vitro* experiments can help pin down specific functional sites, *in vivo* results demonstrate that these effects are relevant to the biology of individuals in the wild.

Thus, although identifying the genetic basis for phenotypic variation in wild primates poses substantial challenges, we present this study as a model to motivate additional evolutionary genetic research on natural primate populations. This work is essential if we hope to integrate an evolutionary and functional genetic perspective into the rich tradition of organismal research on these species. Our results demonstrate that patterns of variation in nonhuman primates can provide unique insights into the influence of ecological and environmental factors on genetic and trait variation in humans. Integrative research on nonhuman primates should also help us develop a better understanding of the evolution of our own species.

METHODS SUMMARY

We screened for *Hepatocystis* in 190 baboons, using *Hepatocystis* mitochondrial DNA specific primers. We sequenced the region homologous to the human *FY cis*-regulatory region in 174 of these baboons. We identified predictors of *Hepatocystis* infection in these 174 individuals using a generalized linear mixed model with a binomial error structure, fitting genotype and an estimate of genetic background as fixed effects and study group as a random effect.

Allelic imbalance was assessed via pyrosequencing using RNA collected from the study population, and was measured as the $\log_2$ transformed ratio of the expression of one allele versus the other allele, within individuals. We modelled allelic imbalance using a general linear mixed model and evaluated the significance of the genotype effect on allelic imbalance using a permutation test.

We co-transfected human erythroleukaemic cells with experimental constructs or a control construct and a normalization construct. A total of 24 measurements were made per construct in three independent experiments. We compared the measurements for each pair of constructs separately using general linear mixed models.

For the F_{st} comparison, we genotyped up to 36 polymorphic microsatellite loci (35 neutral markers plus one located near the *FY cis*-regulatory region[23]) in ten baboons from the Masai Mara Reserve, Kenya, 20 baboons from Mikumi National Park, Tanzania, and 12 baboons from the Amboseli basin, Kenya (Supplementary Information). We compared the F_{st} value for the *FY*-linked microsatellite locus with those for the 35 neutral loci. The probability of observing a more extreme level of genetic differentiation at the *FY*-linked locus was assessed by fitting a gamma distribution to the F_{st} values for the neutral loci, and calculating the average value of *P* based on different parameters of the gamma (weighted by the probability of each parameter set, given the data). Tajima's *D* values were calculated for samples from Amboseli using DnaSP version 4.9 (ref. 27).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 March; accepted 15 May 2009.
Published online 24 June 2009.

1. Barnwell, J. W., Nichols, M. E. & Rubinstein, P. *In vitro* evaluation of the role of the Duffy blood group in erythrocyte invasion by *Plasmodium vivax*. *J. Exp. Med.* **169,** 1795–1802 (1989).
2. Miller, L. H., Mason, S. J., Clyde, D. F. & Mcginniss, M. H. Resistance factor to *Plasmodium vivax* in blacks—Duffy blood group genotype *Fyfy*. *New Engl. J. Med.* **295,** 302–304 (1976).
3. Miller, L. H., Mason, S. J., Dvorak, J. A., Mcginniss, M. H. & Rothman, I. K. Erythrocyte receptors for *Plasmodium knowlesi* malaria—Duffy blood group determinants. *Science* **189,** 561–563 (1975).
4. Garnham, P. C. C. *Malaria Parasites and Other Haemosporidia* (Blackwell, 1966).
5. Myers, B. J. & Kuntz, R. E. A checklist of parasites reported for the baboon. *Primates* **6,** 137–194 (1965).
6. Hamblin, M. T. & Di Rienzo, A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet.* **66,** 1669–1679 (2000).
7. Hamblin, M. T., Thompson, E. E. & Di Rienzo, A. Complex signatures of natural selection at the Duffy blood group locus. *Am. J. Hum. Genet.* **70,** 369–383 (2002).
8. Sabeti, P. C. *et al.* Positive natural selection in the human lineage. *Science* **312,** 1614–1620 (2006).
9. Tournamille, C., Colin, Y., Cartron, J. P. & Levankim, C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy negative individuals. *Nature Genet.* **10,** 224–228 (1995).

10. Michon, P. *et al.* Duffy-null promoter heterozygosity reduces DARC expression and abrogates adhesion of the *P.vivax* ligand required for blood-stage infection. *FEBS Lett.* **495,** 111–114 (2001).
11. Zimmerman, P. A. *et al.* Emergence of FY*A(null) in a *Plasmodium vivax*-endemic region of Papua New Guinea. *Proc. Natl Acad. Sci. USA* **96,** 13973–13977 (1999).
12. Alberts, S. C., Buchan, J. C. & Altmann, J. Sexual selection in wild baboons: from mating opportunities to paternity success. *Anim. Behav.* **72,** 1177–1196 (2006).
13. Buchan, J. C., Alberts, S. C., Silk, J. B. & Altmann, J. True paternal care in a multi-male primate society. *Nature* **425,** 179–181 (2003).
14. Tung, J., Charpentier, M. J. E., Garfield, D. A., Altmann, J. & Alberts, S. C. Genetic evidence reveals temporal change in hybridization patterns in a wild baboon population. *Mol. Ecol.* **17,** 1998–2011 (2008).
15. Perkins, S. L. & Schall, J. J. A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. *J. Parasitol.* **88,** 972–978 (2002).
16. Garnham, P. C. C., Heisch, R. B., Minter, D. M., Phipps, J. D. & Ikata, M. *Culicoides adersi* Ingram and Macfie, 1923, a presumed vector of *Hepatocystis* (= *Plasmodium*) *kochi* (Laveran, 1899). *Nature* **190,** 739–741 (1961).
17. Wittkopp, P., Haerum, B. & Clark, A. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430,** 85–88 (2004).
18. Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B. & Kinzler, K. W. Allelic variation in human gene expression. *Science* **297,** 1143 (2002).
19. Reich, D. *et al.* Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genet.* **5,** e1000360 (2009).
20. Loisel, D. A., Rockman, M. V., Wray, G. A., Altmann, J. & Alberts, S. C. Ancient polymorphism and functional variation in the primate MHC-DQA1 5′ *cis*-regulatory region. *Proc. Natl Acad. Sci. USA* **103,** 16331–16336 (2006).
21. Wooding, S. *et al.* Independent evolution of bitter-taste sensitivity in humans and chimpanzees. *Nature* **440,** 930–934 (2006).
22. Loisel, D. *Evolutionary Genetics of Immune System Genes in a Wild Primate Population*. PhD thesis, Duke Univ. (2007).
23. Seixas, S., Ferrand, N. & Rocha, J. Microsatellite variation and evolution of the human Duffy blood group polymorphism. *Mol. Biol. Evol.* **19,** 1802–1806 (2002).
24. Moore, J. A. & Kuntz, R. E. *Entopolyploides macaci* Mayer, 1934 in the African baboon (*Papio cynocephalus* L. 1766). *J. Med. Primatol.* **4,** 1–7 (1975).
25. Brem, R. B., Storey, J. D., Whittle, J. & Kruglyak, L. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436,** 701–703 (2005).
26. Smith, E. N. & Kruglyak, L. Gene-environment interaction in yeast gene expression. *PLoS Biol.* **6,** e83 (2008).
27. Rozas, J., Sanchez-Del Barrio, J. C., Messeguer, X., Rozas, R. & Dna, S. P. DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19,** 2496–2497 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Office of the President of the Republic of Kenya and the Kenya Wildlife Service for permission to work in Amboseli National Park, and the Institute of Primate Research for local sponsorship. We thank the wardens and staff of Amboseli National Park, and the pastoral communities of Amboseli and Longido for cooperation. We thank J. Altmann for providing access to long-term data and contributing samples, and J. Altmann and Y. Gilad for providing comments on the manuscript. R. S. Mututua, S. Sayialel, and J. K. Warutere assisted with sample collection. The Integrated Primate Biomaterials and Information Resource, the Coriell Institute, J. Rogers and R. Sapolsky provided access to DNA samples from Mikumi and Masai Mara respectively. G. Gibson, T. F. C. Mackay, L. Goering and D. Tan provided access to a pyrosequencer at NC State University. M. Akinyi assisted with sample collection and analysis. A. D. Pfefferle assisted with sequencing. S. Mukherjee advised and assisted with statistical tests. Financial support came from the National Science Foundation (to S.C.A. and J.T.); the American Society of Primatologists (to J.T.); Duke University and the Duke chapter of Sigma Xi (to J.T.); and the Duke Institute for Genome Sciences and Policy (to G.A.W.).

Author Contributions J.T., S.C.A. and G.A.W. designed the study, analysed the results, and wrote the paper. J.T. and S.C.A. collected blood samples; S.C.A. provided the long-term data on Amboseli National Park. J.T. gathered the allelic imbalance and sequence data; A.P., T.F.S. and J.T. collected the transfection assay data; T.F.S. and J.T. collected the genotyping data; A.J.B. and J.T. collected the *Hepatocystis* data. G.A.W. and S.C.A. provided funding support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Sequence data have been deposited in NCBI GenBank under the accession numbers FJ952954–FJ955880, FJ955882–FJ955885, FJ955887–FJ955896 and FJ955899–FJ956699. Correspondence and requests for materials should be addressed to J.T. (jt5@duke.edu).

METHODS

DNA and RNA sampling. Blood samples for DNA extraction and *Hepatocystis* screening were collected from 190 Amboseli baboons between 1989 and 2008 (Supplementary Fig. 2)[28]. DNA was extracted using standard methods. DNA for some individuals was whole genome amplified (Qiagen Repli-G Kit). Blood samples for RNA extraction were collected in PaxGene RNA tubes from 101 adult Amboseli baboons darted between 2004 and 2008. RNA was extracted using the PaxGene RNA Blood kit (Qiagen), and reverse transcribed into complementary DNA (ABI High Capacity cDNA Archive Kit).

Sequencing. We amplified and sequenced the region homologous to the annotated *FY cis*-regulatory region in humans in 174 individuals and sequenced or genotyped the two pyrosequencing assay SNPs in 150 individuals. To assess congruence between the *in vitro* and *in vivo* gene expression results for the C/T SNP, we inferred haplotype phasing using PHASE 2.1.1 (ref. 29).

Hepatocystis screen and association with *FY*. We screened for *Hepatocystis* in all 190 baboons using *Hepatocystis* mtDNA specific primers, which produced a band of approximately 251 base pairs on an agarose gel in the presence of *Hepatocystis* (Supplementary Methods).

We then fitted the following generalized linear mixed model for 150 individuals, using a binomial error structure:

$$P(y_{ij} = 1 | G_{ij}) = \text{logit}\left(\beta G_{ij} + \sum_{u=1}^{5} D_{iu} v_u + S_j + b + \varepsilon\right)$$

where the logit link function has been used to change the expression in parentheses into the probability of observing the parasite, given the explanatory variables. y is *Hepatocystis* infection status ($y = 1$ corresponds to infected; $y = 0$ corresponds to uninfected), individuals are indexed by i, and study group is indexed by j. β is a fixed effect of genotype, G_{ij}; v_u is the fixed effect of the projection D_{iu} on the uth principal component of population structure; S_j is a random effect of study group; b is the intercept; and ε represents model error (Supplementary Information). Principal-components analysis of population structure was conducted on data from 47 unlinked loci (33 SNPs and 14 microsatellites, Supplementary Table S4). We also included study group as a random effect because group is a common source of structure in baboons, and infection rates clearly differed between groups. We evaluated the significance of β, the SNP effect, as evidence for association between infection and *cis*-regulatory variation.

Pyrosequencing. We designed pyrosequencing assays based on two variable SNPs in the transcribed region of the *FY* gene. 38 individuals for which cDNA samples were available were also heterozygous at one or both of these sites. For each of these individuals, we performed six to eight pyrosequencing reactions across two plates (mean number of measurements per individual = 7.05, range = 3–8, excluding failed reactions). The resulting values were expressed as the $\log_2$ transformed ratio of the expression of transcripts carrying one versus the other allele at one of the assay SNPs (values based on the alternative assay SNP were converted on the basis of linkage between the two sites).

We identified an effect of one of the upstream *cis*-regulatory sites by modelling variation in allelic imbalance using the following general linear mixed model:

$$y_{ij} = \beta G_{ij} + Y_j + b + \varepsilon$$

where y is allelic imbalance, indexed by individual i and year of sampling j; β is a fixed effect of homozygous or heterozygous genotype, G_{ij}; Y_j is a random effect of year of sampling (2006, 2007 or 2008; one individual was sampled in 2005 and grouped with the 2006 samples); b is the intercept; and ε is the model error. We assessed the significance of β using a permutation test: all measurements for an individual were grouped as a block and permuted 1,000 times over individual identity. We assigned a P value to the original SNP parameter estimate by ranking it among the corresponding estimates for the permuted data sets. The estimate of variance explained by the C/T site is based on modelling the residuals of allelic imbalance on year of sampling, using the C/T site alone.

Three pairs of individuals in the allelic imbalance analysis were related at $r = 0.5$; removing any set of three individuals so that no individuals were closely related did not qualitatively change our results (see also Supplementary Methods).

Transfection assays. Human erythroleukaemic cells (HEL 92.1.7) were maintained using the ATCC protocol. The wells of 24-well cell culture plates were seeded with 2×10^5 cells in 500 µl media, transfected, incubated for 48 h, and lysed. Cells were co-transfected with experimental constructs or empty firefly luciferase vector as control (pGL4.10; 180 ng per well) and the CMV Renilla normalization construct (pGL4.75, 20 ng per well; Promega) using Fugene 6 (Roche). Expression levels were measured with a dual-luciferase reporter assay (DLR1000 assay kit, Promega) and reported as relative ratios of luminescence (firefly:Renilla). Eight replicate wells were transfected for each experimental and control vector within an assay, with the assay repeated three times ($n = 24$ total measurements per construct).

We compared the measurements for each pair of constructs (A versus G for the *Hepatocystis*-associated SNP and C versus T for the SNP associated with allelic imbalance) separately using the following model:

$$y_{ijk} = \beta C_j + E_k + b + \varepsilon$$

where y is the relative ratio of luminescence for replicate i of construct j in experiment k; β is a fixed effect of construct, C_j; E_k is a random effect of experiment; b is the intercept; and ε is model error.

Signature of selection. For comparison of the F_{st} values, we genotyped up to 36 polymorphic microsatellite loci in ten baboons from the Masai Mara Reserve, Kenya; 20 baboons from Mikumi National Park, Tanzania; and 12 baboons from Amboseli National Park (Supplementary Information). One locus is a polymorphic microsatellite located ~3.7 kb upstream of the sequenced *FY cis*-regulatory region[23]; the other 35 loci reside in putatively neutral sites dispersed around the baboon genome[12,13]. We calculated F_{st} values for each of these loci independently using Arlequin 3.1 (ref. 30), and compared the F_{st} value for the *FY*-linked microsatellite locus with the F_{st} values for the 35 neutral loci (Supplementary Figs 3, 4 and Supplementary Methods). For comparison of Tajima's D, we resequenced 21 other regions in and around genes in the same or a subset of the individuals resequenced at the *FY cis*-regulatory region (Supplementary Fig. 3 and Supplementary Table 5). We calculated Tajima's D for all loci using the program DnaSP version 4.9 (ref. 27), assuming no recombination.

28. Altmann, J. *et al.* Behavior predicts genetic structure in a wild primate group. *Proc. Natl Acad. Sci. USA* **93**, 5797–5801 (1996).
29. Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
30. Excoffier, L., Laval, G. & Schneider, S. Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50 (2005).

LETTERS

Rapamycin fed late in life extends lifespan in genetically heterogeneous mice

David E. Harrison[1]*, Randy Strong[2]*, Zelton Dave Sharp[3], James F. Nelson[4], Clinton M. Astle[1], Kevin Flurkey[1], Nancy L. Nadon[5], J. Erby Wilkinson[6], Krystyna Frenkel[7], Christy S. Carter[8]†, Marco Pahor[8]†, Martin A. Javors[9], Elizabeth Fernandez[2] & Richard A. Miller[10]*

Inhibition of the TOR signalling pathway by genetic or pharmacological intervention extends lifespan in invertebrates, including yeast, nematodes and fruitflies[1–5]; however, whether inhibition of mTOR signalling can extend lifespan in a mammalian species was unknown. Here we report that rapamycin, an inhibitor of the mTOR pathway, extends median and maximal lifespan of both male and female mice when fed beginning at 600 days of age. On the basis of age at 90% mortality, rapamycin led to an increase of 14% for females and 9% for males. The effect was seen at three independent test sites in genetically heterogeneous mice, chosen to avoid genotype-specific effects on disease susceptibility. Disease patterns of rapamycin-treated mice did not differ from those of control mice. In a separate study, rapamycin fed to mice beginning at 270 days of age also increased survival in both males and females, based on an interim analysis conducted near the median survival point. Rapamycin may extend lifespan by postponing death from cancer, by retarding mechanisms of ageing, or both. To our knowledge, these are the first results to demonstrate a role for mTOR signalling in the regulation of mammalian lifespan, as well as pharmacological extension of lifespan in both genders. These findings have implications for further development of interventions targeting mTOR for the treatment and prevention of age-related diseases.

Because incidences of most diseases rise rapidly with age[6], interventions that delay ageing would greatly benefit health[7,8]. So far, dietary additives that delay ageing and increase lifespan in rodent models have shown only weak effects[9–11]. Before clinical studies are considered, anti-ageing interventions must be repeatable and effective in many mouse genotypes, and not merely postpone strain-specific diseases[12–14].

The National Institute on Aging Interventions Testing Program (ITP) evaluates agents that may delay ageing and increase lifespan in genetically heterogeneous mice[15–17]. Agents are chosen as summarized at http://www.nia.nih.gov/ResearchInformation/ScientificResources/InterventionsTestingProgram.htm. Studies are simultaneously replicated at three test sites: The Jackson Laboratory (TJL), the University of Michigan (UM), and the University of Texas Health Science Center (UT). BALB/cByJ × C57BL/6J F1 (CB6F1) females and C3H/HeJ × DBA/2J F1 (C3D2F1) males are supplied to each site by The Jackson Laboratory, and mated to produce genetically heterogeneous populations in which each animal is genetically unique, but a full sibling of all other mice in the population[18]. Sufficient mice are used to provide 80% power to detect a 10% increase (or decrease) in mean lifespan with respect to unmanipulated controls of the same sex, even if data from one of the three test sites were to be unavailable. Here we report that dietary encapsulated rapamycin increases mouse survival, including survival to the last decile, a measure of maximal lifespan.

Rapamycin reduces function of the rapamycin target kinase TOR and has anti-neoplastic activities; genetic inhibition of TOR extends lifespan in short-lived model organisms. In male and female mice at each of three collaborating research sites, median and maximum lifespan were extended by feeding encapsulated rapamycin starting at 600 days of age (Fig. 1). We analysed the data set as of 1 February 2009, with 2% (38 of 1,901) of mice still alive. For data pooled across sites, a log-rank test rejected the null hypothesis that treatment and control groups did not differ ($P < 0.0001$); mice fed rapamycin were longer lived than controls ($P < 0.0001$) in both males and females. Expressed as mean lifespan, the effect sizes were 9% for males and 13% for females in the pooled data set. Expressed as life expectancy at 600 days (the age of first exposure to rapamycin), the effect sizes were 28% for males and 38% for females. Mice treated with other agents (enalapril and CAPE (caffeic acid phenethyl ester)) evaluated in parallel did not differ from controls at the doses used (Supplementary Fig. 1).

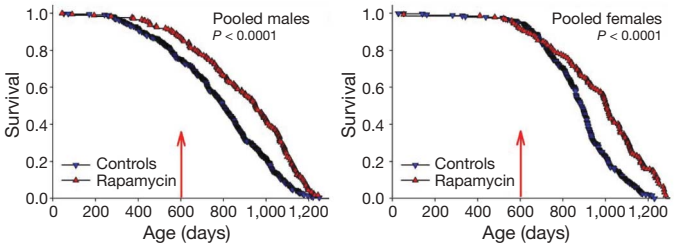


Figure 1 | Survival plots for male and female mice, comparing control mice to those fed rapamycin in the diet starting at 600 days of age, pooling across the three test sites. *P* values were calculated by the log-rank test. Four per cent of the control mice and three per cent of rapamycin-assigned mice were removed from the experiment for technical reasons. Only five animals (three controls, two rapamycin) were removed after the start of rapamycin treatment at 600 days. Thus, there were no significant differences between groups in censoring.

[1]The Jackson Laboratory, Bar Harbor, Maine 04609, USA. [2]Geriatric Research, Education and Clinical Center and Research Service, South Texas Veterans Health Care System, Department of Pharmacology, and Barshop Institute for Longevity and Aging Studies, The University of Texas Health Science Center at San Antonio, Texas 78229, USA. [3]Institute of Biotechnology/Department of Molecular Medicine, and Barshop Institute for Longevity and Aging Studies, The University of Texas Health Science Center at San Antonio, Texas 78245, USA. [4]Department of Physiology and Barshop Institute for Longevity and Aging Studies at The University of Texas Health Science Center at San Antonio, Texas 78229, USA. [5]Division of Aging Biology, National Institute on Aging, Bethesda, Maryland 20892, USA. [6]Unit for Laboratory Animal Medicine, University of Michigan School of Medicine, Ann Arbor, Michigan 48109-2200, USA. [7]Environmental Medicine, NY University School of Medicine, New York 10016, USA. [8]Wake Forest University School of Medicine, Department of Internal Medicine: Section on Gerontology and Geriatrics Winston-Salem, North Carolina 27157, USA. [9]Department of Psychiatry, The University of Texas Health Science Center at San Antonio, Texas 78229, USA. [10]Department of Pathology and Geriatrics Center, University of Michigan, and Ann Arbor VA Medical Center, Ann Arbor, Michigan 48109-2200, USA. †Present address: Department of Aging and Geriatric Research, College of Medicine, Institute on Aging, University of Florida, Gainesville, Florida 32611, USA. *These authors contributed equally to this work.

Rapamycin-fed and control mice were then compared separately for each combination of site and gender. Rapamycin had a consistent benefit, compared with controls, with *P* values ranging from 0.03 to 0.0001 (Fig. 2).

Female mice at all three sites had improved survival after rapamycin feeding (Fig. 2). Mean lifespan increases for females were 15%, 16% and 7% (TJL, UM and UT, respectively), and life expectancy at 600 days increased by 45%, 48% and 22% for females at the three sites. Median lifespan estimates of control females were consistent across sites (881–895 days), and were similar to values noted in Cohort 2004, which ranged from 858 to 909 days[15]. Thus, the improvement in survival seen in the rapamycin-fed females is not an artefact of low survival for the control females.

Male mice at all three sites also had improved survival after rapamycin feeding (Fig. 2). Mean lifespan increases for males were 5%, 8% and 15% (TJL, UM and UT, respectively), and male life expectancy at 600 days increased by 16%, 23% and 52%. Interpretation is complicated by differences among sites in survival of control males, and because mice assigned to the rapamycin-fed group at UT and perhaps at UM had lower mortality before 600 days than controls. Control mice at UT and UM differed from those fed rapamycin not only in exposure to rapamycin from 600 days of age but also in specific formulation of the mouse chows (all based on the NIH-31 standard) used between weaning and 600 days. We thus cannot rule out the possibility that improved survival among males in the rapamycin group, at UT and at UM, might reflect differences in nutritional or health status between control and rapamycin groups before 600 days, rather than solely the effects of rapamycin. Notably, the significant benefits of rapamycin on male (and female) survival at TJL could not have been affected by diet before drug administration, because at TJL both control and rapamycin-fed mice received the same chow (Purina 5LG6) throughout this period.

Maximum lifespan was increased by rapamycin feeding. Table 1 shows the ages at the 90th percentile for control and rapamycin-treated mice, along with the 95% upper confidence bound for the controls. For each site and sex, the 90th percentile age for rapamycin-treated mice is higher than the upper limit for the corresponding control group, showing that rapamycin increases the age for 90th percentile survival.

To determine whether increases in maximal lifespan due to rapamycin feeding are statistically significant, we compared the proportion of living mice in each group after 90% had died in the joint life table[19] (details in Supplementary Table 1). Summing across the three sites, 4.8% of the female control mice were alive at these ages, compared with 21.5% of the rapamycin-treated females ($P < 0.0001$). For males, the corresponding values were 5.9% of controls and 20.2% of rapamycin-treated mice ($P < 0.0001$). The site-specific calculations documented a significant effect on females at both TJL ($P = 0.0006$) and UM ($P = 0.0001$); for males, we noted a significant effect at both TJL ($P = 0.008$) and UT ($P = 0.0001$), with a marginal effect at UM ($P = 0.07$). Rapamycin feeding initiated at 600 days of age thus leads to a significant increase in maximal lifespan.

To test if the spectrum of lesions was altered by dietary rapamycin, complete necropsies were conducted on 31 control and 40 rapamycin-fed mice that were either found dead or killed when moribund (details in Supplementary Table 2). Although rapamycin postpones death, it did not change the distribution of presumptive causes of death.

A separate group of mice was used to evaluate the effects of encapsulated rapamycin initiated at 270 days of age (Fig. 3a). At the time of analysis, 51% of the females and 68% of the males had died, and a stratified log-rank test showed significantly lower mortality risk in the rapamycin-treated mice compared to controls, pooling across the three test sites ($P = 0.0002$ for males and $P < 0.0001$ for females). When each site was evaluated separately, the beneficial effect of rapamycin for females was significant at each site ($P < 0.005$); for males, the effect was significant ($P < 0.025$) at UM and UT, but not at TJL. Rapamycin seems to reduce mid-life mortality risk when started at 270 days of age, but additional data are needed to provide an accurate estimate of effect size, and to evaluate effects on maximal longevity.

To document biochemical effects of rapamycin at the dose used for the lifespan studies, we evaluated the phosphorylation status of ribosomal protein subunit S6 (rpS6)—a target substrate of S6 kinase 1 in the mTOR signalling pathway[20]—in visceral white adipose tissue (a sensitive indicator of mTOR inhibition by rapamycin treatment *in vivo*). Figure 3b shows that rapamycin feeding reduced the levels of phosphorylated rpS6 4–5-fold when fed from 270 to about 800 days of age. Blood levels of rapamycin in the treated mice were equivalent in males and females, between 60 and 70 ng ml^{-1}.

Initial evidence that reduced TOR function can extend longevity came primarily from studies in yeast[1,2] and invertebrates[3–5]. Beneficial effects of diet restriction[21] and dwarf mutations, both of which extend lifespan in rodents, may, to some degree, result from

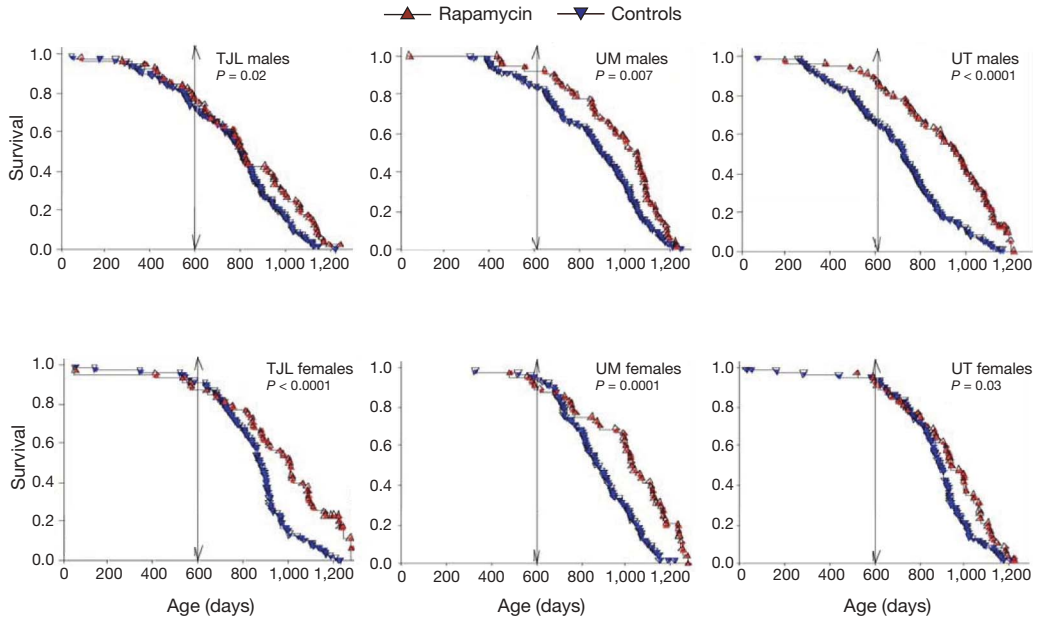


Figure 2 | Survival of control and rapamycin-treated mice for males and females for each of the three test sites separately. *P* values represent results of log-rank calculations. Vertical lines at age 600 days indicate the age at which the mice were first exposed to rapamycin.

Table 1 | The effect of rapamycin on maximum lifespan

Comparison	Sites	Age in days at 90th percentile for controls (upper confidence limit)*	Age in days at 90th percentile for rapamycin-treated mice	Percentage increase
Females				
Rapamycin versus controls	All sites	1,094 (1,136)	1,245	14
Rapamycin versus controls	TJL	1,100 (1,165)	1,282	17
Rapamycin versus controls	UM	1,094 (1,149)	1,250	14
Rapamycin versus controls	UT	1,089 (1,159)	1,179	8
Males				
Rapamycin versus controls	All sites	1,078 (1,111)	1,179	9
Rapamycin versus controls	TJL	1,035 (1,091)	1,142	10
Rapamycin versus controls	UM	1,141 (1,177)	1,188	4
Rapamycin versus controls	UT	1,020 (1,101)	1,179	16

* The upper limit of the 95% confidence interval for control mice is indicated in parentheses. For example, in the top row, for females pooled across sites, the 95% confidence interval for controls goes up to 1,136 days, and the estimate for 90th percentile survival for the rapamycin-treated mice is 1,245 days. This gives good evidence that the 90th percentile survival for rapamycin-treated mice (1,245) is substantially above that for controls (1,094).

repression of the mTOR complex 1 (mTORC1) pathway[22,23]. It is not yet known to what extent inhibition of mTOR will recapitulate other aspects of the phenotypes associated with diet restriction or dwarf mutations. Our demonstration that rapamycin feeding increases lifespan even when started late in life, as well as the absence of changes in body weight (data not shown), distinguishes our results from studies using diet restriction: in all cases diet restriction reduces body weight, and in most reports[21], although not all[24], diet restriction produces little, if any, benefit if started after about 550 days of age.

Rapamycin may extend lifespan in old, genetically heterogeneous mice through a combination of anti-neoplastic effects[25,26] and effects on cellular stress resistance and response to nutrient dynamics[27–29]. The increase in both median and maximum lifespan seen in rapamycin-fed mice is consistent with the hypothesis that inhibiting the mTORC1 pathway retards mammalian ageing, but is not compelling proof that ageing rates are altered; this would require testing whether the intervention decelerates age-dependent changes in multiple organs, cell types and intracellular and extracellular processes[14]. Comparing the effects of rapamycin treatment and other models of decelerated ageing will help narrow the list of possible mechanisms for longevity extension.

At the cellular level, mTORC1 helps to coordinate growth and survival responses induced by alterations in nutrient availability, energy status, growth factor stimuli and exposure to potentially lethal cell stresses[27–29]; this strategic position at the nexus of nutrient/stress sensing pathways may contribute to the importance of TOR function in regulating lifespan in invertebrates and in mammals as well. It is especially noteworthy that rapamycin feeding can extend mouse lifespan even when started late in life; in terms of the percentage of the maximal lifespan, a 600-day-old mouse is roughly the equivalent of a 60-year-old person[14]. An effective anti-ageing intervention that could be initiated later than the midpoint of the lifespan could prove to be

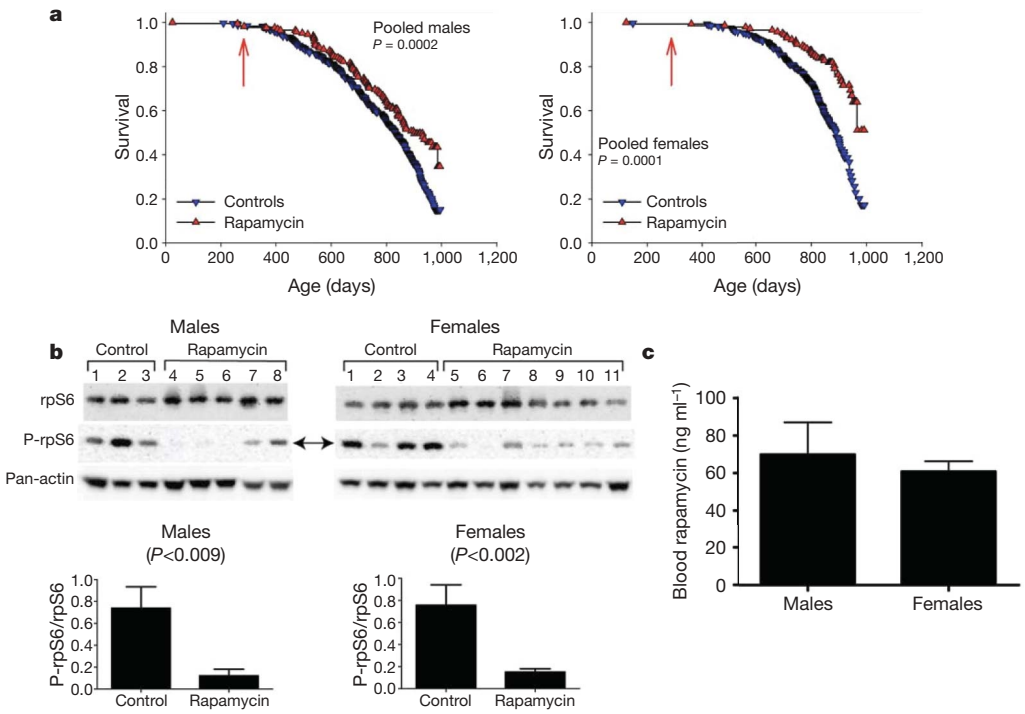


Figure 3 | Characterization of mice receiving rapamycin from 270 days of age. a, Survival plots for male and female mice, comparing control mice to rapamycin-treated mice of a separate (Cohort 2006) population, in which mice were treated with rapamycin from 270 days of age. Because at the time of the interim analysis all live mice were between 800 and 995 days of age, we have only limited information about the shape of the survival curve at ages above 900 days, and the apparent change in slope at the oldest ages (>990 days) reflects this experimental uncertainty. *P* values were calculated by the log-rank test. **b**, Effects of dietary rapamycin on an mTORC1 effector in the visceral fat pads from 750-day-old to 880-day-old male and female mice. Ribosomal subunit protein S6 (rpS6) and its phosphorylation status (P-rpS6, double arrow) were immunoassayed in tissue lysates prepared from mice consuming microencapsulated rapamycin-containing or control diets. Antibodies used are shown to the left. The ratio of intensity values for P-rpS6/rpS6 is shown in the graphs for female and male mice. Pan-actin was also immunoassayed in the blots to provide an indication of protein loading for each lane. **c**, Whole blood rapamycin content in 750-day-old to 880-day-old male and female mice. In **b** and **c**, error bars show standard errors of the mean.

especially relevant to clinical situations, in which the efficacy of anti-ageing interventions would be particularly difficult to test in younger volunteers. Our data justify special attention to the role of the TOR pathway in control of ageing in mammals and in the pathogenesis of late-life illnesses.

METHODS SUMMARY

Mice. Specific pathogen-free (SPF) mice were produced at each of the three test sites by mating CB6F1 females with C3D2F1 males to produce a genetically heterogeneous population[15]. Weanlings at the three sites were fed similar diets (but not identical; see Supplementary Information for details) until they were started on food containing rapamycin, in Purina 5LG6, at 600 days of age. A second independent study was begun 1 year later, with rapamycin initiated at 270 days rather than at 600 days of age. The principal endpoint was age at death (for mice found dead at daily inspections) or age at euthanasia (for mice deemed unlikely to survive for more than an additional 48 h).

Diet preparation. Rapamycin (from LC Labs) was microencapsulated by Southwest Research Institute (San Antonio, Texas), using a spinning disk atomization coating process with the enteric coating material Eudragit S100 (Röhm Pharma). This coating increased the fraction of rapamycin that survived the food preparation process by three- to fourfold, and protected the agent from digestion in the stomach[16]. Encapsulated rapamycin was then incorporated into 5LG6 mouse chow and distributed to all three test sites. Rapamycin in blood and diet was measured by high-performance liquid chromatography (HPLC) with ultraviolet detection as detailed in Supplementary Information.

Rapamycin effectiveness. To test effects of rapamycin on mTORC1 targets, we measured phosphorylation of ribosomal protein S6 (at Ser 240 and Ser 244), a substrate of S6 kinase 1, in visceral adipose tissue. Adipose tissue was dissected from mice that had been fed a rapamycin diet for 420 days; lysates were loaded on a 4–12% gradient PAGE and electrophoresed overnight at 5V. After transfer to membranes, blocking and incubation with appropriate antibodies, total amounts of rpS6 and amounts of phosphorylated rpS6 were quantified by chemiluminescence. Techniques are detailed in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 April; accepted 24 June 2009.
Published online 8 July; corrected 16 July 2009 (see full-text HTML version for details).

1. Kaeberlein, M. *et al.* Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. *Science* **310,** 1193–1196 (2005).
2. Powers, R. W. III, Kaeberlein, M., Caldwell, S. D., Kennedy, B. K. & Fields, S. Extension of chronological life span in yeast by decreased TOR pathway signaling. *Genes Dev.* **20,** 174–184 (2006).
3. Jia, K., Chen, D. & Riddle, D. L. The TOR pathway interacts with the insulin signaling pathway to regulate *C. elegans* larval development, metabolism and life span. *Development* **131,** 3897–3906 (2004).
4. Kapahi, P. *et al.* Regulation of lifespan in *Drosophila* by modulation of genes in the TOR signaling pathway. *Curr. Biol.* **14,** 885–890 (2004).
5. Vellai, T. *et al.* Genetics: influence of TOR kinase on lifespan in *C. elegans*. *Nature* **426,** 620 (2003).
6. Kohn, R. R. *Principles of Mammalian Aging* 2nd edn 151 (Prentice-Hall, 1978).
7. Miller, R. A. Extending life: scientific prospects and political obstacles. *Milbank Q.* **80,** 155–174 (2002).
8. Olshansky, S. J., Perry, D., Miller, R. A. & Butler, R. N. In pursuit of the longevity dividend. *Scientist* **20,** 28–35 (2006).
9. Schneider, E. L. & Miller, R. A. in *Brockelhurst's Textbook of Geriatric Medicine* (eds Tallis, R., Fillit, H. & Brockelhurst, J. C.) 193–199 (Churchill Livingstone, 1998).
10. Archer, J. R. & Harrison, D. E. L-deprenyl treatment in aged mice slightly increases lifespans, and greatly reduces fecundity by aged males. *J Gerontol. Biol. Sci.* **51A,** B448–B453 (1996).
11. Schneider, E. L. & Reed, J. D. Jr. Life extension. *N. Engl. J. Med.* **312,** 1159–1168 (1985).
12. Phelan, J. P. & Austad, S. N. Selecting animal models of human aging. Inbred strains often exhibit less biological uniformity than F1 hybrids. *J. Gerontol.* **49,** B1–B11 (1994).
13. Klebanov, S. E. *et al.* Maximum life spans in mice are extended by wild strain alleles. *Exp. Biol. Med.* **226,** 854–859 (2001).
14. Flurkey, K., Currer, J. M. & Harrison, D. E. in *The Mouse in Biomedical Research* 2nd edn, Vol. III (eds Fox, J. G. *et al.*) 637–672 (Academic, 2007).
15. Miller, R. A. *et al.* An aging interventions testing program: study design and interim report. *Aging Cell* **6,** 565–575 (2007).
16. Nadon, N. L. *et al.* Design of aging intervention studies: the NIA interventions testing program. *AGE* **30,** 187–199 (2008).
17. Strong, R. *et al.* Nordihydroguaiaretic acid and aspirin increase lifespan of genetically heterogeneous male mice. *Aging Cell* **7,** 641–650 (2008).
18. Roderick, T. H. Selection for radiation resistance in mice. *Genetics* **48,** 205–216 (1963).
19. Wang, C., Li, Q., Redden, D. T., Weindruch, R. D. & Allison, B. Statistical methods for testing effects on "maximum lifespan". *Mech. Ageing Dev.* **125,** 629–632 (2004).
20. Petroulakis, E., Mamane, Y., Le Bacquer, O., Shahbazian, D. & Sonenberg, N. mTOR signaling: implications for cancer and anticancer therapy. *Br. J. Cancer* **96** (Suppl.), R11–R15 (2007).
21. Masoro, E. J. Overview of caloric restriction and ageing. *Mech. Ageing Dev.* **126,** 913–922 (2005).
22. Sharp, Z. D. & Bartke, A. Evidence for down-regulation of phosphoinositide 3-kinase/Akt/mammalian target of rapamycin (PI3K/Akt/mTOR)-dependent translation regulatory signaling pathways in Ames dwarf mice. *J. Gerontol. A* **60,** 293–300 (2005).
23. Hsieh, C. C. & Papaconstantinou, J. Akt/PKB and p38 MAPK signaling, translational initiation and longevity in Snell dwarf mouse livers. *Mech. Ageing Dev.* **125,** 785–798 (2004).
24. Dhahbi, J. M. *et al.* Temporal linkage between the phenotypic and genomic responses to caloric restriction. *Proc. Natl Acad. Sci. USA* **101,** 5524–5529 (2004).
25. Garber, K. Rapamycin's resurrection: a new way to target the cancer cell cycle. *J. Natl Cancer Inst.* **93,** 1517–1519 (2001).
26. Lorberg, A. & Hall, M. N. TOR: the first 10 years. *Curr. Top. Microbiol. Immunol.* **279,** 1–18 (2004).
27. Wullschleger, S., Loewith, R. & Hall, M. N. TOR signaling in growth and metabolism. *Cell* **124,** 471–484 (2006).
28. Reiling, J. H. & Sabatini, D. M. Stress and mTORture signaling. *Oncogene* **25,** 6373–6383 (2006).
29. Sonenberg, N. & Hinnebusch, A. G. New modes of translational control in development, behavior, and disease. *Mol. Cell* **28,** 721–729 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by NIA grants AG022303 (R.A.M.), AG025707 and AG022308 (D.E.H.), AG022307 (R.S.) and AG13319 (J.F.N. and R.S.), and the Department of Veterans Affairs (R.A.M. and R.S.) and DoD W81XWH-07-1-0605 (Z.D.S.). We wish to thank P. J. Krason, P. J. Harrison, E. Adler, V. Diaz, J. Sewald, L. Burmeister, B. Kohler, M. Han, M. Lauderdale and D. Jones for reliable technical assistance, S. Pletcher and A. Galecki for statistical assistance, and H. Warner and S. N. Austad for scientific counsel.

Author Contributions D.E.H., R.S. and R.A.M. serve as the principal investigators at the three collaborating institutions; they were responsible for project design, supervision of technical personnel, interpretation of results, and preparation of manuscript drafts. Z.D.S. proposed rapamycin for the study, and was responsible for the measures of mTOR function. J.F.N. and K. Flurkey provided advice on experimental design and interpretation, and comments on the manuscript. Lab manager C.M.A. provided advice, and supervised laboratory procedures and data collection at The Jackson Laboratory site. N.L.N. served as the project officer for the National Institute on Aging, and contributed to program development, experimental design and analysis. J.E.W. conducted and helped interpret the necropsy analyses. K. Frenkel recommended CAPE for the study, and advised on dose and route of administration. C.S.C. and M.P. recommended enalapril for the study, and advised on dose and route of administration. M.A.J. was responsible for the pharmacological analyses. E.F. supervised and conducted laboratory procedures and data collection at the University of Texas site.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.E.H. (david.harrison@jax.org).

METHODS

Mouse production, maintenance and estimation of lifespan. Mice were produced at each of the three test sites by mating CB6F1 females with C3D2F1 males to produce a genetically heterogeneous population. Details of the methods used for health monitoring were provided previously[15]; in brief, each of the three colonies was evaluated four times each year for infectious agents, including pinworm. All such tests were negative throughout the entire study period. Each test site enrolled approximately equal numbers of 19- to 21-day-old weanlings each month over a 6-month period, housing three males or four females per cage. Each site used diets that the manufacturer stated were based on the NIH-31 standard for breeding cages, and the period between weaning and the initiation of experimental diets is given below. For breeding cages, UM used Purina 5008, UT used Teklad 7912 and TJL used Purina 5K52. For weanlings before 120 days of age, UM used Purina 5008, UT used Teklad 7912 and TJL used Purina 5LG6. Starting when 120 days old, mice in the control, enalapril and CAPE groups received Purina 5LG6 at all three sites, without additives (control group) or with the test agent. Mice in the rapamycin group remained on the weanling diet until they began to receive rapamycin, in Purina 5LG6, at 600 days of age. Separate cohorts of control and rapamycin-treated mice were established in the same way 1 year later, again at each test site, but with rapamycin initiated at 270 days rather than at 600 days of age. Additional husbandry details, including accounts of tests for T-cell subset distribution and activity administered to a subset of each group, are provided elsewhere[16,17].

Removal of mice from the longevity population. The Cohort 2005 study population, distributed almost equally among the three test sites, consisted initially of 1,960 mice, of which 674 were assigned to the control group and 317–328 to each of the four treatment groups. Of these, 51 mice were removed from the study because of fighting (31 mice), accidental death (such as chip implantation or cage flooding; 13 mice), or because of technical error (error in gender assignment or diet selection; 7 mice). For survival analyses, mice were treated as alive at the date of their removal from the protocol, and lost to follow-up thereafter. These censored mice were not included in calculations of median longevity.

Estimation of age at death (lifespan). Mice were examined at least daily for signs of ill health, and were killed for humane reasons if they were so severely moribund that they were considered, by an experienced technician, unlikely to survive for more than an additional 48 h. A mouse was considered severely moribund if it showed more than one of the following clinical signs: (1) inability to eat or to drink; (2) severe lethargy, as indicated by a lack of response such as a reluctance to move when gently prodded with a forceps; (3) severe balance or gait disturbance; (4) rapid weight loss over a period of 1 week or more; or (5) a severely ulcerated or bleeding tumour. The age at which a moribund mouse was killed was taken as the best available estimate of its natural lifespan. Mice found dead were also noted at each daily inspection. Bodies were fixed for later necropsy analysis.

Control and experimental diets. TestDiet (Richmond) prepared batches of Purina 5LG6 food containing each of the test substances, as well as control diet batches, at intervals of approximately 120 days, and shipped each batch of food at the same time to each of the three test sites. Enalapril was purchased from Sigma (catalogue E6888-5G) and used at 120 mg per kg food; on the assumption that the average mouse weighs 30 g and consumes 5 g of food per day, this dose supplies 20 mg enalapril per kg body weight per day. CAPE, caffeic acid phenethyl ester, was purchased from Cayman (catalogue 70750), and used at either of two doses: the high dose was 300 mg per kg food (50 mg per kg body weight per day), and the low dose was 30 mg per kg food (5 mg per kg body weight per day). Enalapril was tested because in aged humans and in rodent models of hypertension, obesity, diabetes and congestive heart failure, it has been reported to improve many of these conditions. CAPE was tested because this agent has been reported to possess antioxidant, anti-inflammatory and immunomodulatory capabilities, as well as specific toxicity to transformed and tumour cells. Lifespans of mice given enalapril or CAPE are compared with controls and those given rapamycin in Supplementary Fig. 1. Rapamycin was purchased from LC Labs. The rapamycin was microencapsulated by Southwest Research Institute (San Antonio, Texas), using a spinning disk atomization coating process with the enteric coating material Eudragit S100 (Röhm Pharma). This thermoplastic coating material increased the fraction of rapamycin that survived the food preparation process by three- to fourfold[16]. Because the coating material is water soluble only in non-acidic conditions, the encapsulated rapamycin is released in the small intestine rather than in the stomach. A pilot study showed that encapsulated rapamycin led to blood concentrations approximately tenfold higher than achieved by equivalent doses of non-encapsulated rapamycin[16]. The encapsulated rapamycin was administered at 14 mg per kg food (2.24 mg of rapamycin per kg body weight per day).

Measurement of rapamycin. Rapamycin was obtained from LC Laboratories. 32-desmethoxyrapamycin (32-RPM) was obtained from Sigma Chemical Company. HPLC grade methanol and acetonitrile were purchased from Fisher. All other reagents were purchased from Sigma Chemical Company. Milli-Q water was used for preparation of all solutions. The HPLC system consisted of a Waters 510 HPLC pump, Waters 717 autosampler, Waters 2487 UV detector, and Waters Empower chromatographic software (Waters). The HPLC analytical column was a Grace Alltima C18 (4.6 × 150 mm, 5 μm) purchased from Alltech. The mobile phase was 64% (v/v) acetonitrile and 36% water. The flow rate of the mobile phase was 1.5 ml min^{-1} and the wavelength of absorbance was 278 nm. The temperature of the HPLC analytical column was maintained at 70 °C during the chromatographic runs using an Eppendorf CH-30 column heater.

Rapamycin and 32-RPM powder were dissolved in methanol at a concentration of 1 mg ml^{-1} and stored in aliquots at −80 °C. A working stock solution was prepared each day from the methanol stock solutions at a concentration of 1 μg ml^{-1} and used to spike the calibrators. Calibrator samples were prepared daily by spiking either whole blood or mouse food with stock solutions to achieve final concentrations of 0, 4, 8, 12, 24, 100 and 200 ng ml^{-1}.

Rapamycin was quantified in mouse blood using HPLC with ultraviolet detection. Briefly, 0.5 ml of calibrators and unknown samples were mixed with 75 μl of 1.0 μg ml^{-1} 32-desmethoxy rapamycin (internal standard), 1.0 ml ZnSO4 (50 g l^{-1}) and 1.0 ml of acetone. The samples were vortexed vigorously for 20 s, then centrifuged at 2,600*g* at 23 °C temperature for 5 min (subsequent centrifugations were performed under the same conditions). Supernatants were transferred to clean test tubes, then 200 μl of 100 mM NaOH was added, followed by vortexing. Then, 2 ml of 1-chlorobutane was added and the samples were capped, vortexed (1 min), and centrifuged. The supernatants were transferred to 10 ml glass tubes and dried to residue under a stream of nitrogen at ambient temperature. The dried extracts were dissolved in 750 μl of mobile phase and then 2 ml of hexane was added to each tube. The tubes were capped, vortexed for 30 s, and centrifuged for 2 min. The hexane layers were removed and discarded. The remaining extracts were dried under nitrogen and reconstituted in 250 μl of mobile phase, and then 200 μl of the final extracts were injected into the HPLC. The ratio of the peak area of rapamycin to that of the internal standard (response ratio) for each unknown sample was compared against a linear regression of calibrator response ratios to quantify rapamycin. The concentration of rapamycin was expressed as ng per ml whole blood.

Rapamycin content of mouse chow was verified using HPLC with ultraviolet detection. Briefly, 100 mg of chow for spiked calibrators and unknown samples was crushed with a mortar and pestle, then vortexed vigorously with 20 μl of 100 μg ml^{-1} 32-RPM (internal standard) and 0.5 ml methanol. The samples were then mechanically shaken for 10 min. Next, 0.5 ml of Millipore water was added and the samples were vortexed vigorously for 20 s. The samples were centrifuged for 10 min and then 40 μl were injected into the HPLC. The ratio of the peak area of rapamycin to that of the internal standard (response ratio) was compared against a linear regression of calibrator response ratios at rapamycin concentrations of 0, 2, 4, 8, 10 and 20 ng mg^{-1} of food to quantify rapamycin. The concentration of rapamycin in food was expressed as ng per mg food (parts per million).

Rapamycin effectiveness. To assay for the status of an mTORC1 downstream effector, we measured phosphorylation of ribosomal protein S6 (Ser 240 and Ser 244), a substrate of S6 kinase 1, in visceral adipose tissue lysates in mice fed an encapsulated rapamycin diet for 420 days or a control diet with empty microcapsules. Tissues were dissected and snap frozen in liquid nitrogen for storage at −80 °C, ground into powder under liquid nitrogen and dissolved in 10 volumes of buffer (50 mM Tris-HCl (pH 7.5), 120 mM NaCl, 1% NP-40, 1 mM EDTA, 50 mM NaF, 40 mM 2-glycerophosphate, 0.1 mM Na orthovanadate (pH 10), 1 mM benzamidine, and 1× Complete protease inhibitor cocktail (Roche)). After sonication and microcentrifugation, lysates were quantified[30]. Forty micrograms of soluble protein from each extract was loaded on a 4–12% gradient PAGE and electrophoresed overnight at 5V. Gels were then transferred to nitrocellulose membranes (dry procedure), blocked and incubated with the primary antibodies (S6 ribosomal protein (5G10) rabbit monoclonal antibody, catalogue number 2217; phospho-S6 ribosomal protein (Ser235/236) antibody, catalogue number 2215; pan-actin antibody, catalogue number 4968; Cell Signaling Technologies), followed by secondary antibody (anti-rabbit IgG (H+L) peroxidase conjugated antibody, catalogue number 31460; Pierce) for detection by chemiluminescence. Signal intensities for each immunoblot were captured using a Kodak Image Station, and these were analysed using Kodak 1D image analysis software.

30. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).

Vol 460|16 July 2009|doi:10.1038/nature08130

LETTERS

A conserved ubiquitination pathway determines longevity in response to diet restriction

Andrea C. Carrano[1], Zheng Liu[1,2,3], Andrew Dillin[1,2,3] & Tony Hunter[1]

Dietary restriction extends longevity in diverse species, suggesting that there is a conserved mechanism for nutrient regulation and prosurvival responses[1]. Here we show a role for the HECT (homologous to E6AP carboxy terminus) E3 ubiquitin ligase WWP-1 as a positive regulator of lifespan in *Caenorhabditis elegans* in response to dietary restriction. We find that overexpression of *wwp-1* in worms extends lifespan by up to 20% under conditions of *ad libitum* feeding. This extension is dependent on the FOXA transcription factor *pha-4*, and independent of the FOXO transcription factor *daf-16*. Reduction of *wwp-1* completely suppresses the extended longevity of diet-restricted animals. However, the loss of *wwp-1* does not affect the long lifespan of animals with compromised mitochondrial function or reduced insulin/IGF-1 signalling. Overexpression of a mutant form of WWP-1 lacking catalytic activity suppresses the increased lifespan of diet-restricted animals, indicating that WWP-1 ubiquitin ligase activity is essential for longevity. Furthermore, we find that the E2 ubiquitin conjugating enzyme, UBC-18, is essential and specific for diet-restriction-induced longevity. UBC-18 interacts with WWP-1 and is required for the ubiquitin ligase activity of WWP-1 and the extended longevity of worms overexpressing *wwp-1*. Taken together, our results indicate that WWP-1 and UBC-18 function to ubiquitinate substrates that regulate diet-restriction-induced longevity.

HECT E3 ligases promote the ubiquitination of proteins that are essential in a variety of cellular events. The mammalian WWP1, WWP2 and ITCH family of WW domain HECT ligases (WWP ligases) were initially identified in a search for new proteins containing WW domains, which are modular protein-interaction domains recognizing short proline motifs in their partners[2]. WWP ligases have an amino-terminal C2 domain, which is a phospholipid membrane interaction motif, followed by four WW domains. To identify cellular pathways in which WWP E3 ligases are required, we have taken advantage of *C. elegans* as a model organism, which contains a single HECT WWP E3 ligase orthologue, *wwp-1* (Y65B4BR.4). Disruption of *wwp-1* using RNA interference (RNAi) yields a lethal phenotype late in embryogenesis characterized by abnormal embryogenesis despite normal cell proliferation[3]. The *wwp-1*(*ok1102*) mutant allele has a partially penetrant embryonic lethal phenotype[4]. Independent of the early developmental function of *wwp-1*, we found that the loss of *wwp-1* decreased stress resistance during adulthood (Supplementary Fig. 1a, b, e), leading us to investigate a possible role in longevity. Loss of *wwp-1* function by RNAi or mutation reduced lifespan at 25 °C (Supplementary Fig. 2a, b), but not at 20 °C (Supplementary Fig. 3a, b), consistent with a role for *wwp-1* in stress resistance. To investigate whether increased expression of *wwp-1* extended longevity in N2 (wild-type) worms, we created stable transgenic lines that express an N-terminal green fluorescent protein (GFP)–WWP-1 fusion protein, under the control of the endogenous *wwp-1* promoter in which expression of *wwp-1* messenger RNA is increased by approximately 50% (Supplementary Fig. 4). The overexpressing *wwp-1* transgenic lines (GFP::WWP-1) lived up to 20% longer than controls expressing GFP under the same promoter (Fig. 1a), indicating that *wwp-1* is a positive regulator of lifespan.

When diet is restricted, lifespan is extended in diverse species, suggesting that there is a conserved mechanism for nutrient regulation of ageing. Dietary restriction in worms can be reproduced genetically using *eat-2*(*ad1116*) mutant worms[5,6]. Reduced levels of *wwp-1* completely suppressed the extended longevity of *eat-2* mutant animals (Fig. 1b). The suppression of diet-restriction-extended lifespan by *wwp-1* depletion is unlikely to be due to increased food intake, because no difference in pharyngeal pumping rates with loss or knockdown of *wwp-1* in N2 or *eat-2*(*ad1116*) worms was observed (Supplementary Table 4).

We tested whether the loss of *wwp-1* suppressed the extended longevity of diet-restricted animals by reduced food intake imposed by bacterial dilution in liquid culture. N2 animals exhibited a bell-shaped curve for lifespan in response to varying bacterial concentrations (Fig. 1c)[7–9]. The lifespan of N2 animals grown under diet-restricted conditions was more than double that of animals fed *ad libitum* (Fig. 1c, d). In contrast, the lifespan of *wwp-1*(*ok1102*) mutant worms across the entire food concentration range did not change noticeably (Fig. 1c, d), indicating that WWP-1 has an essential role in regulating the response to nutrient intake and longevity. Similar results were seen using another method of dietary restriction, solid-plate dietary restriction[10] (Supplementary Fig. 5). To determine whether dietary restriction could affect *wwp-1* expression, we used quantitative PCR (qPCR) to quantify *wwp-1* mRNA and found no difference in *wwp-1* expression in animals grown under diet-restricted and *ad libitum* conditions (Supplementary Fig. 6). Furthermore, expression of a GFP::WWP-1 transgene partially rescued the suppression of dietary restriction longevity in *wwp-1*(*ok1102*) mutants (Supplementary Figs 4 and 7). Because the loss of *wwp-1* prevented the extension of lifespan of animals grown using three different diet-restriction methods, we conclude that *wwp-1* is essential for the increased longevity response to dietary restriction.

The FOXA transcription factor PHA-4 is required to specifically mediate diet-restriction-induced longevity in *C. elegans*[8]. RNAi reduction of *pha-4* suppressed the increased longevity of worms overexpressing *wwp-1* (Fig. 1e), but not when these worms were fed bacteria expressing double-stranded (ds)RNA against *daf-16*—the forkhead transcription factor required for the increased longevity due to reduced insulin/IGF1 signalling[11,12] (Fig. 1f). Mutations in the iron sulphur component of complex III, *isp-1*, increase longevity by reducing mitochondrial function[13–15]. *wwp-1* RNAi did not suppress the extended lifespan of *isp-1*(*qm150*) mutant animals (Fig. 1g), and

[1]Molecular and Cell Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA. [2]The Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. [3]The Glenn Center for Aging Research, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.

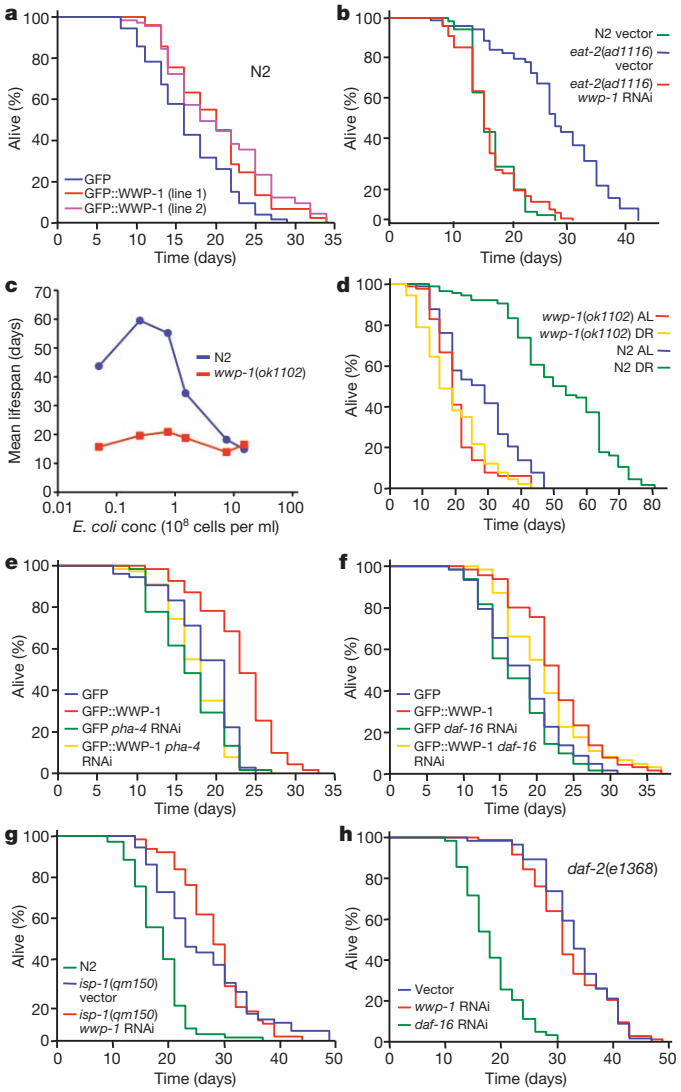


Figure 1 | *wwp-1* is required and specific for the extension of lifespan by dietary restriction. Lifespan values are given in Supplementary Tables 1 and 2. Two-way analysis of variance (ANOVA) results are presented in Supplementary Table 7. Supplementary Fig. 15a, c shows data confirming specific knockdown of *wwp-1* expression by RNAi. **a**, Two independent *wwp-1* overexpressing strains (GFP::WWP-1) can extend longevity compared to control worms expressing GFP. **b**, Lifespan analysis of *eat-2*(*ad1116*) mutant animals fed bacteria expressing *wwp-1* dsRNA or control vector. **c**, Lifespans of N2 and *wwp-1*(*ok1102*) mutant worms grown in S basal buffer with different *E. coli* concentrations. **d**, Lifespan analysis of N2 and *wwp-1*(*ok1102*) mutant worms grown in diet-restricted (DR) or *ad libitum* (AL) *E. coli* concentrations. **e**, **f**, Lifespan analysis of *wwp-1* overexpressing worms (GFP::WWP-1) or a control line fed bacteria expressing *pha-4* dsRNA (**e**), or *daf-16* dsRNA (**f**). **g**, **h**, Lifespan analysis of *isp-1*(*qm150*) (**g**) and *daf-2*(*e1368*) (**h**) fed bacteria expressing *wwp-1* dsRNA or control vector.

had only minor suppressive effects on lifespan extension of another mitochondrial mutant, *clk-1*(*qm30*), and in *cyc-1* RNAi-treated worms (Supplementary Fig. 8). Partial loss of function mutations in the insulin/IGF-1 receptor homologue, DAF-2, increase lifespan in a *daf-16*-dependent, *pha-4*-independent manner[8,16]. RNAi depletion of *wwp-1* had no effect on the long lifespan of *daf-2* mutant animals (Fig. 1h and Supplementary Fig. 9a). Our results indicate that loss of *wwp-1* does not make animals sick, but rather specifically regulates the response to dietary restriction that results in extended longevity.

Ubiquitination by HECT ligases requires the intermolecular transfer of ubiquitin from an associated E2 to the E3 ligase, before transfer to a lysine in the target protein[17]. These transfers depend on the formation of a thioester bond between ubiquitin and a conserved cysteine in the HECT domain. Mutation of this cysteine renders HECT ligases catalytically inactive, and the mutants act as dominant negatives *in vivo*[18]. We established an *in vitro* ubiquitination assay for WWP-1 ligase activity using *C. elegans* embryo extract as a source of substrates. In the presence of extract, bacterially expressed glutathione *S*-transferase (GST)–WWP-1 had very robust ligase activity, which was abolished by mutation of the catalytic cysteine (C762A) of WWP-1 (Fig. 2a). We then compared the longevity of *eat-2*(*ad1116*) transgenic animals that overexpress a GFP–WWP-1(C762A) fusion protein driven by the *wwp-1* promoter to a control line expressing GFP under the same promoter. Two independent *eat-2*(*ad1116*) transgenic lines expressing the dominant negative *wwp-1* had a significantly shorter lifespan, comparable to wild-type animals (Fig. 2b). Therefore, the ubiquitin ligase activity of WWP-1 is essential for diet-restriction-induced longevity.

UBC-18 is a putative E2 that regulates pharyngeal morphogenesis during early embryonic development[19–21]. UBC-18 is homologous to human UBCH7 (also known as UBE2L3), and similar to *Saccharomyces cerevisiae* Ubc5 and Ubc4 (ref. 21). Recently, a two-hybrid screen using UBC-18 as a bait identified WWP-1 and the RING finger E3 ligases ARI-1 and F56D2.2 as UBC-18 interactors[19]. Unlike *ubc-18* and *ari-1* dsRNA, inactivation of *wwp-1* by dsRNA treatment failed to produce a pharynx-unattached phenotype in *pha-1*(*e2123*) animals, suggesting that WWP-1 may function with UBC-18 to ubiquitinate targets not involved in pharyngeal development[19]. Consistent with this, RNAi of either *ari-1* or F56D2.2 did not affect the long lifespan of the *eat-2* mutant animals (Supplementary Fig. 10).

We found that UBC-18 is indeed a functional E2. UBC-18 formed a thiol ester bond with ubiquitin (Fig. 3a), and recombinant WWP-1 ubiquitin ligase activity required UBC-18 and E1 *in vitro* (Fig. 3b). Extracts prepared from worms mutant for *ubc-18*, *ubc-18*(*ku354*)[21], greatly reduced WWP-1-dependent ubiquitin ligase activity, which was restored by the addition of recombinant UBC-18 (Fig. 3b). Finally, we confirmed that UBC-18 and WWP-1 associate *in vitro* (Fig. 3c).

We tested whether *ubc-18* was essential for diet-restriction-induced longevity. Like *wwp-1*, *ubc-18* has a role in stress resistance in *C. elegans* (Supplementary Fig. 1c–e). However, we found that overexpression of *ubc-18* was unable to extend lifespan in *C. elegans* (Supplementary Fig. 11). Possibly, UBC-18 is not limiting for WWP-1 function in lifespan. Loss of *ubc-18* function reduced lifespan at 25 °C (Supplementary Fig. 2c), but only slightly at 20 °C (Supplementary

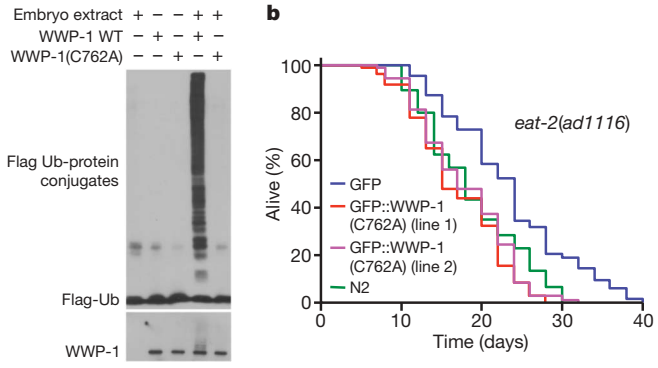


Figure 2 | WWP-1 ubiquitin ligase activity is essential for diet-restriction-induced longevity. **a**, Mutation of the conserved catalytic cysteine of WWP-1 abolishes ubiquitin (Ub) ligase activity. *In vitro* ubiquitination assay of recombinant wild-type (WT) WWP-1 or mutant WWP-1(C762A) using *C. elegans* embryo extract. **b**, *eat-2*(*ad1116*) mutant worms expressing a dominant negative *wwp-1*(C762A) have significantly shorter lifespans than control worms expressing GFP. Lifespan values are given in Supplementary Table 1.

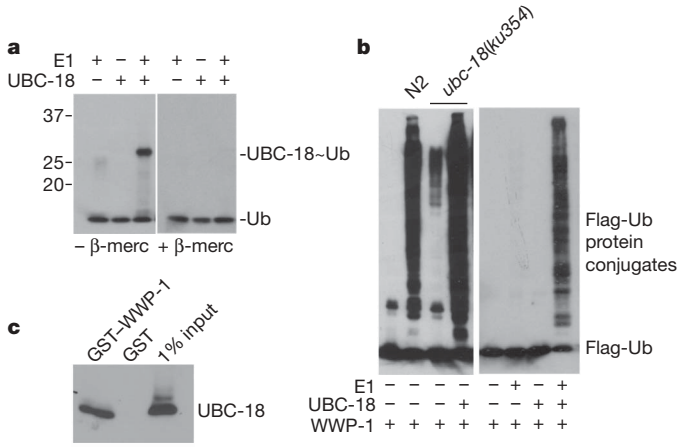


Figure 3 | WWP-1 exhibits ubiquitin ligase activity in a UBC-18 dependent manner *in vitro*. **a**, UBC-18 forms thiol-sensitive adducts with ubiquitin. *In vitro* ubiquitin (Ub) conjugation reaction in which samples were subjected to SDS–PAGE with or without β-mercaptoethanol (β-merc). **b**, UBC-18 is essential for ubiquitin ligase activity *in vitro*. Left panel, *in vitro* ubiquitination assay of wild-type WWP-1 using N2 or *ubc-18* mutant (*ubc-18*(*ku354*)) embryo extract. Recombinant UBC-18 was added in the last lane. Right panel, *in vitro* ubiquitination assay using purified components. **c**, GST pull-down assay in which GST–WWP-1 (or GST alone) bound to glutathione-agarose beads was incubated with cell lysates expressing *ubc-18*.

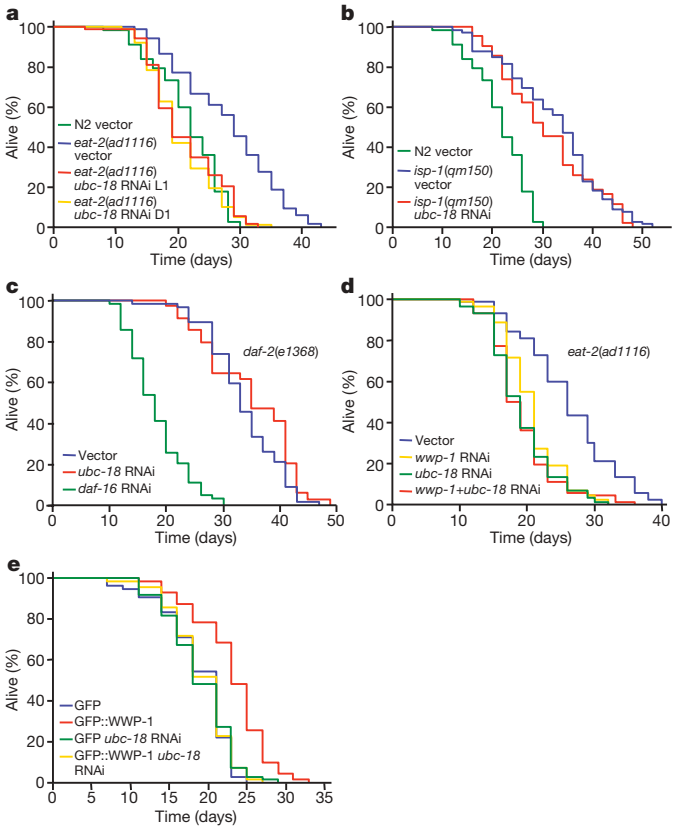


Figure 4 | WWP-1 and UBC-18 function together to regulate diet-restriction-induced longevity. Lifespan values are given in Supplementary Table 1. Two-way ANOVA results are presented in Supplementary Table 7. Knockdown of *ubc-18* expression by RNAi is shown in Supplementary Fig. 15d. **a**, Lifespan analysis of *eat-2*(*ad1116*) mutant worms fed bacteria expressing *ubc-18* dsRNA or control vector initiated after hatching of eggs (L1) or day 1 adults (D1). **b**, **c**, Lifespan analysis of *isp-1*(*qm150*) (**b**) and *daf-2*(*e1368*) (**c**) fed bacteria expressing *ubc-18* dsRNA or control vector. **d**, Lifespan analysis of *eat-2*(*ad1116*) mutant animals fed bacteria expressing *wwp-1* dsRNA and vector (*wwp-1* RNAi), *ubc-18* dsRNA and vector (*ubc-18* RNAi), *wwp-1* and *ubc-18* dsRNA (*wwp-1* + *ubc-18* RNAi), or control vector. **e**, Lifespan analysis of *wwp-1* overexpressing worms (GFP::WWP-1) or control worms (GFP) fed bacteria expressing *ubc-18* dsRNA or control vector.

Fig. 3c, d). However, RNAi depletion of *ubc-18* completely suppressed the increased longevity of *eat-2* mutants (Fig. 4a). This decreased lifespan is unlikely to be due to impaired pharynx function, because RNAi was initiated at the L1 stage when the pharynx is completely developed, and *ubc-18* RNAi initiated at the first day of adulthood also suppressed the increased longevity of *eat-2*(*ad1116*) animals. Like *wwp-1* depletion, we did not see a difference in pharyngeal pumping rates with loss of *ubc-18* (Supplementary Table 4), and RNAi depletion of *ubc-18* had no effect on the long lifespan of *isp-1*(*qm150*) (Fig. 4b) or *daf-2* mutant animals (Fig. 4c and Supplementary Fig. 9b). Moreover, epistasis analysis of *wwp-1* and *ubc-18* indicated that the combined knockdown of both genes by RNAi in *eat-2*(*ad1116*) animals did not shorten lifespan any further than RNAi of either single gene (Fig. 4d). Finally, knockdown of *ubc-18* suppressed the extended lifespan of *wwp-1* overexpressing animals (Fig. 4e).

In summary, the UBC-18–WWP-1 complex functions to specify the longevity response of diet-restricted animals. Because E2s often function with several E3s, it is surprising to find that *ubc-18* was not only essential, but also specific for the response to dietary restriction. M7.1 (also known as UBC-2 and LET-70) is most homologous to UBCH5 (also known as UBE2D1), a mammalian E2 that associates with HECT ubiquitin ligases. Unlike *ubc-18*, loss of *ubc-2* did not specifically suppress the extended longevity of *eat-2* mutants and resulted in general sickness of animals (Supplementary Fig. 12). The other E3s that interact with UBC-18 may be dedicated instead to the developmental function of UBC-18, as is the case for ARI-1 (ref. 19). It is interesting that expression of WWP-1 and UBC-18 is observed in several neurons localized in the head and tail of adult animals (Supplementary Fig. 13), because many recent studies in *C. elegans* and *Drosophila* suggest that signals derived from the nervous system can control longevity[22–25]. Although it is intriguing to speculate that a few key neuronal cells in the nervous system are the site of action of WWP-1–UBC-18 to regulate longevity, expression is not confined to a few neurons, as is the case for the dietary restriction regulator, SKN-1B[7]. Furthermore, expression of WWP-1–UBC-18 is found in intestinal cells, another site where longevity cues are expressed in the worm[26].

Because several transcription factors have been identified as targets for the mammalian orthologues of *wwp-1* (ref. 17), we investigated whether WWP-1 may target one of the two transcription factors essential for dietary restriction longevity in the worm: PHA-4 and SKN-1B[7,8]. The genetic epistasis analysis of *ubc-18/wwp-1* suggested that PHA-4 may be a target for ubiquitination. Using our *in vitro* ubiquitination assay for WWP-1, we were unable to detect ubiquitinated conjugates for either PHA-4 or SKN-1B in a purified system (Supplementary Fig. 14). Recently it has been shown that *pha-4* and the CeTor (also known as *let-363*) pathway antagonize one another to regulate longevity in adults[27]. Our results might suggest that *wwp-1* may feed into the CeTor pathway as well. The identification of the targets of UBC-18–WWP-1 is needed to allow precise placement of this complex in the diet-restriction pathway.

Our study uncovers for the first time, to our knowledge, a role of the ubiquitin pathway in longevity in response to dietary restriction. Given the strong conservation of *wwp-1* with mouse and human WWP1, an attractive hypothesis is that the mammalian orthologue will also be critical for diet-restriction-induced longevity. A detailed understanding of the pathways that mediate the benefits of dietary restriction may lead to new therapies for age-related diseases.

METHODS SUMMARY

***Caenorhabditis elegans* methods.** The *wwp-1* mutant strain was generated by backcrossing RB1178 (*wwp-1*(*ok1102*)) to N2 three times (Supplementary Fig. 15b). Nematodes were handled using standard methods[28].

Lifespan analysis. Lifespan analyses were performed as described[29]. Bacterial dilution diet-restricted lifespans were performed as described[8] with the following modifications: synchronized populations of eggs were hatched and grown at 20 °C on NG agar plates containing OP50 *E. coli* until the L4 larval stage when they were transferred to plates of OP50 containing 100 μg ml^{-1} FUDR. At day 1 of adulthood, worms were transferred into liquid culture. All lifespans were performed at 20 °C unless noted.

Protein extraction for *in vitro* ubiquitination assay. *Caenorhabditis elegans* embryos were isolated using an alkaline hypochlorite solution from gravid N2 worms grown at 20 °C[30]. The embryos were resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 0.75 mM EDTA, 1.5 mM dithiothreitol (DTT), 2.5 mM PMSF, 1 μg ml^{-1} aprotinin, 1 μg ml^{-1} leupeptin, 1 μg ml^{-1} pepstatin A) and homogenized with 30 strokes in a Dounce homogenizer. The extract was centrifuged 15,000*g* at 4 °C and stored at −70 °C.

***In vitro* ubiquitination assay.** The ubiquitination assay was carried out by incubating 1 μg Flag-tagged ubiquitin, 0.5 μg GST–WWP-1 (wild-type or C762A mutant), 0.1 μg UBC-18 and 15–20 μg embryo extract in 30 μl reaction buffer (50 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1 mM DTT and 5 mM ATP) for 1 h at 30 °C. The reaction was stopped with sample buffer and run on denaturing protein gels. Ubiquitinated substrates were identified by anti-Flag (M2, Sigma) immunoblotting. For *in vitro* ubiquitination assays using purified components, similar conditions were administered except 0.5 μg UBC-18 and 1 μg E1 were used. To measure UBC-18 ubiquitin conjugation, a similar reaction was performed and the reaction was stopped with sample buffer lacking β-mercaptoethanol.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 October 2008; accepted 11 May 2009.
Published online 24 June 2009.

1. Mair, W. & Dillin, A. Aging and survival: the genetics of life span extension by dietary restriction. *Annu. Rev. Biochem.* **77,** 727–754 (2008).
2. Pirozzi, G. *et al.* Identification of novel human WW domain-containing proteins by cloning of ligand targets. *J. Biol. Chem.* **272,** 14611–14616 (1997).
3. Huang, K. *et al.* A HECT domain ubiquitin ligase closely related to the mammalian protein WWP1 is essential for *Caenorhabditis elegans* embryogenesis. *Gene* **252,** 137–145 (2000).
4. Astin, J. W., O'Neil, N. J. & Kuwabara, P. E. Nucleotide excision repair and the degradation of RNA pol II by the *Caenorhabditis elegans* XPA and Rsp5 orthologues, RAD-3 and WWP-1. *DNA Repair (Amst.)* **7,** 267–280 (2008).
5. Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95,** 13091–13096 (1998).
6. Mair, W., Panowski, S. H., Shaw, R. J. & Dillin, A. Optimizing dietary restriction for genetic epistasis analysis and gene discovery in *C. elegans*. *PLoS One* **4,** e4535 (2009).
7. Bishop, N. A. & Guarente, L. Two neurons mediate diet-restriction-induced longevity in *C. elegans*. *Nature* **447,** 545–549 (2007).
8. Panowski, S. H., Wolff, S., Aguilaniu, H., Durieux, J. & Dillin, A. PHA-4/Foxa mediates diet-restriction-induced longevity of *C. elegans*. *Nature* **447,** 550–555 (2007).
9. Klass, M. R. Aging in the nematode *Caenorhabditis elegans*: major biological and environmental factors influencing life span. *Mech. Ageing Dev.* **6,** 413–429 (1977).
10. Greer, E. L. *et al.* An AMPK-FOXO pathway mediates longevity induced by a novel method of dietary restriction in *C. elegans*. *Curr. Biol.* **17,** 1646–1656 (2007).
11. Lin, K., Hsin, H., Libina, N. & Kenyon, C. Regulation of the *Caenorhabditis elegans* longevity protein DAF-16 by insulin/IGF-1 and germline signaling. *Nature Genet.* **28,** 139–145 (2001).
12. Ogg, S. *et al.* The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389,** 994–999 (1997).
13. Dillin, A. *et al.* Rates of behavior and aging specified by mitochondrial function during development. *Science* **298,** 2398–2401 (2002).
14. Feng, J., Bussiere, F. & Hekimi, S. Mitochondrial electron transport is a key determinant of life span in *Caenorhabditis elegans*. *Dev. Cell* **1,** 633–644 (2001).
15. Lee, S. S. *et al.* A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature Genet.* **33,** 40–48 (2003).
16. Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* **277,** 942–946 (1997).
17. Bernassola, F., Karin, M., Ciechanover, A. & Melino, G. The HECT family of E3 ubiquitin ligases: multiple players in cancer development. *Cancer Cell* **14,** 10–21 (2008).
18. Hoppe, T. *et al.* Activation of a membrane-bound transcription factor by regulated ubiquitin/proteasome-dependent processing. *Cell* **102,** 577–586 (2000).
19. Qiu, X. & Fay, D. S. ARI-1, an RBR family ubiquitin-ligase, functions with UBC-18 to regulate pharyngeal development in *C. elegans*. *Dev. Biol.* **291,** 239–252 (2006).
20. Fay, D. S. *et al.* The coordinate regulation of pharyngeal development in *C. elegans* by *lin-35/Rb*, *pha-1*, and *ubc-18*. *Dev. Biol.* **271,** 11–25 (2004).
21. Fay, D. S., Large, E., Han, M. & Darland, M. *lin-35/Rb* and *ubc-18*, an E2 ubiquitin-conjugating enzyme, function redundantly to control pharyngeal morphogenesis in *C. elegans*. *Development* **130,** 3319–3330 (2003).
22. Wolkow, C. A., Kimura, K. D., Lee, M. S. & Ruvkun, G. Regulation of *C. elegans* life-span by insulinlike signaling in the nervous system. *Science* **290,** 147–150 (2000).
23. Apfeld, J. & Kenyon, C. Regulation of lifespan by sensory perception in *Caenorhabditis elegans*. *Nature* **402,** 804–809 (1999).
24. Parkes, T. L. *et al.* Extension of *Drosophila* lifespan by overexpression of human SOD1 in motorneurons. *Nature Genet.* **19,** 171–174 (1998).
25. Alcedo, J. & Kenyon, C. Regulation of *C. elegans* longevity by specific gustatory and olfactory neurons. *Neuron* **41,** 45–55 (2004).
26. Libina, N., Berman, J. R. & Kenyon, C. Tissue-specific activities of *C. elegans* DAF-16 in the regulation of lifespan. *Cell* **115,** 489–502 (2003).
27. Sheaffer, K. L., Updike, D. L. & Mango, S. E. The Target of Rapamycin pathway antagonizes *pha-4/FoxA* to control development and aging. *Curr. Biol.* **18,** 1355–1364 (2008).
28. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77,** 71–94 (1974).
29. Dillin, A., Crawford, D. K. & Kenyon, C. Timing requirements for insulin/IGF-1 signaling in *C. elegans*. *Science* **298,** 830–834 (2002).
30. Hope, I. *C. elegans: A Practical Approach* Ch. 4 62–63 (Oxford Univ. Press, 1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Dillin laboratory for discussion, and P. Marks and H. Sun for their help. We thank A. Brunet for the solid-plate diet-restriction protocol. *wwp-1*(*ok1102*) was generated by the *C. elegans* Gene Knockout Consortium, and some strains were provided by the *Caenorhabditis* Genetics Center, which is funded by the National Institutes of Health (NIH) National Center for Research Resources (NCRR). This work was supported by grants from the National Cancer Institute (CA 14195, CA 54418 and CA 82683) to T.H., and grants from the National Institute of Diabetes and Digestive and Kidney Diseases (DK 070696) and the National Institute on Aging (AG 027463 and AG 032560), and the Ellison Medical and Glenn Medical Foundations to A.D. A.C.C. was supported by an American Cancer Society Postdoctoral Fellowship and The Rossi Endowment. T.H. is a Frank and Else Schilling American Cancer Society Professor.

Author Contributions A.C.C. designed the experiments and analysed the data. A.C.C. and Z.L. performed the experiments. A.D. and T.H. supervised the design and data interpretation. The manuscript was written by A.C.C. and edited by A.D. and T.H. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.D. (Dillin@salk.edu).

METHODS

Statistical analysis. JMP IN 5.1 software was used to determine means and percentiles. *P* values were calculated using the log-rank (Mantel-Cox) method. To assess differences in mortality between N2 and *wwp-1*(*ok1102*) strains in diet-restriction lifespans, a Cox proportional hazards regression model was generated in JMP IN 5.1 software. Two-way ANOVA was performed using Prism Version 5.0 software. Mean lifespan ± s.e.m. and *n* of all experiments presented in Supplementary Tables 1 and 5 were considered with strain and RNAi treatment as the factors.

Creation of WWP-1 and UBC-18 constructs for expression in *C. elegans*. To construct the plasmid expressing GFP::WWP-1 driven by the *wwp-1* endogenous promoter, sequences 644 bases upstream of the *wwp-1* coding region were amplified from genomic DNA by PCR and inserted upstream of GFP sequences in the worm expression vector pPD95.77. This construct (*Pwwp-1*::GFP), which expresses GFP under the control of the endogenous *wwp-1* promoter, was used as a control. To generate an N-terminal GFP-fusion construct, point mutagenesis (Stratagene) was performed on the GFP construct to eliminate the stop codon in the GFP complementary DNA. Full-length *wwp-1* cDNA was amplified from a first-strand cDNA from N2 worms by PCR, and inserted into the N-terminal GFP-fusion construct containing the *wwp-1* endogenous promoter downstream and in frame with the GFP sequence. This construct (*Pwwp-1*::GFP::WWP-1) expresses GFP–WWP-1 fusion protein under the control of the *wwp-1* promoter. To generate *wwp-1*(C762A) dominant negative mutants, point mutagenesis was performed on the *Pwwp-1*::GFP::WWP-1 construct to generate *Pwwp-1*::GFP::WWP-1(C762A). Primers for the *wwp-1* promoter: forward, 5′-GCT CTAGACTTGTTTCCTGATGACCTTG-3′; reverse, 5′-CGGGATCCTCGATC ATGAAACTGGCTG-3′. Primers for *wwp-1* cDNA: forward, 5′-ATGGCTCG TAATGAACCATCATCTCAGCAG-3′; reverse, 5′-CTACTCGTTTCCAAATCC TTCCGTCATCTC-3′. To construct the plasmid expressing GFP driven by the *ubc-18* endogenous promoter (*Pubc-18*::GFP), sequences 911 bases upstream of the *ubc-18* coding region were amplified from genomic DNA by PCR and inserted upstream of GFP sequences in the worm expression vector pPD95.67. To construct the plasmids expressing UBC-18 driven by the *ubc-18* promoter, sequences from 911 bases upstream of the *ubc-18* coding region to the stop codon of the *ubc-18* coding region were amplified from genomic DNA by PCR and inserted upstream of the GFP sequence in the worm expression vector pPD95.77. Primers for *ubc-18* promoter: forward, 5′-GATCTTTC TACTGTAGAC-3′; reverse, 5′-CGTTGATATAAACGCTCTAG-3′. Primers to generate *Pubc-18*::UBC-18::GFP: forward, 5′-GATCTTTCTACTGTAGAC-3′; reverse, 5′-GCGTATTCAGGCCGCTTTTC-3′. Primers to generate *Pubc-18*::UBC-18*:* forward, 5′-GATCTTTCTACTGTAGAC-3′; reverse, 5′-CTA TTCAGGCCGCCGCTTTTC-3′.

Generation of transgenic lines. Extrachromosomal array carrying transgenic strains were generated using standard microinjection methods[31]. For generation of N2 animals overexpressing *wwp-1*, *Pwwp-1*::GFP::WWP-1 plasmid DNA construct was mixed at 20 μg ml^{-1} with 80 μg ml^{-1} of pRF4(*rol-6*) construct[31]. *wwp-1*(*ok1102*) mutant animals overexpressing *wwp-1* were generated by crossing *wwp-1*(*ok1102*) mutants to *Pwwp-1*::GFP::WWP-1 males. For generation of *eat-2*(*ad1116*) worms overexpressing dominant negative WWP-1(C762A), *Pwwp-1*::GFP::WWP-1(C762A) plasmid DNA was mixed at 30 μg ml^{-1} with 80 μg ml^{-1} of pRF4 construct. Worms used as controls against *wwp-1* overexpressing strains contained 30 μg ml^{-1} *Pwwp-1*::GFP expressing construct injected with 30 μg ml^{-1} pRF4. For generation of N2 animals overexpressing *ubc-18*, *Pubc-18*::UBC-18::GFP or *Pubc-18*::UBC-18 plasmid DNA construct was mixed at 75 μg ml^{-1} with 75 μg ml^{-1} pRF4 construct. Worms used as controls against *ubc-18* overexpressing strains contained 75 μg ml^{-1} *Pubc-18*::GFP expressing construct injected with 75 μg ml^{-1} pRF4 construct. Transgenic F_1 progeny were selected on the basis of roller phenotype. Individual transgenic F_2 animals were isolated to establish independent lines.

GFP and DiI localization. Paralysed transgenic animals were assayed for GFP expression at ×10 and ×63 magnification using a Leica 6000B digital microscope. Images were acquired using Leica FW4000 and AF6000 software. For DiI assays, worms were incubated with 5 μg ml^{-1} DiI in M9 at room temperature for 2 h, then transferred to fresh plates for 1 h to destain, and examined under the fluorescence microscope. Intestinal granule autofluorescence was viewed with the 4,6-diamidino-2-phenylindole (DAPI) filter.

Stress assays. For heat-shock assays, eggs were grown on plates seeded with various RNAi treatments to D1 adulthood. Worms were transferred to plates without food and heat shocked at 35 °C and were checked every 2 h for viability. Paraquat (methyl viologen, Sigma) assays were performed as described[29]. For all stress assays a representative experiment from at least two independent experiments is presented.

Pumping-rate assays. Pumping rates were determined by counting pumps of the terminal pharyngeal bulb for 1-min intervals. For each condition, worms were treated with bacteria expressing dsRNA since hatching. The pumping rates of ten D1 adult worms per condition were measured and averaged.

Solid-plate dietary restriction. Lifespans were performed as described[10] with the following modifications: adult worms are transferred every 2–3 days to freshly seeded plates with restrictive amounts of bacteria (5×10^7 to 5×10^{10} bacteria per ml) starting at day 5 of adulthood.

RNA isolation, RT–PCR and quantitative RT–PCR. Total RNA was isolated from asynchronous populations of worms and extracted using Trizol reagent (GIBCO). cDNA was created using Superscript II RT (Invitrogen) and oligo dT primers. For quantitative PCR with reverse transcription (qRT–PCR) experiments, approximately 2,000 young adult worms with roller phenotype were hand picked from extrachromasomal array carrying transgenic lines grown at 20 °C. To determine how dietary restriction could affect *wwp-1* expression, bacterial dilution diet-restriction cultures were performed as described above but at a larger scale to collect 4,000 worms per sample. Total RNA was extracted using Trizol reagent and cDNA was created using Quantitec Reverse Transcriptase kit (Qiagen). SYBRgreen real-time qPCR experiments were performed as described in the manual using ABI Prism 7900 HT (Applied Biosystems). Quantitative PCR primers: *wwp-1* forward, 5′-AAGAAGCGCAGGAGTACGAG-3′; *wwp-1* reverse, 5′-ATTGATCGAAACGCATCTCC-3′; *pmp-3* forward, 5′-GTTCCC GTGTTCATCACTCAT-3′; *pmp-3* reverse, 5′-ACACCGTCGAGAAGCTG TAGA-3′.

Recombinant proteins and GST pull-down experiments. Full-length *wwp-1* cDNA (wild-type and C762A mutant) or *ubc-18* cDNA was inserted into pGEX-KG bacterial expression vector. Full-length *skn-1b* and *pha-4* cDNA were inserted into pHis8 bacterial expression vector. His purification was performed using Talon Metal Affinity Resin (Clontech). GST and His purifications were performed according to the manufacturer's instructions. Flag-tagged ubiquitin has been previously described[32,33]. For active protein, the GST tag on UBC-18 was removed by thrombin cleavage, and the supernatant was incubated with benzamidine Sepharose-4B beads (GE Healthcare) for 15 min at room temperature to remove thrombin. Full-length *ubc-18* cDNA was inserted into pcDNA6/Myc-His mammalian expression vector to generate a Myc–His fusion protein. GST-pull-down experiments were performed using 500 μg NIH3T3 lysates expressing *ubc-18* by transient transfection (Effectene, Qiagen). Lysates were prepared as previously described[34]. Lysates were incubated with GST–WWP-1 or GST bound to glutathione agarose beads for 1.5 h. Beads were washed four times with lysis buffer and analysed my SDS–PAGE. UBC-18 was detected by immunoblotting with an anti-Myc (9E10) antibody.

***In vitro* ubiquitination assay for PHA-4 and SKN-1B.** The ubiquitin ligase assay for PHA-4 and SKN-1B was carried out by incubating 1 μg Flag-tagged ubiquitin, 0.5 μg GST–WWP-1 (wild-type or C762A mutant), 0.5 μg UBC-18, 0.1 μg His–E1 and ~1 μg His–PHA-4/SKN-1B in 30 μl reaction buffer for 1 h at 30 °C. UBC-18 was preincubated with 1 mM PMSF to inhibit any residual thrombin activity. The reaction was stopped with sample buffer and run on a protein gel under denaturing conditions. PHA-4 and SKN-1B were detected by immunoblotting with a monoclonal anti-poly Histidine antibody (Sigma).

RNAi treatment. RNAi-treated strains were fed *E. coli* (HT115) containing an empty vector or *E. coli* expressing dsRNA against the gene of interest. The *wwp-1* and *pha-4* RNAi clones were from J. Ahringer's RNAi library[35]. RNAi clones against *ubc-18*, *ari-1* and F56D2.2 were from M. Vidal's RNAi library[36]. RNAi clones against *daf-16*, *daf-2* and *cyc-1* have been previously described[37,38]. Double RNAi experiments were performed as described[39].

31. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10,** 3959–3970 (1991).
32. Leverson, J. D. *et al.* The APC11 RING-H2 finger mediates E2-dependent ubiquitination. *Mol. Biol. Cell* **11,** 2315–2325 (2000).
33. Xia, Y., Pao, G. M., Chen, H. W., Verma, I. M. & Hunter, T. Enhancement of BRCA1 E3 ubiquitin ligase activity through direct interaction with the BARD1 protein. *J. Biol. Chem.* **278,** 5255–5263 (2003).
34. Carrano, A. C., Eytan, E., Hershko, A. & Pagano, M. SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. *Nature Cell Biol.* **1,** 193–199 (1999).
35. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421,** 231–237 (2003).
36. Rual, J. F. *et al.* Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome Res.* **14,** 2162–2168 (2004).
37. Dillin, A., Crawford, D. K. & Kenyon, C. Timing requirements for insulin/IGF-1 signaling in C. elegans. *Science* **298,** 830–834 (2002).
38. Simmer, F. *et al.* Genome-wide RNAi of *C. elegans* using the hypersensitive *rrf-3* strain reveals novel gene functions. *PLoS Biol.* **1,** e12 (2003).
39. Wolff, S. *et al.* SMK-1, an essential regulator of DAF-16-mediated longevity. *Cell* **124,** 1039–1053 (2006).

Vol 460|16 July 2009|doi:10.1038/nature08117

LETTERS

A reevaluation of X-irradiation-induced phocomelia and proximodistal limb patterning

Jenna L. Galloway[1]*, Irene Delgado[2]*, Maria A. Ros[3] & Clifford J. Tabin[1]

Phocomelia is a devastating, rare congenital limb malformation in which the long bones are shorter than normal, with the upper portion of the limb being most severely affected. In extreme cases, the hands or fingers are attached directly to the shoulder and the most proximal elements (those closest to the shoulder) are entirely missing. This disorder, previously known in both autosomal recessive and sporadic forms, showed a marked increase in incidence in the early 1960s due to the tragic toxicological effects of the drug thalidomide, which had been prescribed as a mild sedative[1,2]. This human birth defect is mimicked in developing chick limb buds exposed to X-irradiation[3–5]. Both X-irradiation[5] and thalidomide-induced phocomelia[5,6] have been interpreted as patterning defects in the context of the progress zone model, which states that a cell's proximodistal identity is determined by the length of time spent in a distal limb region termed the 'progress zone'[7]. Indeed, studies of X-irradiation-induced phocomelia have served as one of the two major experimental lines of evidence supporting the validity of the progress zone model. Here, using a combination of molecular analysis and lineage tracing in chick, we show that X-irradiation-induced phocomelia is fundamentally not a patterning defect, but rather results from a time-dependent loss of skeletal progenitors. Because skeletal condensation proceeds from the shoulder to fingers (in a proximal to distal direction), the proximal elements are differentially affected in limb buds exposed to radiation at early stages. This conclusion changes the framework for considering the effect of thalidomide and other forms of phocomelia, suggesting the possibility that the aetiology lies not in a defect in the patterning process, but rather in progenitor cell survival and differentiation. Moreover, molecular evidence that proximodistal patterning is unaffected after X-irradiation does not support the predictions of the progress zone model.

According to the progress zone model[7], proximodistal structures develop under the influence of a continuous signal, now understood to be a fibroblast growth factor (FGF)[8], produced by the overlying apical ectodermal ridge (AER). Initially, the entire limb mesenchyme has a proximal identity; however, the mesenchymal cells at the tip of the limb bud under the AER—those in the 'progress zone'—are exposed to FGF, allowing them to transition to more distal fates. As the progress zone cells divide and the limb grows, not all of these cells remain within range of the FGF signal. The fate of the cells exiting the progress zone becomes fixed, whereas those in this zone are once again respecified to more distal fates. The progress zone model elegantly explained the effects of X-irradiation on developing limb buds[5] by stating that the X-irradiation-induced cell death and resultantly smaller limb buds cause the proximal cells to remain within range of the AER-produced FGF signal for a longer time than normal, until the limb bud recovers and grows to a sufficient size. Consequently, proximal cells spend an increased amount of time in the progress zone, and ultimately are specified to distal fates[5].

The X-irradiation experiments, however, were performed before the identification of molecular markers for each limb segment. The use of such markers could provide powerful additional data supporting the progress-zone-based interpretation of irradiation-induced phocomelia or, alternatively, drawing into question the model's validity. We therefore decided to reexamine this paradigm with modern molecular tools.

We irradiated embryos at embryonic day (E)3.5 (stage 19–21) and grafted right limb buds onto host wings as described[5]. Skeletal elements were examined at E9.5 (6 days post-irradiation). As in previous studies[3,4], the phenotype was dose-dependent. Whereas unirradiated limbs developed normally on the host after grafting (see a and e in Fig. 1A), minor skeletal malformations such as elbow joint fusion were observed at low doses (see b and f in Fig. 1A). Intermediate doses resulted in the loss of the most proximal element, the humerus, and anterior digits (see c and g in Fig. 1A) and, at the

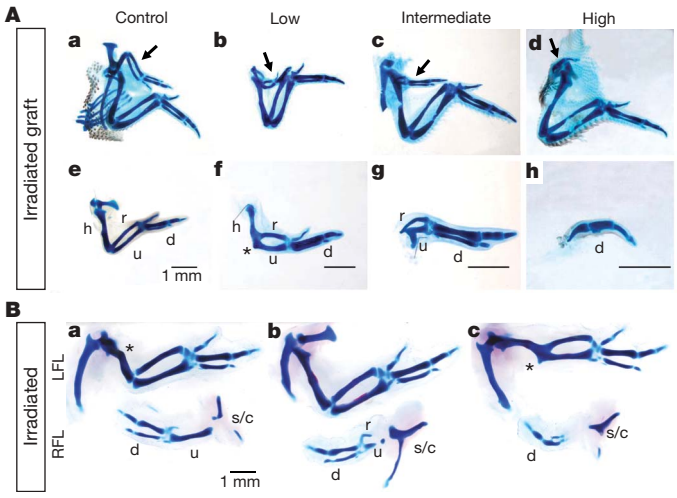


Figure 1 | X-irradiation of chick limbs causes phocomelia. **A**, Limb grafts (**a**–**d**, arrows; **e**–**h**) are shown next to host limbs. Controls (**a**, **e**) develop a humerus (h), radius (r), ulna (u) and digits (d). Low doses cause elbow fusion (asterisk) and reduced anterior digit 2 (**f**). Intermediate-dosed-limbs have no humerus, but form a partial radius and ulna and normal posterior digits (**c**, **g**). At high doses, only digits form (**d**, **h**). **B**, Irradiated RFLs from shielded embryos (lower limb; **a**–**c**) lose proximal elements compared to LFLs (upper limb; **a**–**c**). Some irradiated specimens lack an anterior (**a**, **b**) or posterior digit (**c**, lower limb). Occasionally, LFLs had mild abnormalities (asterisks). **a**–**c** are ventral images at the same scale. s, scapula; c, clavicle; scale bars, 1 mm.

[1]Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. [2]Departamento de Anatomia y Biologia Celular, Universidad de Cantabria, [3]Instituto de Biomedicina y Biotecnología de Cantabria (CSIC-UC-IDICAN), C/ Herrera Oria s/n, 39011 Santander, Spain.
*These authors contributed equally to this work.

highest irradiation dose, only digits formed (see d and h in Fig. 1A). To eliminate the possibility that the irradiation phenotype could be explained by reabsorption of proximal tissue after grafting, we performed a parallel series of experiments irradiating limbs while shielding the rest of the embryo. Using this method, we generated embryos with shorter limbs lacking proximal elements. Thus, at a dose of 17.5–20 grays, 28% of irradiated right forelimbs (RFLs) were missing the humerus, and 36% lacked the humerus and radius/ulna ($n = 36$; see lower limbs in a–c in Fig. 1B).

For the progress zone model to explain the phocomelia phenotype, the growth of the limb bud must be impaired by irradiation, while exposure to the permissive AER-produced FGF signal must not be disrupted. The irradiated limb buds are smaller than unirradiated and contralateral controls at 24, 48 and 72 hours (h) post-irradiation (Supplementary Fig. 2a, c, d). To determine whether FGF signalling continues unabated after X-irradiation, we performed whole-mount *in situ* hybridization (WISH), examining *FGF8* expression in irradiated limbs from grafted or shielded embryos. In all irradiated limbs, at all stages examined, we found normal *FGF8* expression (100%, $n = 18$; Supplementary Fig. 2a, b). Maintenance of normal FGF signalling levels from the AER requires the expression of a second signalling molecule, *sonic hedgehog* (*SHH*), in the posterior limb bud mesenchyme[9,10]. Consistent with this, we observed no loss in *SHH* expression in irradiated limb buds (100%, $n = 13$; Supplementary Fig. 2a, b). Moreover, *TFAP2A* (also known as *AP-2*), a transcription factor dependent on FGF signalling[11], was expressed in irradiated limbs at all stages analysed (100%, $n = 30$; Supplementary Fig. 2c, d), verifying that irradiation does not disrupt AER-produced FGF signalling to limb mesenchyme.

The progress zone model predicts that continual FGF signalling within the geometry of a small limb bud would lead to respecification of proximal tissue to more distal fates, thereby explaining the phocomelia. To verify this hypothesis, we investigated the specification of the three major elements of the limb: the stylopod/humerus, zeugopod/radius-ulna and autopod/digits. The best markers for these limb segments are, respectively, *MEIS1/2*, *HOXA11* and *HOXA13* (Fig. 2C). However, their expression patterns are dynamic at early stages and there is limited evidence that these transcription factors are required for proximodistal specification[12]. Nonetheless, at later stages they uniquely delineate each segment of the limb along the proximodistal axis, providing a molecular indication of their specification. The progress zone model would predict that proximal cells, being respecified to distal fates, would no longer express *MEIS1/2* and instead express distal markers such as *HOXA11* or *HOXA13*.

Surprisingly, WISH shows no difference in the relative domains of expression of these markers in irradiated and unirradiated limbs (Figs 2A, q–s, and B, i, j), although, at an early time point, expression of *MEIS1*, *MEIS2* and *HOXA11* is reduced and *HOXA13* expression was virtually undetectable in both grafted and shielded X-irradiated limbs (Fig. 2A and B; data not shown). This was probably due to a developmental delay as normal forelimb *HOXA13* expression initiates after that of *HOXA11* (ref. 13). We also found that hindlimb *HOXA13* expression was not disrupted after irradiation, consistent with *HOXA13* hindlimb expression preceding its expression in the forelimb (data not shown). After 48 h post-irradiation, *MEIS1*, *MEIS2*, *HOXA11* and *HOXA13* expression returned and demarcated the three limb segments. To confirm that there are three distinct proximodistal domains in irradiated limbs, we performed double WISH with *MEIS1* and *HOXA13*. Similar to contralateral controls, three distinct domains were identifiable in irradiated limbs: a distal *HOXA13* domain, an unstained middle region and a proximal *MEIS1* domain (Fig. 2B, e, f, i, j).

Our data, suggesting that limb segment specification is not affected by X-irradiation, raise the question of why this treatment causes a preferential loss of proximal structures. We reasoned that this could be explained if X-irradiation led to changes in apoptosis, proliferation or vasculature disproportionately in the proximal limb. X-irradiation causes both apoptosis and cell cycle arrest[14]. Consistent with this, we see an increase in apoptosis, using TdT-mediated dUTP nick end labelling (TUNEL) at 3 and 24 h post-irradiation (Fig. 3A and Supplementary Figs 3 and 4), as well as a

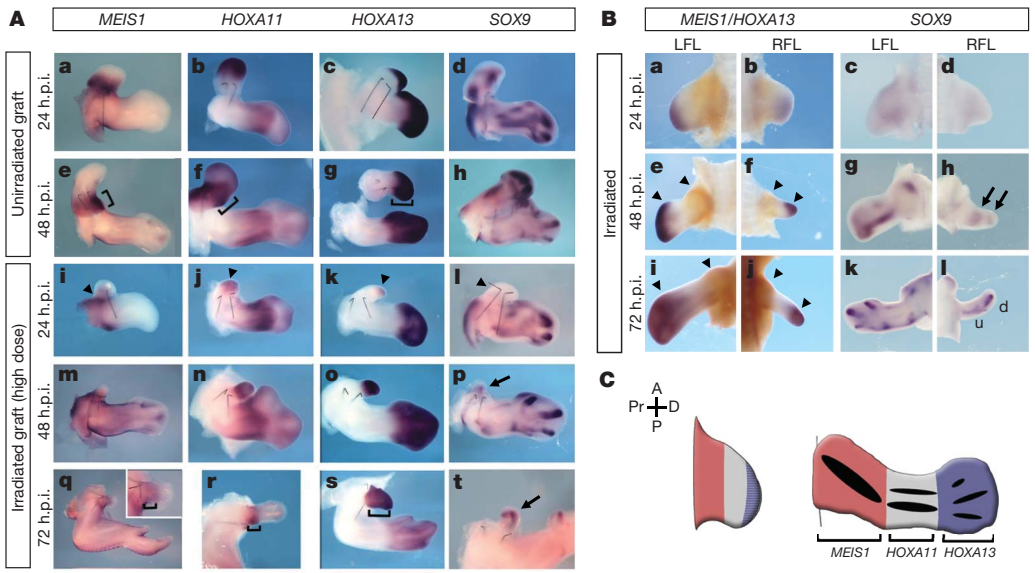


Figure 2 | Expression analysis of irradiated limbs reveals defects in differentiation but not specification of the proximal segments. **A**, In irradiated limb grafts, *MEIS1*, *HOXA11*, *HOXA13* and *SOX9* expression is reduced at 24 h post-irradiation (h.p.i.; **i–l**, arrowheads) compared with controls (**a–d**). At later time points, expression demarcates three limb segments (**m–o**; **q–s**, brackets and *MEIS1* expression in inset of **A**, **q**) as in controls (**e–g**, brackets). *SOX9* expression is found distally (**p**, **t**, arrows) compared with controls (**d**, **h**). (Grafted limbs are located in the anterior margin of host limbs.) **B**, At 24 h.p.i., irradiated RFLs displayed reduced expression of *MEIS1* (brown), *HOXA13* (purple; **b**; 91%, $n = 11$) and *SOX9* (**d**; 100%, $n = 6$) compared with LFLs (**a**, **c**). At 48 h.p.i., *MEIS1* and *HOXA13* expression in irradiated RFLs (**f**, **j**, arrowheads; 100%, $n = 12$) was similar to that in LFLs (**e**, **i**, arrowheads). *SOX9* was only expressed distally in irradiated RFLs (100%, $n = 9$) at 48 h.p.i. (**h**, arrows). **k** and **l** were stained at 96 h.p.i. (u, ulna; d, digit). **C**, At stage 23 (left), *MEIS1* (red) is proximal; *HOXA11* (grey) is distal and partially overlapping with *HOXA13* (purple). After stage 28 (right), expression demarcates three segments (Pr, proximal; D, distal; A, anterior; P, posterior).

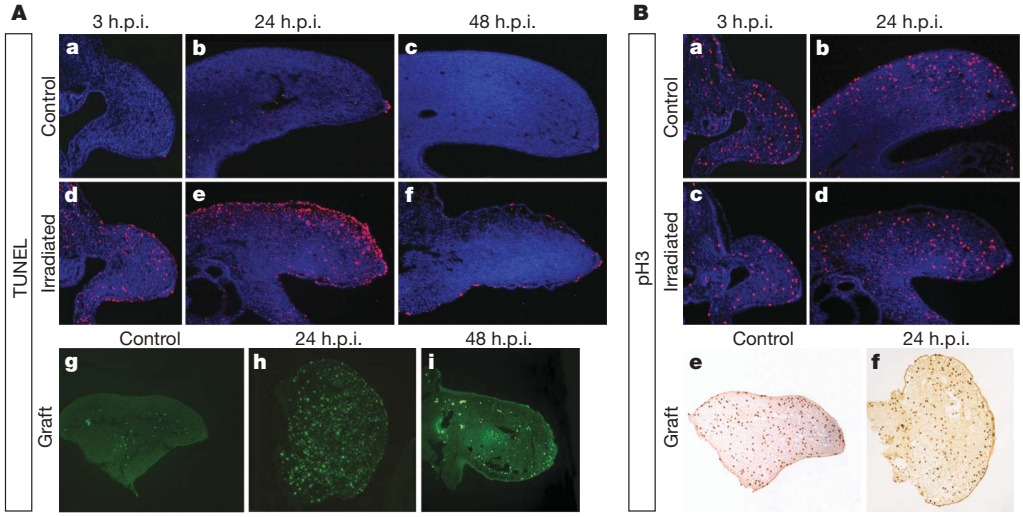


Figure 3 | Changes in cell death and proliferation do not explain the loss of proximal structures. A, An increase in TUNEL-positive cells in irradiated RFLs from shielded embryos (**d**, **e**) and irradiated grafts (**h**) compared with unirradiated controls (**a**, **b**, **g**). Levels are relatively normal at 48 h.p.i. (**c**, **f**, **i**). Autofluorescence in **g** and **i** is due to blood cells. **B**, A decrease in pH3-positive cells at 3 h.p.i. in irradiated RFLs (**c**) compared to controls (**a**). By 24 h.p.i., pH3 staining of irradiated RFLs (**b**) and grafts (**f**) is similar to that in controls (**b**, **e**). Sections are longitudinal except for **h** and **f**, which are frontal.

56% decrease in mitotic cells at 3 h post-irradiation by phosphorylated histone H3 (pH3) staining (Fig. 3B and Supplementary Figs 3 and 4). At later stages, neither apoptosis nor proliferation rates differ from contralateral and unirradiated controls or from control and irradiated grafts (Fig. 3A, c, f, i, and B, b, d, f, and Supplementary Fig. 3). Similar results were reported in previous analyses[5]. Importantly, however, neither the cell death nor the proliferative changes we observed were localized to or enriched in proximal domains of the limb. Similarly, there do not appear to be any differential changes in the vasculature along the proximodistal axis (data not shown). Thus, these effects are not likely to be responsible for the specific failure of the proximal structures to form.

We next examined whether X-irradiation differentially affects differentiation along the proximodistal limb axis. In the limb, condensation of the skeletal elements occurs progressively, with proximal elements condensing before distal ones. *SOX9* is an early marker of condensing mesenchyme and is necessary for cartilage and bone formation[15]. *SOX9* expression is first detected in the limb bud at stage 21, concurrent with or just after irradiation. We find that, at 24 h post-irradiation, *SOX9* expression is markedly reduced or lost in irradiated limb buds compared with unirradiated and contralateral controls (Fig. 2A, d, l, and B, c, d). At 48 h post-irradiation, *SOX9* expression is detected, but, notably, only in more distal regions of irradiated limb buds where condensation is initiated at later stages (Fig. 2A, h, p, t, and B, g, h, k, l). Thus, this finding suggests a model in which X-irradiation depletes the number of cells throughout the mesenchyme. This depletion has dire consequences for the proximal tissue, which is in the early stages of chondrogenic condensation, and, as a result of this loss, cannot form skeletal structures. The failure of the proximal cells to condense is probably attributed to insufficient numbers of chondrocyte precursors. More distal tissue, in contrast, has time to recover before the proximal-to-distal wave of differentiation reaches it. A similar model has been proposed to explain segment loss in FGF mutants[16,17]. Interestingly, the idea of a threshold number of cells required to form skeletal elements was proposed as an explanation as to why structures that have already been specified were affected by increasing doses of irradiation[5].

If this model is correct, irradiation at early stages affects the formation of proximal skeletal elements because cells in that domain initiate condensation first and, hence, undergo this process concurrently with the cellular response to irradiation damage. According to this view, we reasoned that irradiation at later stages should selectively affect the later-condensing distal elements. We irradiated the limbs of shielded embryos at E4.5 (stage 24) and examined their skeletal elements at E8. The irradiated limbs had severely shortened zeugopods whereas other limb segments were minimally affected (see arrow in f in Fig. 4A). As expected, there was an initial reduction in proliferation and an increase in apoptosis throughout the mesenchyme of the late-irradiated limbs (Supplementary Fig. 5a; data not shown). Corresponding to the reduction in zeugopod size, *SOX9* expression was abnormal and reduced in the middle segment of irradiated limb specimens despite the presence of *HOXA11* expression (see arrow in b and e in Fig. 4A).

We saw preferentially distal effects when we irradiated later E5.5 (stage 26) limb buds. Most strikingly, a complete loss of the autopod was observed in these irradiated limbs (see arrow in f in Fig. 4B). The proximal limb cartilage had fewer TUNEL-positive cells, consistent with our finding that the stylopod and zeugopod are only slightly affected in these limbs and in agreement with the idea that differentiated cartilage is resistant to irradiation-induced cell death[18] (Supplementary Fig. 5b). *SOX9* expression was virtually lost in the handplate at 24 and 48 h post-irradiation, in spite of continued *HOXA13* expression (see b and arrow in e in Fig. 4B; not shown). Taken together, these results support the idea that X-irradiation at successively later stages sequentially affects more distal limb segments, where chondrogenic condensation is occurring, and, as in the earlier stage irradiations, segment specification is not affected.

These conclusions depend on the use of markers which, although congruent with the limb segments, may not be involved in their specification. Therefore, we tested our conclusions in a second way. According to the progress zone model where proximal tissue and, indeed, the entire limb bud is respecified as distal after irradiation, all surviving mesenchymal cells of the limb bud, including the proximal ones, should contribute to the distal skeletal structures that form. In contrast, if specification is unaffected, but skeletal elements fail to differentiate from proximal mesenchyme, then proximal tissues should not contribute to the distal elements forming after irradiation.

To follow the fates of proximal tissue after X-irradiation, we injected the lipophilic dye DiI approximately 300 μm from the AER in control and irradiated stage 19–20 limb buds. Consistent with the results of previous fate maps[19], injection at this location in unirradiated limb buds labels the stylopod segment (arrowheads in a, d, g in Fig. 4C). In contrast, DiI injected proximally in irradiated

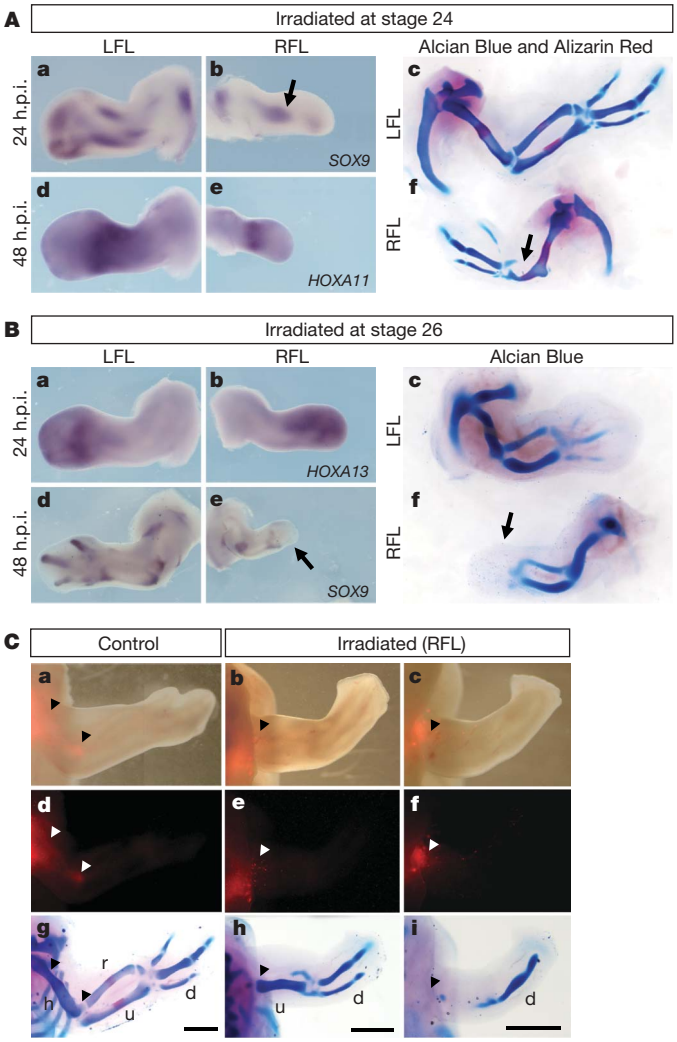


Figure 4 | X-irradiation disrupts chondrogenesis and does not convert proximal cells to distal fates. **A**, Expression of *SOX9* (100%, $n = 8$) and *HOXA11* (100%, $n = 6$), and skeletal preparations of stage-24 irradiated RFLs (**b**, **e**, **f**, arrows) and LFLs (**a**, **c**, **d**). **B**, Expression of *HOXA13* (100%, $n = 3$) and *SOX9* (100%, $n = 3$) and Alcian Blue staining of stage-26 irradiated RFLs (**b**, **e**, **f**; arrows) and LFLs (**a**, **c**, **d**). **C**, DiI (red fluorescence) labels the humerus in controls (**a**, **d**, **g**, arrowheads; 100%, $n = 5$). In irradiated RFLs, DiI did not expand distally (**b**, **e**, **h**, arrowheads; 100%, $n = 4$ and **c**, **f**, **i**, arrowheads; 100%, $n = 2$). Scale bar, 1 mm; h, humerus; r, radius; u, ulna; d, digits.

limbs had limited expansion and did not contribute to the distal skeletal elements that formed (arrowheads in b, c, e, f, h, i in Fig. 4C). Another lipophilic dye, DiO, however, when injected distally within 50 μm of the AER in irradiated limb buds, did expand into the distal regions (Supplementary Fig. 6). This result supports our hypothesis that the loss of proximal skeletal elements is due to failure of those segment progenitors to condense and not because the cells occupy the progress zone for longer periods and are respecified to distal fates.

Our data, in conjunction with earlier studies reevaluating the effect of AER removal[20], eliminate the major experimental support for the progress zone model. Although our study does not address the mechanism by which proximodistal specification does occur, it is consistent with models in which specification is driven by the influence of traditional signalling centres[12,21] rather than by time spent within a progress zone. Most importantly, our data indicate that phocomelia caused by X-irradiation results from a loss of prechondrogenic progenitors and a consequent failure in differentiation, rather than from a defect in proximodistal patterning.

Increased cell death has been thought to underlie thalidomide-induced limb truncations in chick embryos, but whether this is a result of direct activation of caspase pathways[22,23], or an indirect result of angiogenic inhibition, remains unclear[24]. Although thalidomide treatment in chick causes distal truncations, the effects of thalidomide in humans on predominantly proximal segments likewise have been suggested to be due to cell death. Indeed, cell death has been linked to phocomelia in other contexts of teratogenic exposure, such as with nitrogen mustard[25,26], and the phocomelia observed in these experiments has been interpreted mainly as a patterning defect[5,6]. However, our results indicate that in these cases, as in the irradiation experiments, cell death may lead to phocomelia not by producing a smaller limb bud in the context of a progress zone but by eliminating chondrogenic precursors during a time window when proximal condensation is compromised but distal differentiation has not yet commenced.

METHODS SUMMARY

Eggs were obtained, maintained and incubated as described[20]. *In ovo* irradiations were performed at the Hamburger Hamilton stages indicated. For the grafted limb experiments, unirradiated and irradiated stage 19–21 limb buds were dissected and grafted to the anterior wing border of a stage 24–25 host embryo as described[5] to overcome the subsequent death of the embryo. *In ovo* irradiations were performed with eggshell shielding with a 2–3 mm diameter hole, exposing the RFL. This protected the embryo and the blood islands from damage while the RFL was irradiated.

WISH and Alcian Blue and Alizarin Red skeletal preparations were performed as described[27,28]. Chick probes were *MEIS1* and *MEIS2* (ref. 21), *HOXA11* and *HOXA13* (ref. 13), *FGF8* (a gift from J. C. Izpisua Belmonte), *SHH*[29] and *SOX9* (ref. 30). *AP-2* (ChEST765g1) was linearized with NotI and an antisense probe was generated using T3. For each gene expression, irradiation dose and time period, we analysed at least two irradiated grafted specimens (both irradiated leg and wing grafts were used) and at least three irradiated limbs from shielded embryos. *In situ* cell death detection kit, TMR red (Roche) and ApoTag fluorescein *in situ* detection kit (Chemicon Int.) were used for TUNEL. pH3 immunohistochemistry was performed with anti-pH3 (1:200, Millipore) and anti-rabbit AlexaFluor 594 (1:200, Invitrogen) or biotinylated secondary antibody. All images of irradiated RFLs and LFLs are at the same magnification.

The lipophilic dyes DiI and/or DiO (Invitrogen) were prepared and injected into limb buds 1–2 h after irradiation as described[19,20]. After photographing, limbs were stained to visualize the cartilage and skeletal elements and were compared with the previous fluorescent images at the same magnification.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 February; accepted 5 May 2009.
Published online 24 June 2009.

1. Lenz, W. Thalidomide and congenital abnormalities. *Lancet* **279,** 45–46 (1962).
2. Toms, D. Thalidomide and congenital limb abnormalities. *Lancet* **280,** 400–401 (1962).
3. Goff, R. A. The relation of development status of limb formation of X-radiation sensitivity in chick embryos. I. Groso study. *J. Exp. Zool.* **151,** 177–200 (1962).
4. Pinot, M. Developpement de l'ebauche des Membres apres traitement a l'yperite azotee, irradiation aux rayons X et culture *in vitro*. Etude comparative chez l'embryon de poulet. *Annal. Embryol. Morphog.* **3,** 215–234 (1970).
5. Wolpert, L., Tickle, C. & Sampford, M. The effect of cell killing by X-irradiation on pattern formation in the chick limb. *J. Embryol. Exp. Morphol.* **50,** 175–193 (1979).
6. Tabin, C. J. A developmental model for thalidomide defects. *Nature* **396,** 322–323 (1998).
7. Summerbell, D., Lewis, J. H. & Wolpert, L. Positional information in chick limb morphogenesis. *Nature* **244,** 492–496 (1973).
8. Niswander, L., Tickle, C., Vogel, A., Booth, I. & Martin, G. R. FGF-4 replaces the apical ectodermal ridge and directs outgrowth and patterning of the limb. *Cell* **75,** 579–587 (1993).
9. Niswander, L., Jeffrey, S., Martin, G. R. & Tickle, C. A positive feedback loop coordinates growth and patterning in the vertebrate limb. *Nature* **371,** 609–612 (1994).
10. Laufer, E., Nelson, C. E., Johnson, R. L., Morgan, B. A. & Tabin, C. Sonic hedgehog and Fgf-4 act through a signaling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell* **79,** 993–1003 (1994).
11. Shen, H. *et al.* Chicken transcription factor AP-2: cloning, expression and its role in outgrowth of facial prominences and limb buds. *Dev. Biol.* **188,** 248–266 (1997).
12. Tabin, C. & Wolpert, L. Rethinking the proximodistal axis of the vertebrate limb in the molecular era. *Genes Dev.* **21,** 1433–1442 (2007).

13. Nelson, C. E. *et al.* Analysis of Hox gene expression in the chick limb bud. *Development* **122,** 1449–1466 (1996).
14. Choudhury, A., Cuddihy, A. & Bristow, R. G. Radiation and new molecular agents part I: targeting ATM-ATR checkpoints, DNA repair, and the proteasome. *Semin. Radiat. Oncol.* **16,** 51–58 (2006).
15. Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A. & de Crombrugghe, B. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* **16,** 2813–2828 (2002).
16. Sun, X., Mariani, F. V. & Martin, G. R. Functions of FGF signalling from the apical ectodermal ridge in limb development. *Nature* **418,** 501–508 (2002).
17. Mariani, F. V. & Martin, G. R. Deciphering skeletal patterning: clues from the limb. *Nature* **423,** 319–325 (2003).
18. Biggers, J. D. & Gwatkin, R. B. Effect of X-rays on the morphogenesis of the embryonic chick tibiotarsus. *Nature* **202,** 152–154 (1964).
19. Sato, K., Koizumi, Y., Takahashi, M., Kuroiwa, A. & Tamura, K. Specification of cell fate along the proximal–distal axis in the developing chick limb bud. *Development* **134,** 1397–1406 (2007).
20. Dudley, A. T., Ros, M. A. & Tabin, C. J. A re-examination of proximodistal patterning during vertebrate limb development. *Nature* **418,** 539–544 (2002).
21. Mercader, N. *et al.* Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Development* **127,** 3961–3970 (2000).
22. Knobloch, J., Shaughnessy, J. D. Jr & Ruther, U. Thalidomide induces limb deformities by perturbing the Bmp/Dkk1/Wnt signaling pathway. *FASEB J.* **21,** 1410–1421 (2007).
23. Knobloch, J., Schmitz, I., Gotz, K., Schulze-Osthoff, K. & Ruther, U. Thalidomide induces limb anomalies by PTEN stabilization, Akt suppression, and stimulation of caspase-dependent cell death. *Mol. Cell. Biol.* **28,** 529–538 (2008).
24. Therapontos, C., Erskine, L., Gardner, E., Figg, W. & Vargesson, N. Thalidomide induces limb defects by preventing angiogenic outgrowth during early limb formation. *Proc. Natl Acad. Sci. USA.* **106,** 8573–8578 (2009).
25. Salzgeber, B. Comparative study of the effects of nitrogen mustard on mesodermal and ectodermal limb bud components of chick embryos. *J. Embryol. Exp. Morphol.* **22,** 373–394 (1969).
26. Salzgeber, B. Ectodermal-mesodermal interactions in chick embryo limb buds treated with nitrogen mustard. *Dev. Growth Differ.* **17,** 295–296 (1975).
27. Wilkinson, D. G. & Nieto, M. A. Detection of messenger RNA by *in situ* hybridization to tissue sections and whole mounts. *Methods Enzymol.* **225,** 361–373 (1993).
28. McLeod, M. J. Differential staining of cartilage and bone in whole mouse fetuses by alcian blue and alizarin red S. *Teratology* **22,** 299–301 (1980).
29. Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75,** 1401–1416 (1993).
30. Healy, C., Uwanogho, D. & Sharpe, P. T. Regulation and role of Sox9 in cartilage formation. *Dev. Dyn.* **215,** 69–78 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank J. G. Arozamena and D. Connor for help with the irradiation machine and members of the Tabin-Cepko and Ros laboratories for discussions and ideas. Work in the laboratory of C.J.T. is supported by a grant, R37 HD032443, from the NIH. J.L.G. is supported by grant F32HD057701 from the National Institute of Child Health and Human Development. I.D. and M.A.R. are supported by BFU2008-00397 from the Spanish Ministry of Science and Innovation.

Author Contributions J.L.G. performed the shielding experiments. I.D. performed the grafting experiments. J.L.G., I.D., C.J.T. and M.A.R. came up with ideas, analysed data and wrote the paper. C.J.T. and M.A.R. oversaw the research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.J.T. (tabin@receptor.med.harvard.edu).

METHODS

Irradiation. For the grafted limb experiments, a Xylon International Smart 200-E irradiator at 200 kV was used at a dose rate of 2.2 grays (Gy) per min at a height of 18 cm. Three different doses were used (low, 3.6; intermediate, 5.5; high, 7.25 Gy), which were selected according to the resulting phenotypes. For the *in ovo* irradiations, the RFL bud was exposed to 17.5–20 Gy (1,750–2,000 rads) in a Faxitron 43855D irradiator at 120 kV and a height of 30.5 cm. This dose range consistently produced phocomelia in RFLs and never in LFLs. Some irradiated specimens lack anterior (see lower limbs in a, b in Fig. 1B) and posterior (see lower limb in c in Fig. 1B) digits. On a few occasions, the contralateral LFLs had mild phenotypes, including abnormally shaped skeletal elements (Fig. 1B, asterisk; humerus in upper limb of a) or fusion at the elbow (Fig. 1B, asterisk; upper limb in c), but never lost proximal elements. Whereas the phenotypes produced by shielding were more variable than the grafting method (probably due to variation in shield placement), we consistently saw phocomelia phenotypes in the affected limbs. Irradiated limbs from shielded embryos were only collected and used for *in situ* hybridizations, TUNEL or pH3 analysis if they were morphologically smaller and appeared to be affected compared with the contralateral LFLs.

Whole-mount *in situ* hybridizations and skeletal preparations. After overnight fixation in 4% paraformaldehyde/PBS and subsequent dehydration in methanol, *in situ* hybridizations were performed as described[27]. DIG-labelled probes were detected with BCIP (5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt) and NBT (nitro-blue tetrazolium chloride) (Sigma) and FITC-labelled probes were detected with INT/BCIP (Sigma). Cartilage and bone staining were performed after collection at E7–9.5 using Alcian Blue and Alizarin Red as described[28].

Analysis of cell death, proliferation and the vasculature. Limbs were fixed in 4% PFA for 1–2 h, dehydrated in ethanol, cleared in xylene and embedded in paraffin, and 7–10-μm alternate sections were collected and used for TUNEL staining and pH3 immunohistochemistry. TUNEL staining was done using the *in situ* cell death detection kit, TMR red (Roche) or ApoTag fluorescein *in situ* detection kit (Chemicon Int.). For pH3-staining, antigen retrieval was followed sequentially by incubation with anti-pH3 (1:200, Millipore) and anti-rabbit AlexaFluor 594 (1:200, Invitrogen) or biotinylated secondary antibody followed by incubation with ABC solution (Vector Laboratories) and development in diaminobenzidine. Slides for TUNEL and anti-pH3 were counterstained with DAPI to visualize the nucleus. In some irradiated-shielded specimens there were more apoptotic cells on the dorsal side closest to irradiation exposure (upper portion of limb in panel e in Fig. 3A). The number of TUNEL and pH3-positive cells were counted in irradiated RFLs, contralateral LFLs and unirradiated limbs from shielded embryos at 3 h post-irradiation, and this number was divided by the total pixel area of the limb for Supplementary Fig. 4. When calculating the pixel area, all images were at the same magnification and resolution. For both sets of data, two sections from at least four different irradiated specimens were counted.

Using an Eppendorf pressure injector (model 5242), approximately 3 μl of Alexa Fluor 488 acetylated LDL (Invitrogen) was directly injected into the circulation of control and irradiated embryos. Acetylated LDL binds and is internalized by endothelial cells and macrophages and it was used to visualize changes in the vasculature at several time points after irradiation.

Lineage tracing. The lipophilic dyes DiI and/or DiO (Invitrogen) were injected into stage 19–20 limb buds 1–2 h after irradiation as described using a calibrated reticle[19,20]. Embryos were allowed to develop for 4–5 days before collection and imaging for fluorescent signals on a Nikon Eclipse E1000 with Y-2 E/c Texas Red (DiI) and fluorescein isothiocyanate/HYQ (DiO) filters.

LETTERS

The AP-1 transcription factor *Batf* controls T_H17 differentiation

Barbara U. Schraml[1]*, Kai Hildner[1,2]*, Wataru Ise[1,2], Wan-Ling Lee[1], Whitney A.-E. Smith[1], Ben Solomon[1], Gurmukh Sahota[3], Julia Sim[4], Ryuta Mukasa[5], Saso Cemerski[1], Robin D. Hatton[5], Gary D. Stormo[3], Casey T. Weaver[5], John H. Russell[4], Theresa L. Murphy[1] & Kenneth M. Murphy[1,2]

Activator protein 1 (AP-1, also known as JUN) transcription factors are dimers of JUN, FOS, MAF and activating transcription factor (ATF) family proteins characterized by basic region and leucine zipper domains[1]. Many AP-1 proteins contain defined transcriptional activation domains, but BATF and the closely related BATF3 (refs 2, 3) contain only a basic region and leucine zipper, and are considered to be inhibitors of AP-1 activity[3–8]. Here we show that *Batf* is required for the differentiation of IL17-producing T helper (T_H17) cells[9]. T_H17 cells comprise a $CD4^+$ T-cell subset that coordinates inflammatory responses in host defence but is pathogenic in autoimmunity[10–13]. $Batf^{-/-}$ mice have normal T_H1 and T_H2 differentiation, but show a defect in T_H17 differentiation, and are resistant to experimental autoimmune encephalomyelitis. $Batf^{-/-}$ T cells fail to induce known factors required for T_H17 differentiation, such as RORγt[11] (encoded by *Rorc*) and the cytokine IL21 (refs 14–17). Neither the addition of IL21 nor the overexpression of RORγt fully restores IL17 production in $Batf^{-/-}$ T cells. The *Il17* promoter is BATF-responsive, and after T_H17 differentiation, BATF binds conserved intergenic elements in the *Il17a–Il17f* locus and to the *Il17*, *Il21* and *Il22* (ref. 18) promoters. These results demonstrate that the AP-1 protein BATF has a critical role in T_H17 differentiation.

In a gene expression survey (Supplementary Fig. 1a), we identified the basic leucine zipper transcription factor ATF-like[7] *Batf* to be highly expressed in T_H1, T_H2 and T_H17 cells compared to naive T and B cells. BATF and BATF3 (refs 2, 3) form heterodimers with JUN[6,7] and are considered to be repressors of AP-1 activity[3,5,6,8,19]. To assess the role of BATF in T cell differentiation[20], we generated $Batf^{-/-}$ mice (Supplementary Fig. 2a, b). $Batf^{-/-}$ mice lacked detectable BATF protein, were fertile and appeared healthy. BATF protein was low in naive T cells, increased in T_H2 cells, induced by activation (Supplementary Fig. 2), present in the nucleus and cytoplasm, but showed increased nuclear translocation after activation (Fig. 1a and Supplementary Fig. 1b, c). $Batf^{-/-}$ mice had normal thymus, spleen and lymph node development, and normal $CD4^+$ and $CD8^+$ T cell development (Supplementary Figs 3, 4a, b). Although *Batf*-transgenic mice had altered natural killer T cell development[21], $Batf^{-/-}$ mice had normal development of natural killer T cells (Supplementary Fig. 4c), B cells (Supplementary Fig. 4d, e), and conventional and plasmacytoid dendritic cells (Supplementary Fig. 5a, b).

$Batf^{-/-}$ T cells showed normal T_H1 and T_H2 differentiation (Supplementary Fig. 6a). Under T_H17 conditions, $Batf^{-/-}$ T cells, but not $Batf^{+/-}$ T cells, showed a marked reduction in IL17 production, but had normal levels of IL2, IFN-γ and IL10 (Fig. 1b, c). $Batf^{-/-}$ $DO11.10^+$ T cells showed loss of IL17 even after several passages under T_H17 conditions (Supplementary Fig. 6b). $Batf^{-/-}$ $CD8^+$ T cells also failed to produce IL17 (Supplementary Fig. 6c). We generated transgenic mice expressing Flag-tagged *Batf* under the control of the *Cd2* promoter[22]. *Batf*-transgenic $DO11.10^+$ $CD4^+$ T cells and $CD8^+$ T cells had increased IL17 production under T_H17 conditions compared to controls (Supplementary Fig. 6d, e). Lamina propria $CD4^+$ T cells, which constitutively express IL17 in wild-type mice[11], failed to produce IL17 in $Batf^{-/-}$ mice (Supplementary Fig. 6f).

T_H17 cells are the major pathogenic population in experimental autoimmune encephalomyelitis[10] (EAE), although factors other than IL17A and IL17F can contribute to the disease[23]. $Batf^{+/+}$ mice immunized with myelin oligodendrocyte glycoprotein peptide 35–55 (MOG(35–55); Fig. 2) developed EAE, but $Batf^{-/-}$ mice were completely resistant (Fig. 2a). At peak disease, central nervous system (CNS)-infiltrating and splenic $CD4^+$ T cells from $Batf^{+/+}$ mice produced abundant IL17 and IFN-γ, whereas T cells from $Batf^{-/-}$ mice did not produce IL17 (Fig. 2b and Supplementary Fig. 7a). Because IL6-deficient mice are resistant to EAE owing to a compensatory increase in $FOXP3^+$ T regulatory (T_{reg}) cells[14], we analysed splenic $Batf^{+/+}$ and $Batf^{-/-}$ $CD4^+$ T cells for FOXP3 expression before and after MOG(35–55) immunization (Supplementary Fig. 7b, c). $Batf^{-/-}$ mice had lower basal numbers of splenic $FOXP3^+$ T cells compared to $Batf^{+/+}$ mice, but showed no change in $FOXP3^+$ expression after MOG(35–55) immunization (Supplementary Fig. 7b, c), suggesting that their resistance to EAE is not due to an increase in T_{reg} cells. To determine whether the resistance to EAE in $Batf^{-/-}$ mice resulted from a defect within T cells or other immune cells, we injected naive $Batf^{+/+}$ $CD4^+$ T cells or PBS control buffer into mice before MOG(35–55) immunization (Fig. 2c). $Batf^{-/-}$ mice receiving PBS remained resistant to EAE, but $Batf^{-/-}$ mice receiving naive $Batf^{+/+}$ $CD4^+$ T cells developed severe EAE (Fig. 2c and Supplementary Table 1) with CNS-infiltrating IL17-producing $CD4^+$ T cells (Supplementary Fig. 7d). Thus, $Batf^{-/-}$ mice have a T-cell-intrinsic defect preventing EAE.

Batf could control T_H17 development by regulating IL6 or TGF-β signalling. IL6 receptor expression and IL6-induced STAT3 phosphorylation were normal in $Batf^{-/-}$ T cells (Supplementary Fig. 8a, b). TGF-β induced normal levels of FOXP3 in $Batf^{-/-}$ $CD4^+$ T cells (Supplementary Fig. 8d). Whereas $Batf^{-/-}$ T cells failed to fully downregulate FOXP3 in response to IL6 (ref. 12), neutralization of IL2 abrogated increased FOXP3 in $Batf^{-/-}$ T cells, without restoring IL17 production (Supplementary Fig. 8d, e). Thus, $Batf^{-/-}$ T cells

[1]Department of Pathology and Immunology, [2]Howard Hughes Medical Institute, [3]Department of Genetics, [4]Department of Molecular Biology and Pharmacology, Washington University School of Medicine, 660 South Euclid Avenue, Saint Louis, Missouri 63110, USA. [5]Department of Pathology, University of Alabama at Birmingham, University Station, Birmingham, Alabama 35294, USA.
*These authors contributed equally to this work.

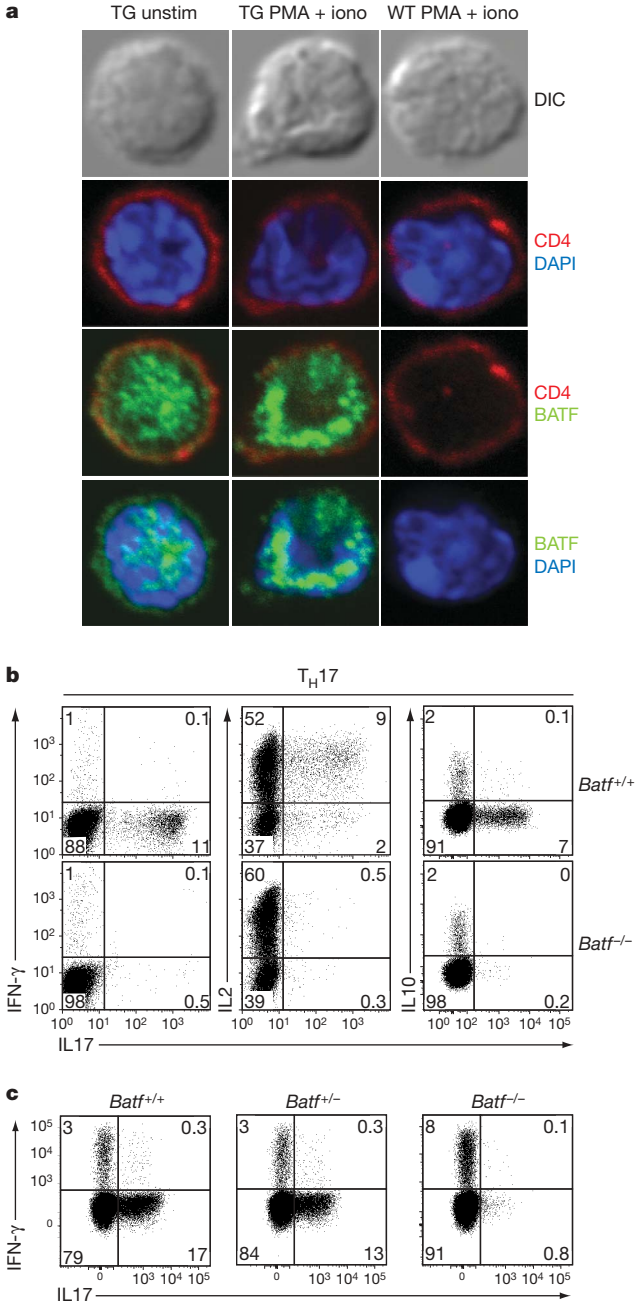


Figure 1 | Loss of IL17 production in *Batf*$^{-/-}$ T cells. **a**, DO11.10$^+$ CD4$^+$ T cells from CD2-N-Flag-*Batf* transgenic mice or littermate controls were cultured with ovalbumin (OVA) and antigen-presenting cells (APCs) under T_H2 conditions for 7 days, and stained with antibodies to CD4 and Flag. DAPI, 4,6-diamidino-2-phenylindole; DIC, differential interference contrast; iono, ionomycin; TG, transgenic, unstim, unstimulated; WT, wild type. **b**, *Batf*$^{+/+}$ and *Batf*$^{-/-}$ CD4$^+$ CD62L$^+$ CD25$^-$ T cells cultured under T_H17 conditions were restimulated with PMA and ionomycin on days 7 (left panel) or 3 (middle and right panels), and stained for IL17, IFN-γ, IL2 and IL10. **c**, IL17 and IFN-γ expression in DO11.10$^+$ CD4$^+$ T cells from *Batf*$^{+/+}$, *Batf*$^{+/-}$ and *Batf*$^{-/-}$ mice activated with OVA and APCs under T_H17 conditions. Data are representative of at least two independent experiments.

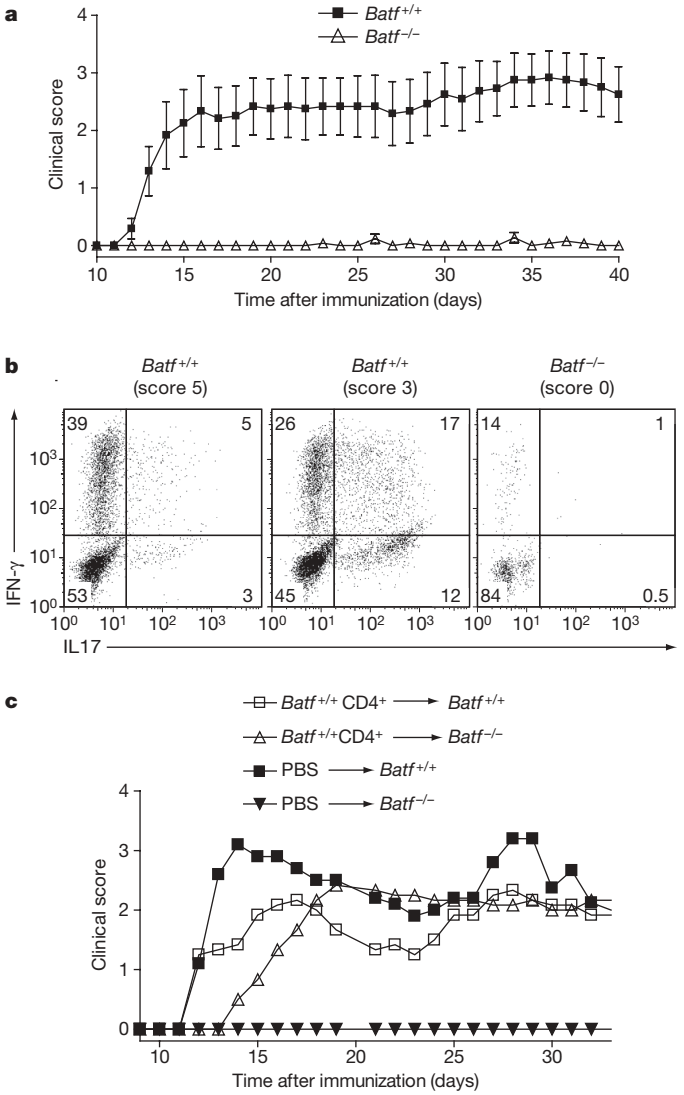


Figure 2 | *Batf*$^{-/-}$ mice are resistant to EAE. **a**, *Batf*$^{+/+}$ ($n = 12$) and *Batf*$^{-/-}$ ($n = 13$) mice were immunized with MOG(33–35) peptide. Mean clinical EAE scores ± s.e.m are shown, and are representative of two independent experiments. **b**, Thirteen days after EAE induction, CNS-infiltrating lymphocytes were stimulated with PMA and ionomycin, gated on CD4$^+$ cells, and stained for intracellular IL17 and IFN-γ (clinical scores are in parentheses, data are representative of 2–3 mice per group). **c**, *Batf*$^{+/+}$ and *Batf*$^{-/-}$ mice were injected with control PBS buffer ($n = 5$) or with 1×10^7 *Batf*$^{+/+}$ CD4$^+$ T cells ($n = 6$) 4 days before EAE induction. Mean clinical scores are shown.

exhibit normal TGF-β signalling and proximal IL6 signalling, indicating *Batf* may regulate downstream target genes.

IL21, an early target of IL6 signalling in CD4$^+$ T cells[17], is required for T_H17 development[14–16]. IL21 was reduced in *Batf*$^{-/-}$ CD4$^+$ T cells activated under T_H17 conditions (Fig. 3a). The addition of IL21 failed to rescue T_H17 development in *Batf*$^{-/-}$ T cells (Fig. 3b), but IL21-induced STAT3 phosphorylation was intact (Supplementary Fig. 8c), suggesting that *Batf* regulates other factors besides IL21 during T_H17 differentiation.

We performed DNA microarrays and quantitative PCR with reverse transcription (qRT–PCR) of *Batf*$^{+/+}$ and *Batf*$^{-/-}$ T cells activated with combinations of IL6 and/or TGF-β (Fig. 3c, d and Supplementary Fig. 9). This analysis identified several genes known to regulate T_H17 development as *Batf*-dependent (Fig. 3c, d, Supplementary Fig. 9c and Supplementary Table 2), including *Rorc*[17], *Rora*[24], the aryl hydrocarbon receptor (*Ahr*)[25,26], *Il22* (ref. 18) and *Il17*. However, *Irf4* (ref. 13) and suppressor of cytokine signalling (*Socs1–7*) gene expression were unchanged in *Batf*$^{-/-}$ T cells (Supplementary Fig. 9b and Supplementary Table 4). Early induction of RORγt was normal in *Batf*$^{-/-}$ T cells, but was not maintained 62 h after stimulation (Supplementary Fig. 11a). *Batf* seemed to be necessary for the expression of a subset of IL6-induced genes, but was not required for the expression of TGF-β-induced genes (Fig. 3c, Supplementary Fig. 9a and Supplementary Tables 2

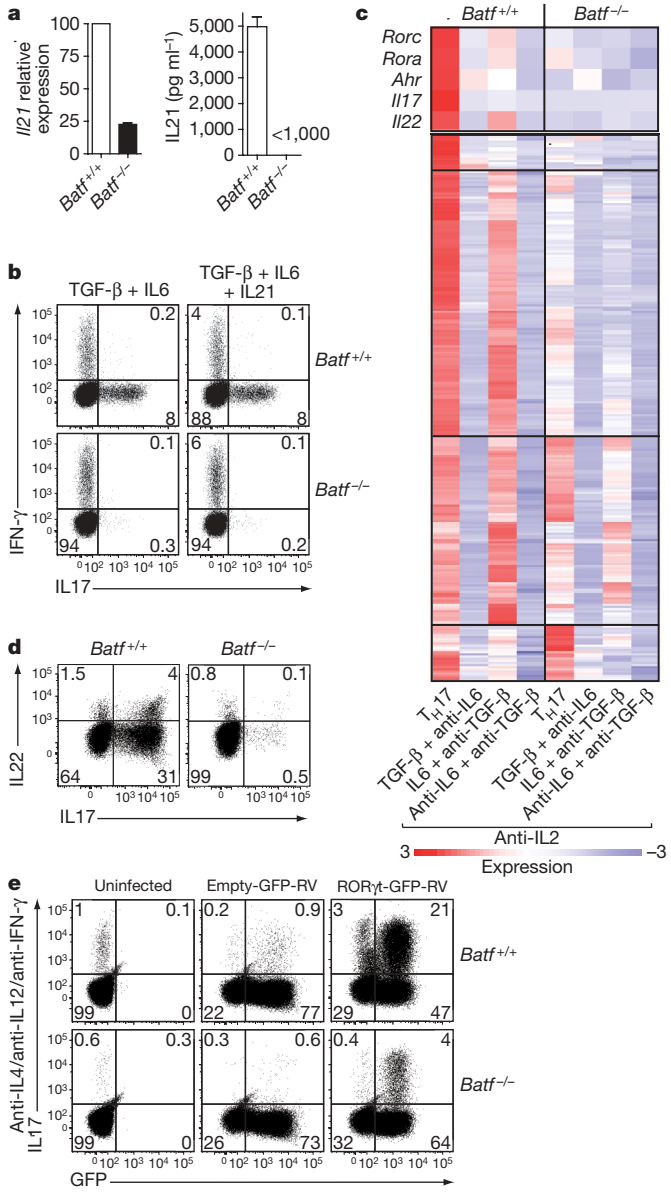


Figure 3 | BATF controls several T_H17-associated genes. **a**, IL21 expression in *Batf*$^{+/+}$ or *Batf*$^{-/-}$ T cells cultured under T_H17 conditions determined by qRT–PCR and ELISA. The mean and s.d. are shown, from 3 mice. **b**, IL17 and IFN-γ expression of CD4$^+$ CD62L$^+$ CD25$^-$ T cells cultured as in **a** in the presence or absence of IL21. **c**, Microarray analysis of anti-CD3/CD28-activated T cells at 72 h, presented as heat maps of genes fivefold-induced in *Batf*$^{+/+}$ T cells under T_H17 conditions. **d**, IL17 and IL22 expression in *Batf*$^{+/+}$ or *Batf*$^{-/-}$ CD4$^+$ T cells activated under T_H17 conditions for 3 days. **e**, Anti-CD3/CD28-activated *Batf*$^{+/+}$ or *Batf*$^{-/-}$ CD4$^+$ T cells were left uninfected or infected with retrovirus (RV) RORγt-GFP-RV or with the control empty-GFP-RV, and stained for IL17. GFP, green fluorescent protein.

and 3). However, *Batf* did not globally affect IL6-induced responses, because IL6-induced liver acute phase responses appeared normal in *Batf*$^{-/-}$ mice (Supplementary Fig. 10).

Because RORγt acts directly on the *Il17* promoter[27,28], we addressed whether RORγt could rescue T_H17 development in *Batf*$^{-/-}$ T cells. In *Batf*$^{+/+}$ T cells, retroviral RORγt expression induced 38% IL17 production, compared to only 1.6% IL17 production induced by control retrovirus (Fig. 3e and Supplementary Fig. 11c)[11,13]. However, in *Batf*$^{-/-}$ T cells, retroviral RORγt expression induced only 5.7% IL17 production (Fig. 3e and Supplementary Fig. 11c). Even under T_H17-inducing conditions, retroviral RORγt expression did not fully restore IL17 production in *Batf*$^{-/-}$ T cells (Supplementary Fig. 11b, c). Retroviral expression of both BATF and RORγt in *Batf*$^{-/-}$ T cells induced 26% IL17 production, compared to only 5% with RORγt alone and 14% with BATF alone (Supplementary Fig. 11d), suggesting that there is potential synergy between RORγt and BATF, and a possible direct action of BATF in transcription of *Il17* and other T_H17-specific genes.

We used a reverse-strand retroviral reporter[29] to examine *Il17* promoter activity in primary *Batf*$^{+/+}$ and *Batf*$^{-/-}$ T cells (Fig. 4a). Three days after activation, *Batf*$^{-/-}$ CD4$^+$ T cells showed considerably less reporter activity than *Batf*$^{+/+}$ T cells, suggesting that the proximal *Il17* promoter is *Batf*-responsive (Fig. 4a). Using chromatin immunoprecipitation (ChIP) analysis of several conserved regions within the *Il17a–Il17f* locus (Supplementary Fig. 12a), we found that BATF specifically bound to the +9.6 kilobase (kb) and +28 kb intergenic regions within 24 h after activation (Fig. 4b and Supplementary Fig. 12b, c). By day 5 after stimulation, BATF bound specifically to several intergenic regions and to the proximal *Il17a* and *Il17f* promoters (Fig. 4b and Supplementary Fig. 12b, c), with distal elements showing more rapid and stronger binding than proximal elements.

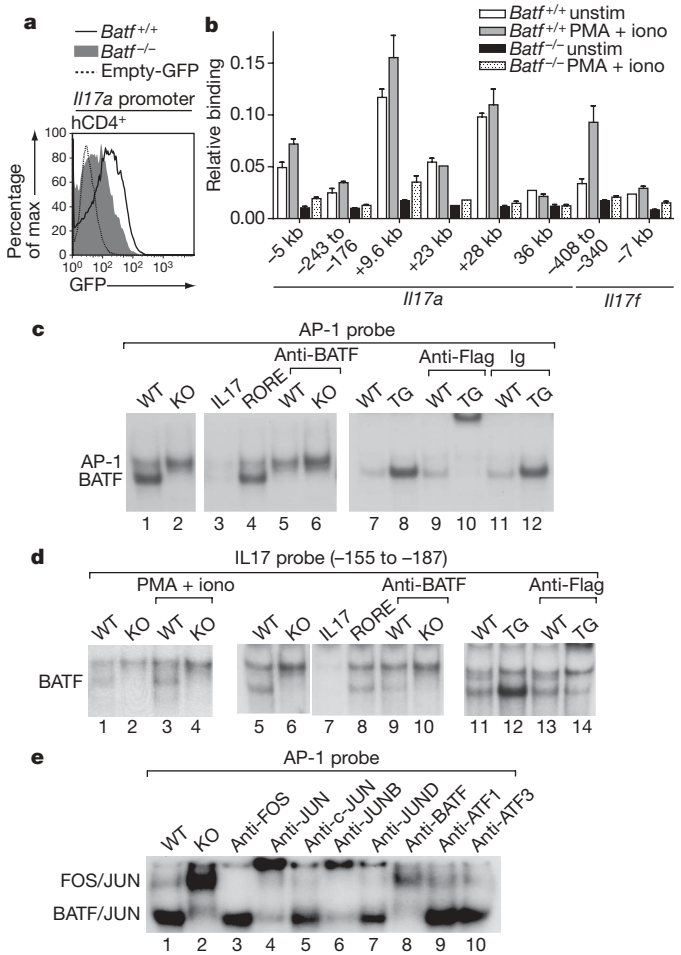


Figure 4 | BATF directly regulates IL17 expression. **a**, *Batf*$^{+/+}$ and *Batf*$^{-/-}$ CD4$^+$ T cells cultured under T_H17 conditions were infected with hCD4-pA-GFP-RV-IL17p reporter virus. GFP expression after PMA and ionomycin restimulation is shown. **b**, *Batf*$^{+/+}$ and *Batf*$^{-/-}$ CD4$^+$ T cells cultured under T_H17 conditions for 5 days were subjected to ChIP analysis of the indicated regions using an anti-BATF antibody (mean and s.d.). **c**–**e**, EMSA supershift analysis of T_H17 whole cell extracts using a consensus AP-1 (**c**, **e**) or the IL17 (−155 to −187) (**d**) probe. *Batf*$^{+/+}$ (wild type, WT), *Batf*$^{-/-}$ (knockout, KO), CD2-N-Flag-*Batf* transgenic (TG), IL17 (−155 to −187) and retinoic-acid-receptor-related orphan receptor (ROR) response element (RORE) probes were used as competitors. Anti-JUN denotes antibodies with pan-JUN specificity.

We next examined BATF binding to a consensus AP-1 probe[6] by electrophoretic mobility shift assays (EMSAs). This probe formed two complexes in *Batf*$^{+/+}$ T_H17 cell extracts (Fig. 4c) that were dependent on stimulation (Supplementary Fig. 13a). Only the upper complex formed in *Batf*$^{-/-}$ T_H17 cells (Fig. 4c). An anti-BATF antibody inhibited the lower complex. In CD2-N-Flag-*Batf*-transgenic T_H17 cell extracts, the lower complex was specifically supershifted by an anti-Flag antibody (Fig. 4c). Thus, only the lower complex binding the consensus AP-1 probe in T_H17 cells contains BATF.

Several potential BATF-binding sites were identified by EMSAs in the *Il17*, *Il21* and *Il22* proximal promoters, including the *Il17* promoter region (−188 to −210) that bound BATF in ChIP (Fig. 4b and Supplementary Fig. 13b–d). Another BATF-binding *Il17* promoter region (−155 to −187) overlapped with a reported RORγt-binding element[27]. As an EMSA probe, this region forms two complexes in T_H17 cells (Fig. 4d), with the lower complex being selectively inhibited by an anti-BATF antibody, absent in *Batf*$^{-/-}$ T_H17 cells, and supershifted by an anti-Flag antibody in *Batf*-transgenic T_H17 extracts (Fig. 4d). We confirmed BATF binding to the *Il21* and *Il22* promoters by ChIP analysis (Supplementary Fig. 13e). The program CONSENSUS[30] determined that the BATF-binding element in the *Il17*, *Il21* and *Il22* promoters resembles canonical AP-1 elements at positions 1 to 3, with variation at the remaining nucleotides (Supplementary Fig. 13f). CONSENSUS did not identify other transcription-factor-binding sites enriched near BATF-binding elements. We determined the composition of the BATF-containing complex using supershift analysis (Fig. 4e). The upper complex supershifted with a pan-anti-FOS antibody, whereas the lower complex supershifted with a pan-anti-JUN and anti-BATF antibodies. Anti-JUNB supershifted most of the lower complex, but antibodies against JUN (also known as c-JUN), JUND, ATF1 or ATF3 did not. Thus, BATF preferentially forms heterodimers with JUNB during T_H17 differentiation.

Although *Batf* and *Batf3* were considered to be AP-1 inhibitors[3–8], we have shown that they are required for the development of specific immune lineages[2]. *Batf* is selectively required for T_H17 development, but unlike *Irf4* (ref. 13), it is not required for T_H2 development. Because *Batf* is also expressed in T_H1 and T_H2 cells, it probably cooperates with other T_H17-specific factors to regulate target genes. Future work will determine whether the actions of BATF involve distinct DNA binding specificity or unique protein–protein interactions with T_H17 specific factors.

METHODS SUMMARY

Mice. *Batf*$^{-/-}$ mice were generated by homologous recombination, deleting exons 1 and 2 of the *Batf* gene on the pure 129SvEv genetic background. The neomycin-resistance cassette was removed from the targeted *Batf* allele in embryonic stem cells before the generation of mice.

T cell differentiation assays. Naive $CD4^+CD62L^+CD25^-$ T cells (also known as $CD4^+SELL^+IL2RA^-$) were isolated by cell sorting and activated with plate-bound anti-CD3 (also known as CD3E) and soluble anti-CD28 antibodies. Cultures were supplemented with anti-IL4 (11B11; hybridoma supernatant), IFN-γ (Peprotech; 0.1 ng ml^{-1}) and IL12 (Genetics Institute; 10 U ml^{-1}) for T_H1; anti-IFN-γ (H22; gift from R.D. Schreiber; 10 μg ml^{-1}), anti-IL12 (Tosh; BioXcell; 10 μg ml^{-1}) and IL4 (Peprotech; 10 ng ml^{-1}) for T_H2; anti-IL4, anti-IL12, anti-IFN-γ, IL6 (Peprotech; 20 ng ml^{-1}) and TGF-β (Peprotech; 0.5 ng ml^{-1}) for T_H17 differentiation. Unless otherwise indicated, 3 days after activation cells were restimulated with phorbol myristate acetate (PMA) and ionomycin for 4 h for intracellular cytokine analysis by flow cytometry.

Intracellular staining. For intracellular cytokine staining, cells were stained for surface markers followed by fixation with 2% formaldehyde for 15 min at room temperature. Cells were then washed once in 0.05% saponin and stained with anti-cytokine antibodies in 0.5% saponin. Anti-phospho-STAT3 antibody (BD Pharmingen) was used according to the manufacturer's recommendations. In brief, cells were stained for surface markers followed by fixation with 90% methanol at −20 °C overnight. Cells were then washed and stained for phosphorylated-STAT3 in PBS containing 3% FCS. FOXP3 staining was performed according to the manufacturer's recommendations using FOXP3 staining buffers (eBioscience).

Induction of EAE. Mice (7–10 weeks old) were immunized subcutaneously with 100 μg MOG(35–55) peptide (Sigma) emulsified in complete Freund's adjuvant (incomplete Freund's adjuvant supplemented with 500 μg *Mycobacterium tuberculosis*). One and three days later, mice were given 300 ng Pertussis Toxin (List Biological Laboratories) intraperitoneally. Clinical scores were assessed as described in Methods. For T-cell transfer experiments, mice were injected with either PBS or 10^7 *Batf*$^{+/+}$ $CD4^+$ T cells 4 days before MOG(35–55) immunization[13].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 April; accepted 5 May 2009.
Published online 5 July 2009.

1. Wagner, E. F. & Eferl, R. Fos/AP-1 proteins in bone and the immune system. *Immunol. Rev.* **208,** 126–140 (2005).
2. Hildner, K. *et al.* Batf3 deficiency reveals a critical role for $CD8\alpha^+$ dendritic cells in cytotoxic T cell immunity. *Science* **322,** 1097–1100 (2008).
3. Iacobelli, M., Wachsman, W. & McGuire, K. L. Repression of IL-2 promoter activity by the novel basic leucine zipper $p21^{SNFT}$ protein. *J. Immunol.* **165,** 860–868 (2000).
4. Blank, V. Small Maf proteins in mammalian gene control: mere dimerization partners or dynamic transcriptional regulators? *J. Mol. Biol.* **376,** 913–925 (2008).
5. Williams, K. L. *et al.* Characterization of murine BATF: a negative regulator of activator protein-1 activity in the thymus. *Eur. J. Immunol.* **31,** 1620–1627 (2001).
6. Echlin, D. R., Tae, H. J., Mitin, N. & Taparowsky, E. J. B-ATF functions as a negative regulator of AP-1 mediated transcription and blocks cellular transformation by Ras and Fos. *Oncogene* **19,** 1752–1763 (2000).
7. Dorsey, M. J. *et al.* B-ATF: a novel human bZIP protein that associates with members of the AP-1 transcription factor family. *Oncogene* **11,** 2255–2265 (1995).
8. Thornton, T. M., Zullo, A. J., Williams, K. L. & Taparowsky, E. J. Direct manipulation of activator protein-1 controls thymocyte proliferation *in vitro*. *Eur. J. Immunol.* **36,** 160–169 (2006).
9. Harrington, L. E. *et al.* Interleukin 17-producing $CD4^+$ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. *Nature Immunol.* **6,** 1123–1132 (2005).
10. Langrish, C. L. *et al.* IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *J. Exp. Med.* **201,** 233–240 (2005).
11. Ivanov, I. I. *et al.* The orphan nuclear receptor RORγt directs the differentiation program of proinflammatory IL-17^+ T helper cells. *Cell* **126,** 1121–1133 (2006).
12. Bettelli, E. *et al.* Reciprocal developmental pathways for the generation of pathogenic effector T_H17 and regulatory T cells. *Nature* **441,** 235–238 (2006).
13. Brustle, A. *et al.* The development of inflammatory T_H17 cells requires interferon-regulatory factor 4. *Nature Immunol.* **8,** 958–966 (2007).
14. Korn, T. *et al.* IL-21 initiates an alternative pathway to induce proinflammatory T_H17 cells. *Nature* **448,** 484–487 (2007).
15. Nurieva, R. *et al.* Essential autocrine regulation by IL-21 in the generation of inflammatory T cells. *Nature* **448,** 480–483 (2007).
16. Wei, L., Laurence, A., Elias, K. M. & O'Shea, J. J. IL-21 is produced by Th17 cells and drives IL-17 production in a STAT3-dependent manner. *J. Biol. Chem.* **282,** 34605–34610 (2007).
17. Zhou, L. *et al.* IL-6 programs T_H17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways. *Nature Immunol.* **8,** 967–974 (2007).
18. Liang, S. C. *et al.* Interleukin (IL)-22 and IL-17 are coexpressed by Th17 cells and cooperatively enhance expression of antimicrobial peptides. *J. Exp. Med.* **203,** 2271–2279 (2006).
19. Bower, K. E., Fritz, J. M. & McGuire, K. L. Transcriptional repression of MMP-1 by $p21^{SNFT}$ and reduced *in vitro* invasiveness of hepatocarcinoma cells. *Oncogene* **23,** 8805–8814 (2004).
20. Hess, J., Angel, P. & Schorpp-Kistner, M. AP-1 subunits: quarrel and harmony among siblings. *J. Cell Sci.* **117,** 5965–5973 (2004).
21. Williams, K. L. *et al.* BATF transgenic mice reveal a role for activator protein-1 in NKT cell development. *J. Immunol.* **170,** 2417–2426 (2003).
22. Zhumabekov, T., Corbella, P., Tolaini, M. & Kioussis, D. Improved version of a human CD2 minigene based vector for T cell-specific expression in transgenic mice. *J. Immunol. Methods* **185,** 133–140 (1995).
23. Haak, S. *et al.* IL-17A and IL-17F do not contribute vitally to autoimmune neuro-inflammation in mice. *J. Clin. Invest.* **119,** 61–69 (2009).
24. Yang, X. O. *et al.* T helper 17 lineage differentiation is programmed by orphan nuclear receptors RORα and RORγ. *Immunity* **28,** 29–39 (2008).
25. Veldhoen, M. *et al.* The aryl hydrocarbon receptor links T_H17-cell-mediated autoimmunity to environmental toxins. *Nature* **453,** 106–109 (2008).
26. Quintana, F. J. *et al.* Control of T_{reg} and T_H17 cell differentiation by the aryl hydrocarbon receptor. *Nature* **453,** 65–71 (2008).
27. Ichiyama, K. *et al.* Foxp3 inhibits RORγt-mediated IL-17A mRNA transcription through direct interaction with RORγt. *J. Biol. Chem.* **283,** 17003–17008 (2008).
28. Zhang, F., Meng, G. & Strober, W. Interactions among the transcription factors Runx1, RORγt and Foxp3 regulate the differentiation of interleukin 17-producing T cells. *Nature Immunol.* **9,** 1297–1306 (2008).
29. Zhu, H. *et al.* Unexpected characteristics of the IFN-γ reporters in nontransformed T cells. *J. Immunol.* **167,** 855–865 (2001).

30. Hertz, G. Z. & Stormo, G. D. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15,** 563–577 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Lallone for anti-BATF antibody preparation, and B. Sleckman for Cre-expressing adenovirus. This work was supported by the Howard Hughes Medical Institute (K.M.M.), and grants from the National Institutes of Health HG00249 and training grant GM07200 (G.D.S.), AI035783 (C.T.W.), AR049293 (R.D.H.), and from Daiichi-Sankyo Co. Ltd (C.T.W.).

Author Contributions B.U.S. generated *Batf*$^{-/-}$ mice, designed and analysed the experiments, interpreted results and wrote the manuscript. K.H. constructed the targeting vector and probes, transgenic vector, and recombinant BATF. W.I. helped with retroviral expression experiments. W.-L.L. helped with reverse-strand reporter analysis. W.A.-E.S. helped with mouse generation. B.S. helped with EMSA analysis. G.S. and G.D.S. performed bioinformatics analysis for the BATF binding elements. J.S. and J.H.R. helped with EAE experiments. R.M., R.D.H. and C.T.W. performed ChIP experiments. T.L.M. and S.C. performed confocal microscopy for BATF. K.M.M. directed the study and wrote the manuscript.

Author Information Microarray data are available at Array Express (http://www.ebi.ac.uk/array express/) under the accession numbers E-MEXP-1518, E-MEXP-2152 and E-MEXP-2153. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.M.M (kmurphy@wustl.edu).

METHODS

Generation of *Batf*$^{-/-}$ mice. Murine *Batf* exons 1–2 were deleted by homologous recombination via a targeting vector constructed in pLNTK[31] using a 1 kb genomic fragment (left arm) upstream of the *Batf* exon 1, and a 3.6 kb genomic fragment (right arm) downstream of exon 2. The left arm was generated by PCR from genomic DNA with the use of the oligonucleotides: left arm forward, 5′-ATTA**CTCGAG**TGAAACAAACAGGCAGTCGCAGTG-3′; left arm reverse, 5′-ATTA**CTCGAG**CCTACTACCTTTCAGGGCTACTGC-3′ (bold nucleotides indicate Xho1 restriction-enzyme sites). The right arm was generated by PCR with the use of the oligonucleotides: right arm forward, 5′-ATTA**GTCGAC**-GCATTCTTCATGGTCCTTAGCCTTGG-3′; right arm reverse, 5′-ATTA-**GTCGAC**CAGAGAATGAGAAATGTTGGAGG-3′ (bold nucleotides indicate Sal1 restriction-enzyme sites). EDJ22 embryonic stem cells were transfected with linearized targeting vector and targeted clones were identified by Southern blot analysis using probes A and B located 5′ to the left arm and 3′ to the right arm, respectively. Probe A was generated using the oligonucleotides 5′-CAACTGGGTCTGAGTCAAGAGGT-3′ and 5′-CGTAGCCGCTGATTGTTT-TAGAAC-3′ to generate a 531-bp product. Probe B was generated using the oligonucleotides 5′-ACAGCTTGAACTTCAGAGCCCTCC-3′ and 5′-CACATTTAAGTCACAATAACACTGC-3′ to generate a 772-bp product. The neomycin-resistance cassette was deleted from successfully targeted clones by *in vitro* treatment with Adeno-Cre virus (gift from B. Sleckman, Washington University), and targeted clones with successful neo deletion were identified by Southern blot using probes A and B (Supplementary Fig. 2a, b). Blastocyst injections were performed with two distinct recombinant clones, each of which generated germline transmission of the targeted *Batf* allele. Male chimaeras were crossed with 129SvEv females to establish *Batf* mutants on the pure 129SvEv genetic background. All experiments were performed with mice containing the neo-deleted mutant allele. Homozygous mice were obtained by intercrossing heterozygous siblings, and littermates were used as controls in most experiments. For some experiments 129SvEv wild-type mice purchased from Taconic served as controls. For experiments with DO11.10 transgenic *Batf*$^{-/-}$ mice, mice were crossed to BALB/c mice for at least five generations, and littermates were used as control.

For the generation of transgenic mice, *Batf* complementary DNA was cloned from CD4$^+$ T cell messenger RNA using primers 5′-GGAAGATTAGAACCAT-GCCTC-3′ and 5′-AGAAGGTCAGGGCTGGAAG-3′, and subcloned into the GFP-RV retrovirus[32]. An amino-terminal Flag tag was introduced by Quick Change Mutagenesis kit (Stratagene) using the primers 5′-GGACTACA-AAGACGATGACGACAAGCCTCACAGCTCCGACAGCA-3′ and 5′-CTTG-TCGTCATCGTCTTTGTAGTCCATGGTTCTAATCTTCCAGATC-3′. The underlined sequence indicates nucleotides used to introduce the Flag-tag. The Flag-tagged *Batf* was cloned into the CD2 microinjection cassette[33] via blunt-end strategy into a Sma1-digested CD2 microinjection cassette. Transgene expression in CD4$^+$ T cells was tested by anti-Flag western blot. CD2-N-Flag-*Batf* transgenic mice were crossed to C57BL/6 and BALB/c mice for at least five generations. Transgene-negative littermates were used as control mice. Mice were bred and maintained at the animal facilities at Washington University. All animal experiments were approved by the Animal Studies Committee at Washington University.

Visualization of lymph nodes. To visualize superficial inguinal lymph nodes, mice were injected with 50 μl of 1% Evans Blue dye solution into each hind foot pad. After 1.5 h mice were euthanized and lymph nodes were visualized using a dissecting microscope[34].

Western blot analysis. To test for residual BATF protein expression, total splenocytes from *Batf*$^{+/+}$ and *Batf*$^{-/-}$ 129SvEv mice were stimulated with anti-CD3 for 3 days under T_H17 conditions. Cells were then lysed in RIPA buffer, electrophoresed on 15% polyacrylamide gels, transferred to nitrocellulose, and analysed by western blot with rabbit anti-murine BATF polyclonal serum and HRP-conjugated anti-rabbit immunoglobulin antibody (Jackson ImmunoResearch). Affinity purified rabbit anti-murine BATF polyclonal serum (Brookwood Biomedical) was generated by immunization with full-length recombinant BATF protein. Equal protein loading was assessed by subsequent immunoblotting with antibody to β-actin (Santa Cruz Biotechnology) and HRP-conjugated anti-mouse antibody (Jackson ImmunoResearch).

For analysis of BATF protein expression in naive CD4$^+$ T cells, CD4$^+$ T cells from *Batf*$^{+/+}$ and *Batf*$^{-/-}$ 129SvEv mice were magnetically purified. Equal cell numbers were lysed in RIPA buffer and subjected to western blot analysis as described above.

For analysis of BATF expression in T_H2 cells, magnetically purified CD4$^+$ T cells from *Batf*$^{+/+}$ and *Batf*$^{-/-}$ mice were activated with anti-CD3/CD28 in the presence of IL4, anti-IL12 (Tosh) and anti-IFN-γ (H22). On day 4, cells were left unstimulated or stimulated with PMA and ionomycin for 4 h. Cells were collected by centrifugation, washed with PBS, and resuspended (100×10^6 cells ml^{-1}) in Affymetrix Chip lysis buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.5% IGEPAL, with protease inhibitors (PMSF, aprotinin, leupeptin)). After 5 min at 4 °C, nuclei were collected by centrifugation (800*g* for 3 min, 4 °C) and lysed in RIPA (100×10^6 cell equivalents per ml) with protease inhibitors. Nuclear lysates were centrifuged for 10 min at 4 °C 15,000*g*, diluted with an equal volume of 2× SDS–PAGE sample buffer containing 2-mercaptoethanol, and extracts from equal cell numbers were subjected to western blot analysis using rabbit anti-murine BATF polyclonal serum. Equal protein loading was assessed by subsequent immunoblotting with antibody to Lamin B (Santa Cruz Biotechnology) and HRP-conjugated anti-goat Ig (Jackson ImmunoResearch).

Immunohistochemistry. CD4$^+$ T cells from CD2-N-Flag-*Batf* transgenic mice were isolated by magnetic separation and either left untreated or stimulated with PMA and ionomycin for 4 h. Cells were then allowed to settle on poly-L-lysine-treated slides, fixed with 4% formaldehyde, permeabilized with 0.25% Triton X-100, and stained with an anti-Flag antibody (M2, Sigma Aldrich) according to the manufacturer's recommendations. A goat anti-mouse AF-488 (Invitrogen) antibody was used to detect anti-Flag staining. For analysis of cellular localization of BATF in T_H2 cells, DO11.10 CD4$^+$ T cells from CD2-N-Flag-*Batf* transgenic mice were isolated and differentiated with OVA and APCs under T_H2 conditions for 7 days. On day 7, cells were either left untreated or stimulated with PMA and ionomycin for 4 h. Cells were stained with anti-Flag antibody as described earlier. Cells were also stained with anti-CD4 APC antibody (BD Biosciences). Confocal images were obtained with the Olympus FV1000 microscope and software using a ×60 oil objective. The pinhole was set to 110 μm. The excitation/emission settings used for DAPI, Alexa 488 and Alexa 633 were 405/461 nm, 488/520 nm and 635/668 nm, respectively.

Further methods can be found in the Supplementary Information.

31. Gorman, J. R. *et al.* The Igκ enhancer influences the ratio of Igκ versus Igλ B lymphocytes. *Immunity* **5,** 241–252 (1996).
32. Ranganath, S. *et al.* GATA-3-dependent enhancer activity in IL-4 gene regulation. *J. Immunol.* **161,** 3822–3826 (1998).
33. Zhumabekov, T., Corbella, P., Tolaini, M. & Kioussis, D. Improved version of a human CD2 minigene based vector for T cell-specific expression in transgenic mice. *J. Immunol. Methods* **185,** 133–140 (1995).
34. Sun, Z. *et al.* Requirement for RORγ in thymocyte survival and lymphoid organ development. *Science* **288,** 2369–2373 (2000).

Vol 460|16 July 2009|doi:10.1038/nature08079

LETTERS

Cohesins form chromosomal *cis*-interactions at the developmentally regulated *IFNG* locus

Suzana Hadjur[1], Luke M. Williams[1], Natalie K. Ryan[1], Bradley S. Cobb[1], Tom Sexton[2], Peter Fraser[2], Amanda G. Fisher[1] & Matthias Merkenschlager[1]

Cohesin-mediated sister chromatid cohesion is essential for chromosome segregation and post-replicative DNA repair[1,2]. In addition, evidence from model organisms[3–6] and from human genetics[7] suggests that cohesin is involved in the control of gene expression[8,9]. This non-canonical role has recently been rationalized by the findings that mammalian cohesin complexes are recruited to a subset of DNase I hypersensitive sites and to conserved noncoding sequences by the DNA-binding protein CTCF[10–13]. CTCF functions at insulators (which control interactions between enhancers and promoters) and at boundary elements (which demarcate regions of distinct chromatin structure)[14], and cohesin contributes to its enhancer-blocking activity[10,11]. The underlying mechanisms remain unknown, and the full spectrum of cohesin functions remains to be determined. Here we show that cohesin forms the topological and mechanistic basis for cell-type-specific long-range chromosomal interactions *in cis* at the developmentally regulated cytokine locus *IFNG*. Hence, the ability of cohesin to constrain chromosome topology is used not only for the purpose of sister chromatid cohesion[1,2], but also to dynamically define the spatial conformation of specific loci. This new aspect of cohesin function is probably important for normal development[3–6] and disease[7].

One model of CTCF function is the formation of chromatin loops, which have been demonstrated at the *H19/IGF2*, β-globin and major histocompatibility complex (MHC) class II loci[15–17]. Whether cohesin

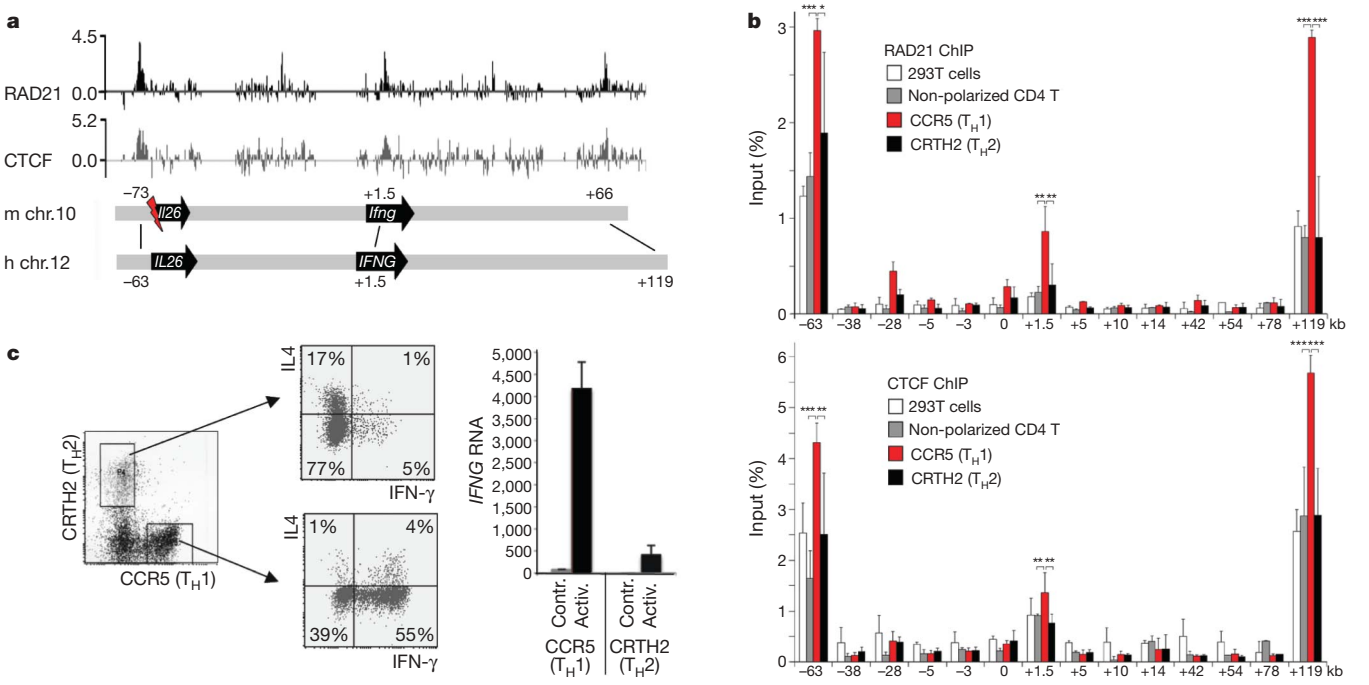


Figure 1 | Developmentally regulated cohesin and CTCF binding at *IFNG*. **a**, ChIP and genomic-tiling array data for the mouse B3 pre-B cell line[10] show log2 enrichment of the cohesin subunit RAD21 (black) and CTCF (grey) at *Ifng*. Schematic representation of the mouse *Ifng* region (m chromosome (chr.) 10) and the human *IFNG* region (h chr.12), the red flash symbol indicates genomic rearrangement at *Il26*, a pseudogene in rodents[20]. **b**, ChIP and real-time PCR mapping of RAD21 (top) and CTCF (bottom) at human *IFNG* in 293T cells (white), non-polarized CD4 T cells (grey), $CCR5^+$ T_H1 (red) and $CRTH2^+$ T_H2 (black) effector memory cells. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.005$; Student's *t*-test. Primer positions are indicated relative to the *IFNG* transcription start site. Human/mouse identity at the −63 kb site is 59% over 291 base pairs (bp) surrounding a canonical CTCF consensus motif. Two evolutionarily conserved regions (http://www.dcode.org) are found at +1.5 kb (70.2% and 69.1% identity over 104 and 217 bp), and three at +119 kb (70.0%, 71.0% and 62.2 and 69.1% identity over 120, 100 and 45 bp). Positive control sites at chr.11: 118,283 kb, chr.11: 118,333 kb, chr.8: 24,891 kb and *CD8* cluster I were included in all experiments (Supplementary Fig. 1), and data are normalized to chr.11: 118,283 kb (mean and s.d., $n = 4$). **c**, Isolation (left) of human effector memory $CCR5^+$ T_H1 and $CRTH2^+$ T_H2 cells and cytokine expression were assessed by intracellular staining and flow cytometry (middle) or real-time RT–PCR analysis (right) before (contr.) and after activation with plate bound anti-CD3 and anti-CD28 (activ.).

[1]Lymphocyte Development Group, MRC Clinical Sciences Centre, Imperial College London, Du Cane Road, London W12 0NN, UK. [2]Laboratory of Chromatin and Gene Expression, The Babraham Institute, Cambridge CB2 4AT, UK.

controls chromosome conformation is important for understanding its contribution to insulator and boundary element function[18], and for determining the significance of cohesin binding to genomic locations that are unlikely to constitute insulators or chromatin boundaries, including sites within the coding regions of active genes[10,11].

Cohesin binding within a developmentally regulated gene is exemplified by *IFNG*, which encodes the cytokine interferon-γ (IFN-γ). A conserved cohesin site is located within the first intron of the mouse *Ifng* and the human *IFNG* coding region[10]. The *Ifng* locus contains several conserved noncoding sequences and putative *cis*-regulatory elements that are located at considerable distances from the coding region[19–21] and can function as enhancers or insulators[20]. Genomic-tiling arrays[10] showed major binding sites for the cohesin subunit RAD21 and for CTCF at −73 and +66 kilobases (kb) in mouse lymphoid cell lines (Fig. 1a). Chromatin immunoprecipitation (ChIP) and real-time PCR analysis of the human 293T cell line and of primary CD4 T cells sorted from human peripheral blood demonstrated strong binding of cohesin and CTCF at orthologous sequences located 63 kb upstream of *IFNG*, and at evolutionarily conserved regions (http://www.decode.org/) at +1.5 and +119 kb in the human *IFNG* region (Fig. 1b, see legend for details). Analysis of genome-wide ChIP-sequencing data[22] confirmed −63, +1.5 and +119 kb as the major CTCF sites in human CD4 T cells, and showed that features of canonical CTCF sites, namely the histone H2A variant H2AZ (also known as H2AFZ) and trimethylated histone H3 lysine 4 (H3K4me3) were present at −63 and +1.5 kb (data not shown).

Naive, non-polarized CD4 T cells are the progenitors of specialized T helper (T_H) cell types, including T_H1 cells (which confer immunity to intracellular pathogens) and T_H2 cells (which aid antibody production and mediate responses to helminths). T_H cell differentiation has become a model for how genetically identical cells acquire distinct gene expression programs by the interplay of transcription factors and chromatin regulators[23]. The *IFNG* locus is prepared for expression during the differentiation of T_H1 cells: DNA methylation declines[20,24], cell-type-specific hypersensitive sites emerge[19–21] and *IFNG* transcription becomes inducible[23]. Conversely, *IFNG* is silenced during T_H2 differentiation[23]. To gain insight into the developmental regulation of cohesin and CTCF at the human *IFNG* locus we isolated T_H1 and T_H2 effector memory cells (Fig. 1c). These cells arise *in vivo*, can be identified by the expression of the chemokine receptor CCR5 and the prostaglandin D2 receptor CRTH2 (also known as CD294 or GPR44), respectively[25], and retain their cytokine profiles when expanded *in vitro* under non-polarizing conditions[25]. Real-time PCR with reverse-transcription (RT–PCR) showed that *IFNG* transcripts were highly induced after activation of $CCR5^+$ T_H1 cells (Fig. 1c), and intracellular staining showed the inducible expression of IFN-γ protein in $CCR5^+$ T_H1 cells, whereas $CRTH2^+$ T_H2 cells expressed interleukin-4 (IL4, Fig. 1c). The positioning of CTCF and cohesin across *IFNG* was similar between non-polarized CD4 T cells, $CCR5^+$ T_H1 cells, $CRTH2^+$ T_H2 cells and non-lymphoid 293T cells (Fig. 1b), consistent with the reported preference of CTCF and cohesin for constitutive hypersensitive sites[10]. Interestingly, however, both cohesin and CTCF were significantly more abundant at the −63, +1.5 and +119 kb sites in $CCR5^+$ T_H1 cells (Fig. 1b). Increased cohesin and CTCF binding in $CCR5^+$ T_H1 cells was selective for the *IFNG* locus, because $CCR5^+$ T_H1 and $CRTH2^+$ T_H2 cells expressed similar levels of RAD21 and CTCF (Supplementary Fig. 1a), and the occupancy of control sites outside the *IFNG* region was equivalent (Supplementary Fig. 1b).

Analysis of activating and repressive post-translational histone modifications across the *IFNG* region suggested boundary element function for the −63 site, but not for the +1.5 or the +119 kb sites in non-polarized human CD4 T cells[26]. We therefore explored whether these sites have additional functions in organizing the *IFNG* region in three-dimensional nuclear space by chromosome conformation capture[27] (3C) experiments on $CCR5^+$ T_H1 and $CRTH2^+$ T_H2 chromatin templates (Supplementary Fig. 2). The inducible nature of *IFNG* expression in $CCR5^+$ T_H1 cells allowed us to define the chromosome topology of the locus in its inducible state and independently of changes in locus conformation that may accompany high-level transcription. Using a HindIII restriction fragment containing the intronic +1.5 kb site as a bait, we looked for long-range interactions with sites across the *IFNG* genomic region. Crosslinking frequencies fell with increasing distance from the bait, but were significantly increased for the CTCF and cohesin sites at −63 and

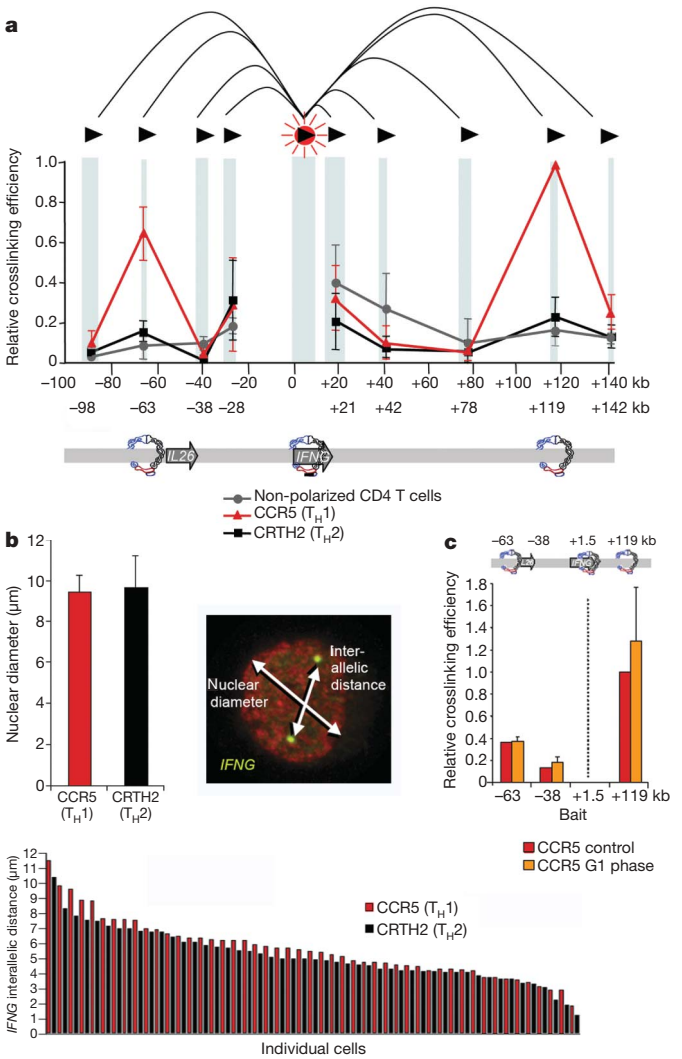


Figure 2 | Cell-type-specific long-range chromosomal interactions at *IFNG* are based on CTCF/cohesin sites. a, Chromosome conformation capture (3C) and Taqman PCR were used to determine interactions between *IFNG* +1.5 kb (beacon) and primers (black arrowheads) placed within the indicated HindIII fragments (shaded blue) in non-polarized CD4 T cells (grey), $CCR5^+$ T_H1 (red) and $CRTH2^+$ T_H2 (black) cells (Supplementary Fig. 2 and Supplementary Table 1). Crosslinking efficiencies are expressed relative to +119 kb in CCR5 cells and normalized to mouse *Acta2* (see Methods and Supplementary Fig. 2) or to human *ACTA2* with similar results (Supplementary Fig. 3a). Cohesin sites and the approximate positions of the *IFNG* and *IL26* genes are indicated (mean ± s.d., $n = 5$ for CCR5 and CRTH2, $n = 4$ for non-polarized CD4 T cells). $P < 0.005$; Student's *t*-test, T_H1 versus non-polarized, and T_H1 versus T_H2. **b**, 3D-FISH to determine the nuclear diameters (top) of $CCR5^+$ T_H1 (red, 9.4 ± 1.1 μm, $n = 50$) and $CRTH2^+$ T_H2 (black, 9.5 ± 1.5 μm, $n = 50$) cells, and the distances between *IFNG* alleles (bottom) in individual CCR5 (red, 4.5 ± 1.6 μm, $n = 50$, minimum 1.5 μm) and CRTH2 (black, 5.0 ± 1.9 μm, $n = 50$, minimum 1.1 μm) cells. **c**, Long-range interactions at *IFNG* in $CCR5^+$ T_H1 cells occur before DNA replication. 3C analysis of *IFNG* in total $CCR5^+$ T_H1 cells (red) compared to $CCR5^+$ T_H1 cells in the G1 phase of the cell cycle (orange). Data are presented as in **a** (mean and s.e.m., $n = 2$).

+119 kb, indicating long-range interactions at *IFNG* that occur selectively in CCR5$^+$ T_H1 cells (Fig. 2a). Nuclear dimensions were similar for CCR5$^+$ T_H1 and CRTH2$^+$ T_H2 cells (Fig. 2b), which rules out spatial constraints as an explanation for increased crosslinking efficiencies in CCR5$^+$ T_H1 cells. Because the *IFNG* loci on homologous chromosomes were separated by at least 1.5 μm in three-dimensional fluorescence *in situ* hybridization (3D-FISH) experiments (Fig. 2b), and long-range interactions in the *IFNG* genomic region were readily detectable in purified G1 cells (that is, before DNA replication and the formation of sister chromatids, Fig. 2c), the interactions recorded by 3C probably occurred *in cis*, not *in trans* between *IFNG* alleles. These data establish that CTCF and cohesin sites form the topological basis for cell-type-specific, long-range interactions at *IFNG* that occur preferentially in human CCR5$^+$ T_H1 cells.

We next explored the relationship of CTCF and cohesin with long-range chromosomal interactions. Depletion of *CTCF* by short interfering RNA (siRNA)-mediated knockdown in CCR5$^+$ T_H1 cells resulted in reduced 3C interactions between the *IFNG* locus (+1.5 kb) and the −63 and +119 kb sites (Supplementary Fig. 4), reminiscent of the role for CTCF in long-range interactions at other loci[15–17]. Importantly, *RAD21* knockdown substantially reduced long-range interactions of the *IFNG* coding region with the −63 and the +119 kb sites (Fig. 3a), demonstrating that the conformation of the locus depended on cohesin as well as CTCF. ChIP analysis of the IFNG region in CCR5$^+$ T_H1 cells showed that *RAD21* RNA interference (RNAi) reduced the binding of RAD21 to its sites at *IFNG*, whereas CTCF remained bound (Fig. 3b). This result indicates that CTCF cannot maintain long-range interactions at *IFNG* in the absence of cohesin. Control experiments established that *RAD21* knockdown did not interfere with technical aspects of the 3C methodology (Supplementary Fig. 2) or with cell cycle progression of human CCR5$^+$ T_H1 cells (Supplementary Fig. 5). Notably, basal *INFG* transcript levels were reduced by half (50 ± 13% (mean ± s.d.), $n = 5$, $P < 0.005$, one-sided *t*-test) in RAD21-depleted CCR5$^+$ T_H1 cells (Fig. 3c), and the inducibility of *INFG* transcripts after T-cell activation was reduced to 76 ± 12% of control levels (Fig. 3c, $n = 4$, $P < 0.05$). These effects were selective because the inducibility of *IL2* was not compromised (115 ± 30%, $n = 3$), and the expression of *IGF2*, which is restricted to the paternal allele by cohesin[11], was

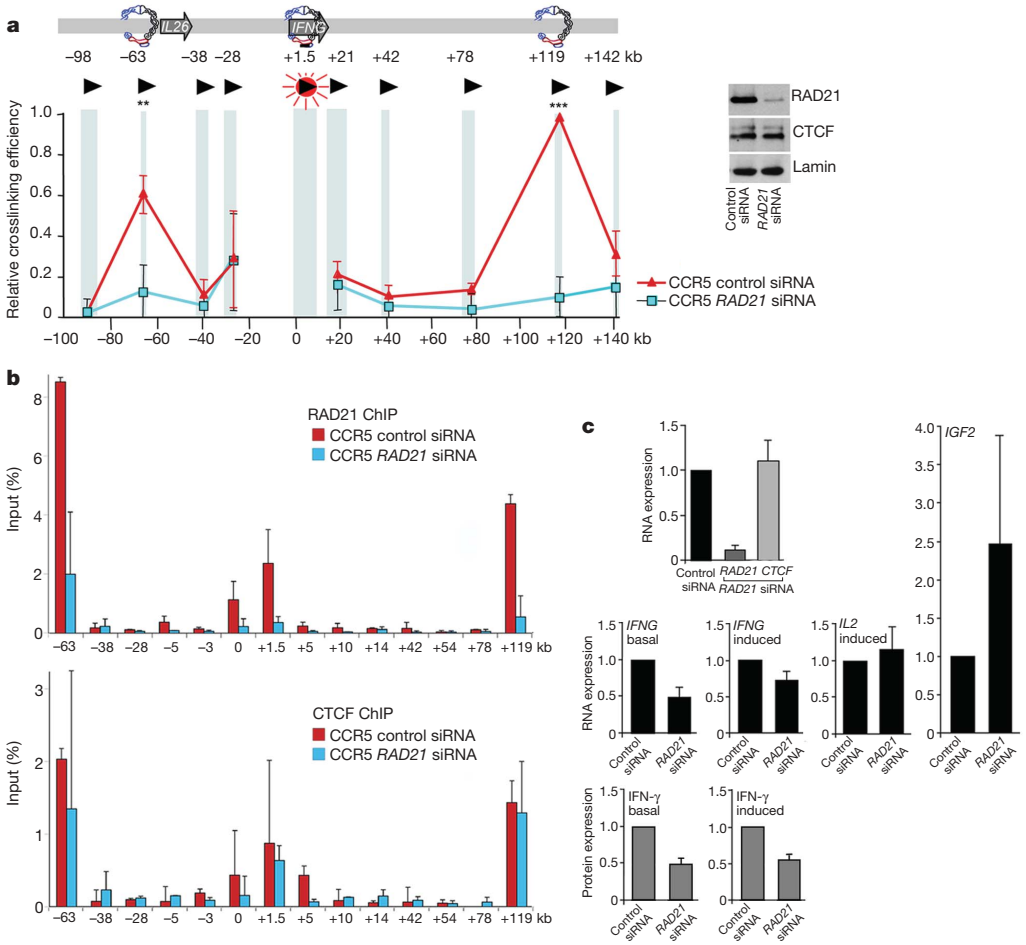


Figure 3 | Long-range chromosomal interactions at *IFNG* require cohesin. **a**, CCR5$^+$ T_H1 cells were transfected with control (red) or *RAD21* (light blue) siRNA oligonucleotides. Long-range interactions at *IFNG* were assessed by 3C (mean ± s.e.m., $n = 4$ for −63, −38 and +119 kb, $n = 2$ for −98, −28, +42, +78 and +142 kb). Data are expressed relative to +119 kb in control-siRNA-treated CCR5$^+$ cells after normalization to mouse *Acta2* in admixed mouse chromatin, which also served to control for technical aspects of the 3C methodology (Supplementary Fig. 2). Normalization to human *ACTA2* gave similar results (Supplementary Fig. 3b). $^{**}P < 0.05$, $^{***}P < 0.005$; Student's *t*-test. The expression of CTCF and RAD21 protein (right) was assessed by immunoblotting 72 h after transfection of CCR5$^+$ T_H1 cells with *RAD21* or control siRNA. Lamin is shown as a control. **b**, ChIP for RAD21 (top) and CTCF (bottom) occupancy of *IFNG* 72 h after transfection of CCR5$^+$ T_H1 cells with control (red) or *RAD21* siRNA (light blue). Primer positions are indicated in kb (mean and s.e.m., $n = 2$). **c**, Real-time RT–PCR analysis of CCR5$^+$ T_H1 cells 72 h after transfection with *RAD21* siRNA or control siRNA. Top panel, *RAD21* and *CTCF* are shown as knockdown controls (mean and s.d., $n = 5$). Middle panels, levels of *IFNG* are shown 1 week (basal, $n = 5$) or 4 h (induced, $n = 4$) after T-cell activation with anti-CD3 and anti-CD28 antibodies. *IL2* is shown as an activation control ($n = 3$) and *IGF2* as a known cohesin target ($n = 3$). Data are normalized to *GAPDH*, *HPRT* and *TBP* expression and shown relative to control-siRNA-transfected CCR5 cells. Bottom panels, enzyme-linked immunosorbent assay (ELISA) for basal ($n = 4$) and activation-induced ($n = 4$) IFN-γ protein secretion. All data in **c** are mean and s.d.

increased (247 ± 148%, $n = 3$, $P < 0.05$) in RAD21-depleted CCR5$^+$ T_H1 cells (Fig. 3c). The basal and inducible expression of IFN-γ protein were also reduced (to 48 ± 9%, $n = 4$, $P < 0.05$, and 57 ± 7%, $n = 4$, $P < 0.05$) by *RAD21* knockdown (Fig. 3c). Hence, cohesin is required for the CCR5$^+$ T_H1 cell-specific conformation of the *IFNG* locus and contributes to its regulation.

Taken together, these data show that the remodelling of the developmentally regulated *IFNG* locus during the differentiation of naive CD4 T cells into T_H1 cells is accompanied by the enhanced association of CTCF and cohesin with a select set of conserved sequence elements. CTCF and cohesin—perhaps in collaboration with other factors or modifications—mediate T_H1-specific long-range interactions between these sites. Recent data provide support for an 'architectural role' of CTCF/cohesin sites in chromosome conformation[28]. Our data show that long-range interactions mediated by cohesin can be dynamically regulated in a developmental context, rather than static in nature. An attractive hypothesis is that CTCF recruits cohesin to specific sites[10–13], and that cohesin in turn mediates chromosomal interactions between these sites *in cis* (this study). This new aspect of cohesin function is probably of general importance in defining local and global genome topology. Cohesin-mediated chromosomal *cis*-interactions provide one possible mechanism for the impact of cohesin on the regulation of *H19*/*IGF2* (refs 9, 11 and J.-M. Peters and A. Murrell, personal communication), of cytokine loci such as *IL4*, *IL5*, *IL13* and *IFNG*[29], MHC class II genes[17] and of other loci[9,11] that depend on CTCF/cohesin for their appropriate expression. We speculate that by constraining the physical conformation of chromatin, cohesin may affect the probability with which gene regulatory elements interact with each other, both negatively and positively, in normal development[3–7] and in disease[7].

METHODS SUMMARY

Naive human CD4 T cells, CCR5$^+$ T_H1 and CRTH2$^+$ T_H2 effector memory cells were isolated by flow cytometry as described[25]. ChIP and real-time PCR, RNA isolation, RT–PCR, siRNA transfection (Amaxa) and immunoblotting have been described[10]. As detailed in the full Methods, human T cells were mixed 5:1 with mouse B3 pre-B cells[10], fixed, and isolated nuclei were digested with HindIII at >80% efficiency at each HindIII site tested. Religated restriction fragments were assayed by TaqMan PCR using custom dual-labelled fluorogenic probes (Sigma). 3C primers for mouse *Acta2*, human *ACTA2* and *IFNG* were designed and tested as detailed in Supplementary Fig. 2. Data were normalized to mouse *Acta2* or human *ACTA2* as indicated. 3D-FISH and distance measurements with Volocity software (Perkin-Elmer/Improvision) have been described[30].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 March; accepted 27 April 2009.
Published online 20 May 2009.

1. Nasmyth, K. & Haering, C. H. The structure and function of SMC and kleisin complexes. *Annu. Rev. Biochem.* **74,** 595–648 (2005).
2. Hirano, T. At the heart of the chromosome: SMC proteins in action. *Nature Rev. Mol. Cell Biol.* **7,** 311–322 (2006).
3. Horsfield, J. A. *et al.* Cohesin-dependent regulation of Runx genes. *Development* **134,** 2639–2649 (2007).
4. Pauli, A. *et al.* Cell-type-specific TEV protease cleavage reveals cohesin functions in *Drosophila* neurons. *Dev. Cell* **14,** 239–251 (2008).
5. Schuldiner, O. *et al. piggyBac*-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. *Dev. Cell* **14,** 227–238 (2008).
6. Zhang, B. *et al.* Mice lacking sister chromatid cohesion protein PDS5B exhibit developmental abnormalities reminiscent of Cornelia de Lange syndrome. *Development* **134,** 3191–3201 (2007).
7. Liu, J. & Krantz, I. D. Cohesin and human disease. *Annu. Rev. Genomics Hum. Genet.* **9,** 303–320 (2008).
8. Hagstrom, K. A. & Meyer, B. J. Condensin and cohesin: more than chromosome compactor and glue. *Nature Rev. Genet.* **4,** 520–534 (2003).
9. Dorsett, D. Roles of the sister chromatid cohesion apparatus in gene expression, development, and human syndromes. *Chromosoma* **116,** 1–13 (2007).
10. Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132,** 422–433 (2008).
11. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451,** 796–801 (2008).
12. Stedman, W. *et al.* Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J.* **27,** 654–666 (2008).
13. Rubio, E. D. *et al.* CTCF physically links cohesin to chromatin. *Proc. Natl Acad. Sci. USA* **105,** 8309–8314 (2008).
14. Wallace, J. A. & Felsenfeld, G. We gather together: insulators and genome organization. *Curr. Opin. Genet. Dev.* **17,** 400–407 (2007).
15. Kurukuti, S. *et al.* CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl Acad. Sci. USA* **103,** 10684–10689 (2006).
16. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes Dev.* **20,** 2349–2354 (2006).
17. Majumder, P., Gomez, J. A., Chadwick, B. P. & Boss, J. M. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J. Exp. Med.* **205,** 785–798 (2008).
18. Gause, M., Schaaf, C. A. & Dorsett, D. Cohesin and CTCF: cooperating to control chromosome conformation? *Bioessays* **30,** 715–718 (2008).
19. Lee, D. U., Avni, O., Chen, L. & Rao, A. A distal enhancer in the interferon-gamma locus revealed by genome sequence comparison. *J. Biol. Chem.* **279,** 4802–4810 (2004).
20. Schoenborn, J. R. *et al.* Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-gamma. *Nature Immunol.* **8,** 732–742 (2007).
21. Hatton, R. D. *et al.* A distal conserved sequence element controls *Ifng* gene expression by T cell and NK cells. *Immunity* **25,** 717–729 (2006).
22. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).
23. Ansel, K. M., Lee, D. U. & Rao, A. An epigenetic view of helper T cell differentiation. *Nature Immunol.* **4,** 616–623 (2003).
24. Jones, B. & Chen, J. Inhibition of IFN-γ transcription by site-specific methylation during T helper cell development. *EMBO J.* **25,** 2443–2452 (2006).
25. Messi, M. *et al.* Memory and flexibility of cytokine gene expression as separable properties of human T_H1 and T_H2 lymphocytes. *Nature Immunol.* **4,** 78–86 (2003).
26. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19,** 24–32 (2009).
27. Dekker, J. The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nature Methods* **3,** 17–21 (2006).
28. Mishiro T. . *et al.* Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J.* doi:10.1038/emboj.2009.81 (in the press).
29. Ribeiro de Almeida, C. *et al.* Critical role for the transcription regulator CCCTC-binding factor in the control of Th2 cytokine expression. *J. Immunol.* **182,** 999–1010 (2009).
30. Hewitt, S. L., High, F. A., Reiner, S. L., Fisher, A. G. & Merkenschlager, M. Nuclear repositioning marks the selective exclusion of lineage-inappropriate transcription factor loci during T helper cell differentiation. *Eur. J. Immunol.* **34,** 3604–3613 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Messi, F. Sallusto and L. Bruno for help and advice with human T cells, L. Aragon for discussions, M. Leleu and M. Spivakov for help with informatics, D. Dormann for help with image analysis and E. Ng, E. O'Connor and P. Hexley for cell sorting. Supported by the Medical Research Council, UK.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.M. (matthias.merkenschlager@csc.mrc.ac.uk).

METHODS

Cell sorting and culture. The use of human cells was approved by the local National Health Service Trust Research Ethics Committee. CD4 T cells were enriched with CD4 microbeads (Miltenyi Biotech). Naive $CD4^+$ $CD45RA^+$ (ImmunoTech, Beckman Coulter) and $CCR5^+$ T_H1 (anti-CCR5, BD Biosciences) or $CRTH2^+$ T_H2 (anti-CD294, Miltenyi Biotech) effector memory cells were isolated by flow cytometry, activated with plate-bound anti-CD3 and anti-CD28 for 4 days, and expanded in media containing IL2 as described[25]. Naive cells were maintained in a non-polarized state by the addition of anti-IL4 and anti-IL12 (eBiosciences). Cytokine expression was induced with plate-bound anti-CD3 and anti-CD28, and evaluated by intracellular staining, real-time RT–PCR or ELISA (IFN-γ Quantikine, R&D Systems).

Real-time RT–PCR. RNA was isolated using RNAbee (Tel-Test) and reverse transcribed. PCR reactions included 2× SYBR PCR Master Mix (Qiagen), 300 nM primers and 2 µl of complementary DNA as a template in 50 µl reaction volume. Cycle conditions were 94 °C for 8 min, 40 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min, followed by plate read. All primers amplified specific cDNAs with at least 95% efficiency. Data were normalized against the average of three housekeeping genes, *GAPDH*, *HPRT* and *TBP*. Primer sequences were (5′-to-3′): *GAPDH* forward (fw) TCTGCTCCTCCTGTTCGACA, reverse (rev) AAAAG CAGCCCTGGTGACC; *HPRT* fw TCCTTGGTCAGGCAGTATAATCC, rev GTCAAGGGCATATCCTACAACAAA; *TBP* fw CCGGCTGTTTAACTTCGC TT, rev TTTTCCTAGAGCATCTCCAGCAC; *IFNG* fw GTTTTGGGTTCTCT TGGCTGTTA, rev ATTATCCGCTACATCTGAATGACCT; *IL4* fw GCGATA TCACCTTACAGGAGATCA, rev GTGTTCTTGGAGGCAGCAAAG; *IL2* fw AACACAGCTACAACTGGAGCATTT, rev AAATGTGAGCATCCTGGTGAG TTT; *IGF2* fw ACACCCTCCAGTTCGTCTGT, rev CGGAAACAGCACTCCT CAAC; *CTCF* fw GGGGAAAATGGAGGAGAAAC, rev GCTCCTCCTCATCC TCATTG; *RAD21* fw TGACTTTGATCAGCCACTGC, rev RT TCCCAACTTC TTCTCTCATGG.

Chromatin immunoprecipitation. Formaldehyde (1%) was added to the culture medium at room temperature, blocked after 10 min with 0.125 M glycine, and cells were washed in cold PBS. Nuclear extracts were prepared and pellets were lysed in 150 mM NaCl, 25 mM Tris-HCl (pH 7.5), 5 mM EDTA, 1% Triton-X, 0.1% SDS, 0.5% deoxycholate for 30 min on ice. Chromatin was sonicated to an average fragment size of 500 bp, centrifuged to pellet debris, incubated at 4 °C with BSA- and salmon sperm-blocked magnetic PGS beads (Dynal) for 2 h, and then overnight with PGS beads coated with rabbit anti-RAD21 (Abcam) or rabbit anti-CTCF (Upstate Biotech). The beads were sequentially washed for 5 min with low-salt buffer (0.1% SDS, 1% Triton-X, 2 mM EDTA, 20 mM Tris-HCl, 150 mM NaCl), high-salt buffer (0.1% SDS, 1% Triton-X, 2 mM EDTA, 20 mM Tris-HCl, 500 mM NaCl) and LiCl buffer (0.25 M LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl). Chromatin was RNase treated and extracted in elution buffer (1% SDS, 0.1 M $NaHCO_3$). DNA crosslinks were reversed at 65 °C overnight, samples were phenol–chloroform extracted and assessed by real-time PCR.

siRNA transfection and chromosome conformation capture. Human T cells were electroporated (Amaxa) with siRNA oligonucleotides and siGLO-red transfection indicator (Dharmacon) or siGLO-red alone using the Amaxa human T cell Nucleofector kit according to the manufacturer's protocol (Amaxa Biosystems). Viable siGLO-red-positive cells were isolated by cell sorting and used for ChIP or 3C experiments. For 3C, 10^7 human T cells were mixed with 2×10^6 mouse B3 pre-B cells[5]. Formaldehyde (1%) was added to the culture medium at room temperature, blocked after 10 min with 0.125 M glycine, and the cells were washed with cold PBS before lysis in 10 mM Tris (pH 7.5), 10 mM NaCl, 5 mM $MgCl_2$, 0.2% NP40 for 20 min on ice. Nuclei were washed and agitated at 37 °C for 1 h in 500 µl of 1.2× NEB buffer 2 with 0.3% SDS, and for 1 h in 1.2× NEB buffer 2 with 2% Triton-X, and overnight with 1,500 units HindIII followed by 30 min at 65 °C in 1.5% SDS. Samples were diluted into 1.1× T4 DNA ligase buffer (NEB) containing 1% Triton-X, and incubated at 37 °C for 1 h. T4 DNA ligase was added (800 units) at 16 °C for 4 h. Crosslinks were reversed at 65 °C overnight and DNA was purified by phenol–chloroform extraction. Real-time PCR was performed in triplicate using TaqMan master mix (Qiagen) and custom dual-labelled fluorogenic probes (Sigma). 3C primers for human *ACTA2* and *IFNG*, as well as mouse *Acta2*, were designed and tested as detailed in Supplementary Fig. 2, and 3C data were normalized to mouse *Acta2* or to human *ACTA2* as indicated. For 3C analysis of cells in the G1 phase of the cell cycle, cells were fixed as described earlier and stained with 50 µg ml^{-1} propidium iodide in the presence of 0.05% NP40 and 1.5 mg ml^{-1} RNaseA for 1 h at room temperature and G1 cells were sorted by flow cytometry. Control cells were left unsorted.

3D-FISH. 3D-FISH and distance measurements with Volocity software (Perkin-Elmer/Improvision) were performed as described[30] using the IFNG BAC CTD2532A1 as a probe.

Untangling the protein web

Researchers have identified thousands of macromolecular interactions within cells. But, as Nathan Blow finds out, joining them up in networks and figuring out how they work still poses a big challenge.

In the spring of 2006, Andrew Emili and Jack Greenblatt from the University of Toronto in Canada and their colleagues published a survey[1] of the global landscape of protein complexes within the yeast *Saccharomyces cerevisiae* in *Nature*. In the same issue, another group of researchers from the drug research company Cellzome in Heidelberg, Germany, also reported[2] on *Saccharomyces* protein complexes. "Those two data sets overlapped nicely, but by no means perfectly," says Mike Tyers, a systems biologist at the University of Edinburgh, UK. "And yet it was essentially the same method and same organism."

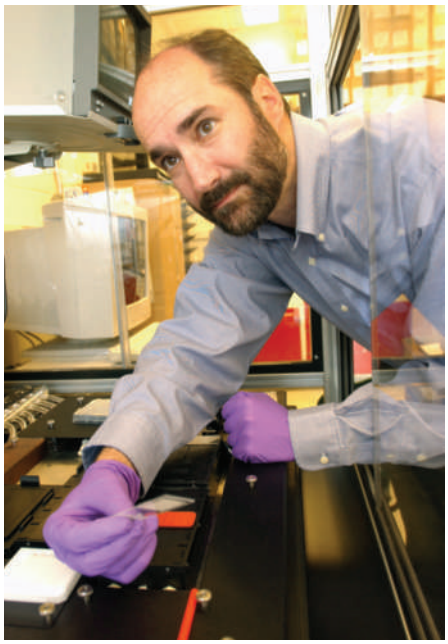
Greenblatt thinks that the two studies highlight something important that is emerging from the current crop of large-scale protein–protein interaction studies. "If you combine data sets you have more information than from any one study alone," he says. This is not to say that one such study is right and the other is wrong: scientists suspect it is more likely that one study often compensates for another's false negatives, revealing true protein interactions that can be missed during a single screen.

"I think the interaction space is very large. Part of the issue is that there is a large range of interaction affinities, and as you start to get down into the weaker interactions those are tougher to detect," says Tyers. He adds that identifying such "moving targets" is not like sequencing DNA, which can be argued to be a more stable target for researchers to aim at.

But the jigsaw pieces are starting to pile up as researchers generate more and more genetic, metabolic and protein-interaction data sets using a diverse array of technologies. This work has been aided in recent years by a number of improved methods and techniques. Add to this recent refinements in computational tools for modelling signalling pathways and it's clear that scientists might be on the cusp of changing the way they look at signalling and information flow in cells.

Embracing diversity

"I think genetic information lays out the blueprints, whereas proteomics is much closer to what is going on in the cell, a molecular manifestation of a phenotype," says Mike Snyder, a biologist at Yale University. When it comes to cataloguing proteins and their interactions, researchers are learning to embrace experimental diversity. "Every approach will usually give an overlapping but distinct set of information," says Snyder. "They all have their strengths and weaknesses."



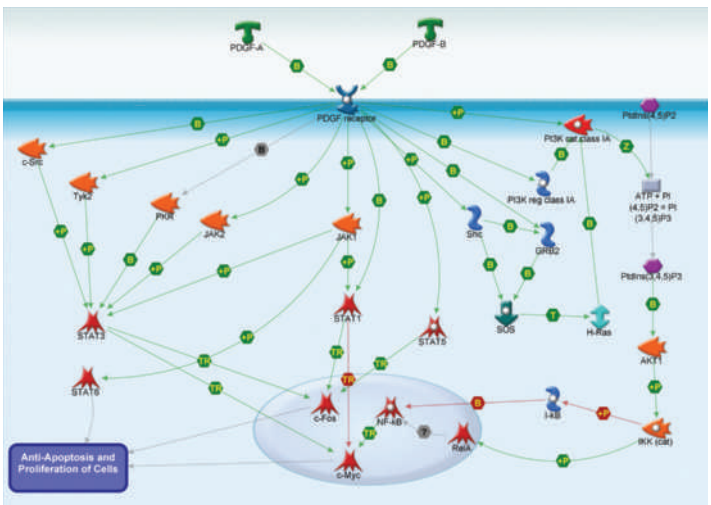
M. SNYDER

Mike Snyder has used protein arrays to explore the yeast interactome.

Tyers and Greenblatt are in a growing group of investigators who are advancing the use of affinity-purification chromatography followed by mass spectrometry to uncover protein interactions in different cell types. In this approach, a protein of interest is tagged with a label that can be used for affinity purification. Although some scientists suspect weaker-interacting protein pairs or transient interactions could be lost during purification, Greenblatt — whose lab relies on tandem affinity purification tags in their purifications — says this is where the use of mass spectrometry helps out. "Mass spectrometry is very sensitive, so even if you lose 90% of the interactor during the affinity purification you can still detect the 10% that is left," he says.

As with the technologies behind protein–protein analysis, researchers are finding that no single labelling tag may be enough to isolate all proteins. Tyers's group recently reinterrogated a section of the yeast proteome using three different tags, each with different properties. "For a number of baits we queried, it made a difference what tag was on it," he says. "Tags can certainly affect the recovery of interactions, consistent with the well-known genetic effects often caused by different tags."

Once a specific protein or protein complex is purified, it is analysed with mass spectrometry. Electro-spray ionization or matrix-assisted laser desorption ionization (MALDI) volatilizes and ionizes peptides, which are analysed on orthogonal or quadrupole time-of-flight (Q-TOF) instruments to identify ions with high mass-to-charge ratio values. Here researchers have benefited greatly from advances by instrument developers. During the American Society for Mass Spectrometry annual conference in Philadelphia, Pennsylvania, in June, Bruker Daltonics of Billerica, Massachusetts, announced its new ultrafleXtreme MALDI TOF/TOF system, and Thermo Fisher Scientific of Waltham, Massachusetts, introduced the LTQ Velos and LTQ Orbitrap Velos devices. Alongside other hardware, such as the Xevo Q-TOF from Waters in Milford, Massachusetts, and the 6500 series of Q-TOF instruments from Agilent Technologies in Santa Clara, California, these machines have improved both the dynamic range and sensitivity of mass analysis; in many cases they also feature integrated upstream separation technologies and improved databases, all of which is making it easier to define a sample's protein composition. For additional detail in the analysis, protein complexes can also be analysed with tandem mass spec-

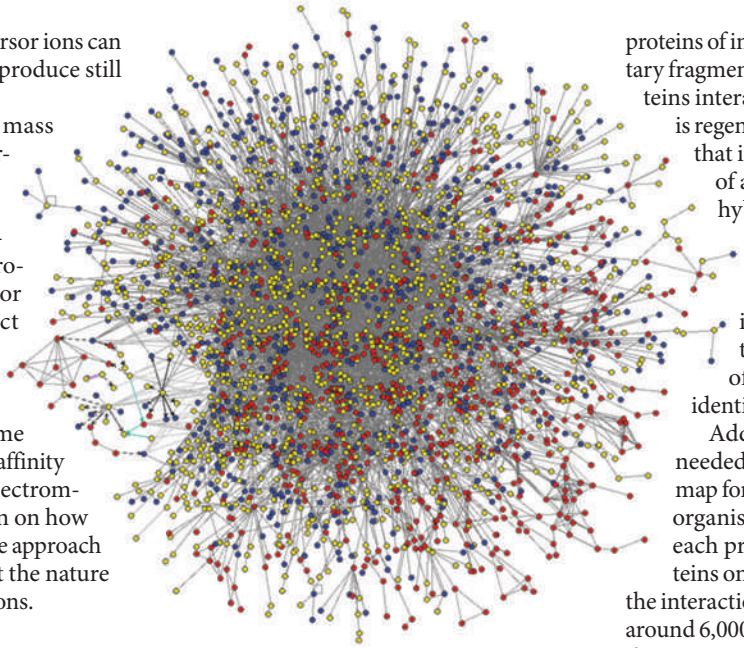


GENEGO

Pathway maps illustrate the complexity of cellular interactions.

trometry, in which selected precursor ions can be smashed into one another to produce still smaller fragments for analysis.

"The big advantage of using mass spectrometry is that it can be performed in a physiological context," says Tyers. Unlike other methods for surveying protein–protein interactions, mass spectrometry can be done on cell lines or even tissue samples, so indirect interactions that depend on more than two proteins or on post-translational protein modifications can be uncovered. Still, some researchers suggest that although affinity purification followed by mass spectrometry gives important information on how proteins interact in complexes, the approach does not reveal everything about the nature and mechanics of those interactions.



M. VIDAL

***Caenorhabditis elegans* interactome map, showing 5,500 protein interactions among 3,000 proteins.**

Yeast shows the way

Binary approaches, such as the yeast two-hybrid assay, can provide different protein interaction information, according to Marc Vidal, a geneticist at the Dana–Farber Cancer Institute in Boston, Massachusetts. Vidal uses the analogy of two football teams facing each other with referees in the middle of the field to explain the differences between the techniques. "The pull-down mass-spectrometry approach will show you the players, referees and field, but not who is passing to whom and in what direction the ball is travelling," he says. "This is where a binary approach comes in."

The yeast two-hybrid assay is arguably the best-known binary approach. It relies on a split transcription factor in which one portion is placed on each of the two proteins being tested for interaction. If the proteins interact, the transcription factor will be regenerated and a reporter gene transcribed, providing a read-out. The assay allows for more the testing of dynamics of protein–protein interactions, such as dissociation rates. "Physical interactions are not everything: you need both edges and arrows to know the dissociation rates as well as other logical aspects of the relationships. Pull-down mass spectrometry is a little short when it comes to those interactions," says Vidal.

The other advantage of the yeast two-hybrid approach is that it presents a more high-throughput solution to studying protein interactions. "The two-hybrid approach is reasonably high-throughput," says Snyder, noting that with robotics a large number of proteins can be tested for potential interactions in a two-by-two format.

Other approaches have also been rising to the surface. "From the probing that we have done, we have picked up interactions that you definitely do not see with other methods," says Snyder of his experience using protein microarrays to explore protein–protein interactions. Protein arrays, which are sold by a number of companies including Invitrogen in Carlsbad, California, RayBiotech in Norcross, Georgia, and R&D Systems in Minneapolis, Minnesota, have not been used as often for large-scale protein-interaction studies as either mass spectrometry or the binary-interaction approaches. "They have had impact in certain areas. Part of the problem is that they have been somewhat expensive, which might be the reason that they have not caught on as much for large-scale studies," says Snyder. Given the potential of protein microarrays to identify unique interactions, he hopes that costs will fall, which could increase their use in large-scale interaction studies.

An orthogonal approach to the yeast two-hybrid assay for detecting protein–protein interactions is the protein-fragment complementation assay (PCA), in which two proteins of interest are attached to complementary fragments of a reporter protein. If the proteins interact with one another the reporter is regenerated providing a direct read-out that is not dependent on transcription of another gene as in the yeast-two-hybrid assay. Steven Michnick and his colleagues used the PCA approach last year[3] to explore the yeast-protein interactome, identifying nearly 2,800 interactions among 1,124 proteins, many of which had not previously been identified by other approaches.

Additional work and tools could be needed to define a complete interaction map for even the most well-characterized organisms. Snyder suspects that in yeast each protein 'sees' about five other proteins on average. But at the moment all of the interactions identified for yeast, which has around 6,000 proteins, add up to far fewer than the potentially 30,000 predicted. "So, there is still a way to go," he says.

Clear pathways

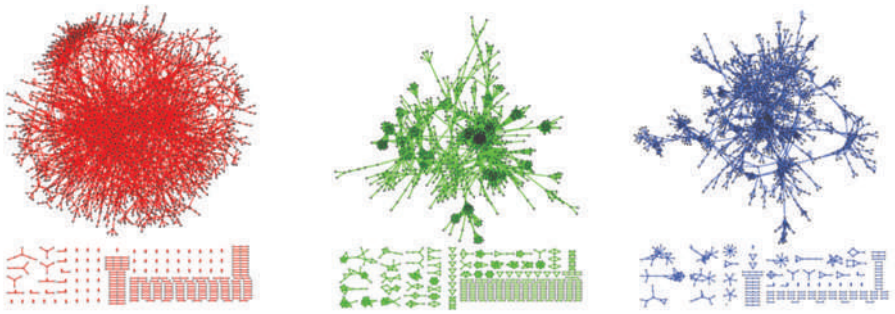
Finding which macromolecules interact is only the first step to figuring out signalling pathways. Researchers also need methods to assemble those interactions into cellular networks, which is where bioinformatics enters the picture. "It is like building a bicycle — you have the wheels, a seat and handlebars, but we provide the steps to put the parts together," says Julie Bryant, vice-president of business development at GeneGo in St Joseph, Michigan, a company specializing in the development of software for cell-signalling and metabolic analysis. GeneGo is not alone here: a growing number of developers are creating tools for the analysis of signalling networks — from those that build model networks based on existing data to systems that use data sets and models to make predictions about the activity of different signalling networks.

"We can take in any kind of experimental data — genomic, proteomic, metabolomic — and overlay them on cell-signalling pathways,"

THERMO FISHER SCIENTIFIC

Advances in mass spectrometry technology are benefiting protein–protein interaction studies.

M. VIDAL



Different approaches for identifying protein–protein interactions often reveal unique information.

says Bryant, describing GeneGo's MetaCore software. Being able to overlay a variety of different experimental data from different sources requires careful database curation, she says. At the moment, GeneGo employs 50 scientists to manually mine and curate published literature for studies on protein interaction, gene expression, metabolism and drugs to expand and update its internal database, which now contains more than 120,000 multi-step interaction pathways, each averaging 11 steps, with information on direction, mechanism and feedback along the pathways, along with direct links to literature evidence.

Literature mining is important for building larger interaction databases, but Bryant says it can be especially difficult if the experimental descriptions underlying the results have not been published. Another problem, according to Vidal, is that researchers sometimes have "sociological" biases in terms of which proteins and interactions they will work on and report. "We have learned a lot about the rules of how macromolecules interact, but when you ask how much of the network we have, or what the size of the interactome of a particular species is, if you only used the literature it would be tough to answer those questions," he says.

Tyers is involved with the publicly funded BioGRID (Biological General Repository for Interaction Datasets) initiative, an internationally curated database of molecular interactions. Three years ago, there was an effort to back-curate all the yeast literature for protein and genetic interactions, but now the database contains protein-interaction data from yeast, worms, flies, plants and even humans along with some genetic-interaction data as well. For Tyers, the goal is to accurately mirror the primary literature and distil it into a format that can be used in network biology. "We make no judgement calls on the method or even, within reason, the quality of the data themselves," he says, giving researchers the opportunity to

PLAYING BY THE RULES

When researchers at Plectix BioSystems in Somerville, Massachusetts, began to use their new Cellucidate software to model the epidermal growth factor receptor pathway, they calculated that there were 10^{33} potential states — including all protein complexes and phosphorylation states — for the system. "This is the kind of complexity that scientists have to grapple with when it comes to cell-signalling networks," says Gordon Webster, vice-president of biology at Plectix.

Although not all these potential states necessarily occur in that pathway, when it comes to creating more manageable models for understanding cell signalling researchers face a difficult question: what interaction data do they use in their models? Although many commercial and public databases still rely heavily on the small-scale protein–protein-interaction studies that appear in peer-reviewed literature, the emergence of high-throughput experimental approaches that generate very large interaction data sets is creating the need for a new set of rules.

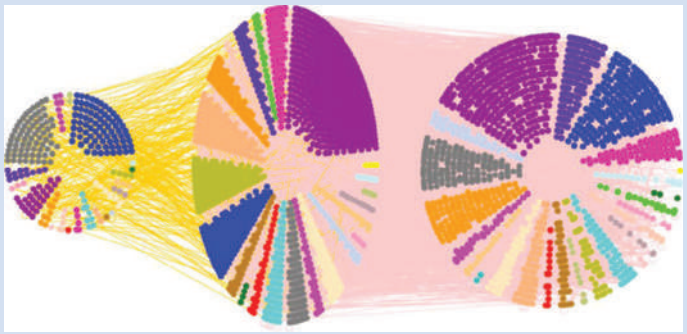
"In practice, what comes out of these high-throughput studies is not a yes/no thing — 'these interact, and these don't' — but in fact they generate a list of interactions and associated probabilities," says Jack Greenblatt from the University of Toronto in Canada. To generate such probabilities for his mass-spectrometry studies, Greenblatt applied a 'gold standard' for protein interactions — a set of protein complexes or interactions in which there is a strong amount of confidence according to the literature — as well as a set of proteins not known to interact with one another as a negative standard. He then tackled the question of whether or not data sets generated by mass spectrometry stacked up against protein-interaction reports seen in peer-reviewed literature.

"What we did in the end was to use the same gold standard to look at the molecular-biology literature," says Greenblatt. After adjusting the cut-off point so that the average confidence score from a high-throughput study matched the confidence score of interactions reported in the literature, he says the interaction data from such studies are no better or worse than what is in the literature.

Marc Vidal, a geneticist at the Dana–Farber Cancer Institute in Boston, Massachusetts, wants to see a similar approach taken with yeast two-hybrid and other binary screens. "Let's roll up our sleeves and decide on a positive and negative gold standard," he says. "But let's also use orthogonal assays to give confidence scores to the interactions."

In January, Vidal and his colleagues published a series of papers[6–9] suggesting the use of new binary interaction assays to build confidence in basic networks produced using yeast two-hybrid data sets. "You say 'OK, this is basic network' and then push that into a framework where all interactions are going to be tested by two or three orthogonal assays. And not only that, but do that under conditions where you have a positive and negative gold standard," says Vidal, adding that the high-scoring interactions can then serve as hypotheses for researchers to test.

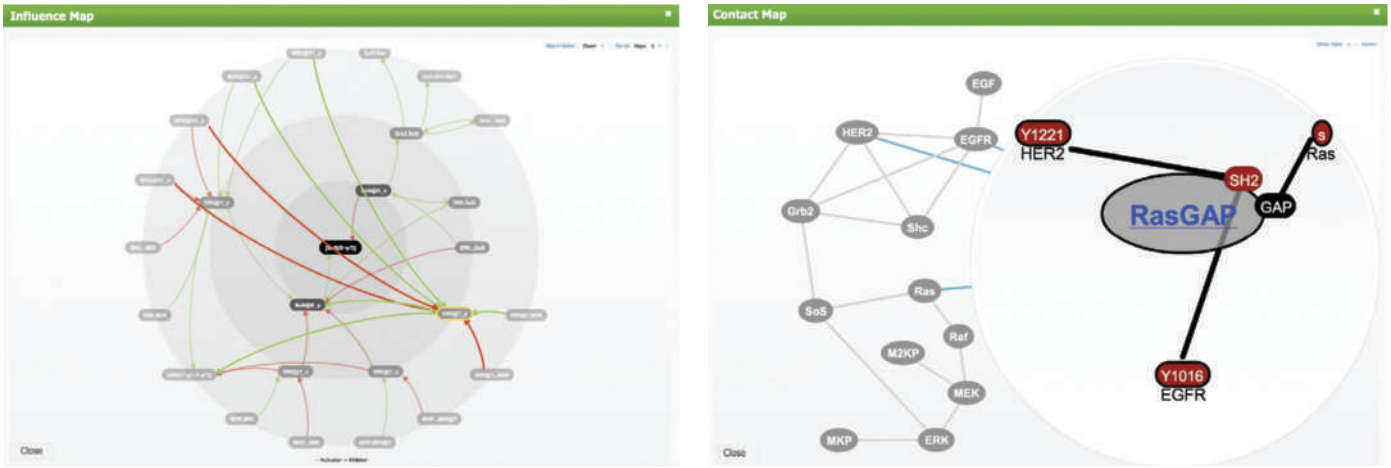
M. TYERS



Graphical representation of the current budding-yeast interaction network.

Whether or not these efforts and standards will lead researchers to rely more on large-scale data sets and mine them more deeply will only be known in time. For some, even with confidence measures, large-scale data sets lack information often found in smaller studies. "This is one of the paradoxes that we find when people talk about systems biology. With technology it is very easy to generate spreadsheets of interaction data, but that alone does not represent any knowledge," says Webster.

But for Greenblatt and others, large-scale data sets represent a starting point for further research efforts. "To me, high-throughput studies are just like the conventional literature," he says, "providing a gold mine for people to dig into." N.B.



PLECTIX BIOSYSTEMS

Cell-signalling software packages allow researchers to model and test cellular interaction networks.

extract the maximum amount of information.

A different angle in modelling signalling networks was recently described by Walter Fontana from Harvard University and his colleagues[4]. It uses sets of rules to define relationships between cellular components instead of the more conventional method of defining specific interactions and species using differential equations. Fontana co-founded a company called Plectix BioSystems in Somerville, Massachusetts, which has employed this approach in a web-based system called Cellucidate.

"The system is represented at a very granular level where the participants are allowed to do *in silico* what they would do in real life," says Paul Edwards, chief executive at Plectix. Imagine the city-building computer game SimCity reworked for complex cellular networks, but here the agents of the cell — proteins and other molecules — are the automata instead of colourful animated people. "In that way the model mirrors the behaviour of the living system it represents: the biology that emerges from our models is the combinatorial expression of all these automata doing their own little thing — just the way it is in the cell," says Gordon Webster, vice-president of biology at Plectix.

Complexity from simplicity

According to Edwards, the advantage of the Cellucidate approach is that a simple set of rules for each agent can result in complex biological behaviour when agents interact during the course of a simulation, unlike modelling in other formats, where the complexity has to be defined before a simulation can be executed. "The level of granularity also means that rules and agents can be easily recycled from one model to another," he notes. Like the GeneGo platform and the BioGRID initiative, Plectix relies on literature mining from various sets of experimental data to create the rules for a model system (see 'Playing by the rules', page 417).

"Mapping all interactions is important, but so is understanding the dynamics behind those interactions," says Snyder. To understand the dynamics of the information flow in cells, researchers not only need more knowledge of protein–protein interaction networks, but they also need to understand protein–DNA interactions, the effects of microRNAs and epigenetic changes on gene expression, and how other macromolecules such as metabolites affect the output of signalling networks. "It is the whole system together that determines the final output and activity," says Snyder.

Vidal thinks that technological improvements — especially in nanotechnology, to generate more data, and microscopy, to explore interaction inside cells, along with increased computer power — are required to push systems biology forward. "Combine all this and you can start to think that maybe some of the information flow can be captured," he says.

But when it comes to figuring out the best way to explore information flow in cells, Tyers jokes that it is like comparing different degrees of infinity. "The interesting point coming out of all these studies is how complex these systems are — the different feedback loops and how they cross-regulate each other and adapt to perturbations are only just becoming apparent," he says. "The simple pathway models are a gross oversimplification of what is actually happening."

Paul Nurse of Rockefeller University in New York wrote about understanding the cell's information flow last year[5]. He noted that "our past successes have led us to underestimate the complexity of living organisms", an oversight that is rapidly disappearing within the world of systems biology and will probably never happen again. ■

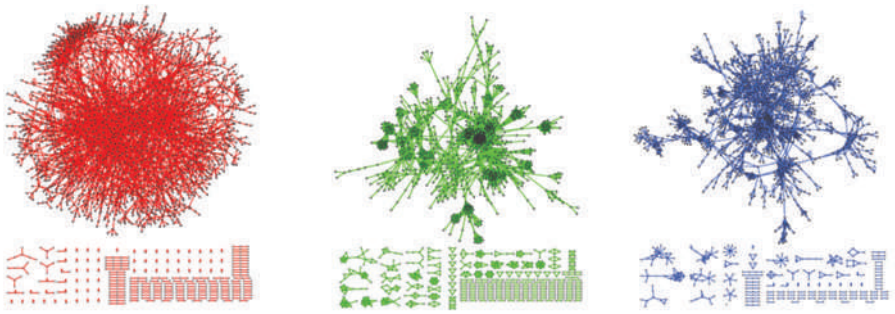
M. TYERS

Mike Tyers uses mass spectrometry to identify protein–protein interactions.

Nathan Blow is technology editor for *Nature* and *Nature Methods*.

1. Krogan, N. J. *et al. Nature* **440,** 637–643 (2006).
2. Gavin, A.-C. *et al. Nature* **440,** 631–636 (2006).
3. Tarassov, K. *et al. Science* **320,** 1465–1470 (2008).
4. Feret, J., Danos, V., Krivine, J., Harmer, R. & Fontana, W. *Proc. Natl Acad. Sci. USA* **106,** 6453–6458 (2009).
5. Nurse, P. *Nature* **454,** 424–426 (2008).
6. Vidal, M. *et al. Nature Methods* **6,** 39–46 (2009).
7. Vidal, M. *et al. Nature Methods* **6,** 47–54 (2009).
8. Vidal, M. *et al. Nature Methods* **6,** 83–90 (2009).
9. Vidal, M. *et al. Nature Methods* **6,** 91–97 (2009).

M. VIDAL



Different approaches for identifying protein–protein interactions often reveal unique information.

says Bryant, describing GeneGo's MetaCore software. Being able to overlay a variety of different experimental data from different sources requires careful database curation, she says. At the moment, GeneGo employs 50 scientists to manually mine and curate published literature for studies on protein interaction, gene expression, metabolism and drugs to expand and update its internal database, which now contains more than 120,000 multi-step interaction pathways, each averaging 11 steps, with information on direction, mechanism and feedback along the pathways, along with direct links to literature evidence.

Literature mining is important for building larger interaction databases, but Bryant says it can be especially difficult if the experimental descriptions underlying the results have not been published. Another problem, according to Vidal, is that researchers sometimes have "sociological" biases in terms of which proteins and interactions they will work on and report. "We have learned a lot about the rules of how macromolecules interact, but when you ask how much of the network we have, or what the size of the interactome of a particular species is, if you only used the literature it would be tough to answer those questions," he says.

Tyers is involved with the publicly funded BioGRID (Biological General Repository for Interaction Datasets) initiative, an internationally curated database of molecular interactions. Three years ago, there was an effort to back-curate all the yeast literature for protein and genetic interactions, but now the database contains protein-interaction data from yeast, worms, flies, plants and even humans along with some genetic-interaction data as well. For Tyers, the goal is to accurately mirror the primary literature and distil it into a format that can be used in network biology. "We make no judgement calls on the method or even, within reason, the quality of the data themselves," he says, giving researchers the opportunity to

PLAYING BY THE RULES

When researchers at Plectix BioSystems in Somerville, Massachusetts, began to use their new Cellucidate software to model the epidermal growth factor receptor pathway, they calculated that there were 10^{33} potential states — including all protein complexes and phosphorylation states — for the system. "This is the kind of complexity that scientists have to grapple with when it comes to cell-signalling networks," says Gordon Webster, vice-president of biology at Plectix.

Although not all these potential states necessarily occur in that pathway, when it comes to creating more manageable models for understanding cell signalling researchers face a difficult question: what interaction data do they use in their models? Although many commercial and public databases still rely heavily on the small-scale protein–protein-interaction studies that appear in peer-reviewed literature, the emergence of high-throughput experimental approaches that generate very large interaction data sets is creating the need for a new set of rules.

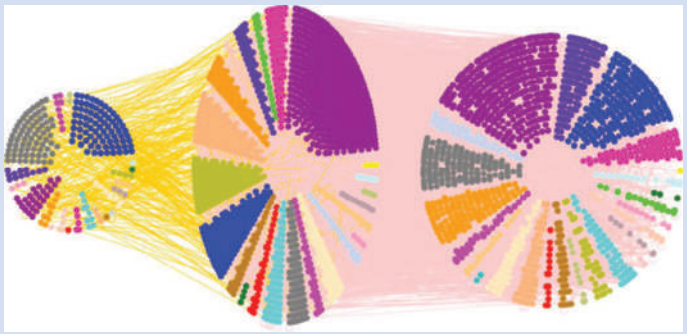
"In practice, what comes out of these high-throughput studies is not a yes/no thing — 'these interact, and these don't' — but in fact they generate a list of interactions and associated probabilities," says Jack Greenblatt from the University of Toronto in Canada. To generate such probabilities for his mass-spectrometry studies, Greenblatt applied a 'gold standard' for protein interactions — a set of protein complexes or interactions in which there is a strong amount of confidence according to the literature — as well as a set of proteins not known to interact with one another as a negative standard. He then tackled the question of whether or not data sets generated by mass spectrometry stacked up against protein-interaction reports seen in peer-reviewed literature.

"What we did in the end was to use the same gold standard to look at the molecular-biology literature," says Greenblatt. After adjusting the cut-off point so that the average confidence score from a high-throughput study matched the confidence score of interactions reported in the literature, he says the interaction data from such studies are no better or worse than what is in the literature.

Marc Vidal, a geneticist at the Dana–Farber Cancer Institute in Boston, Massachusetts, wants to see a similar approach taken with yeast two-hybrid and other binary screens. "Let's roll up our sleeves and decide on a positive and negative gold standard," he says. "But let's also use orthogonal assays to give confidence scores to the interactions."

In January, Vidal and his colleagues published a series of papers[6–9] suggesting the use of new binary interaction assays to build confidence in basic networks produced using yeast two-hybrid data sets. "You say 'OK, this is basic network' and then push that into a framework where all interactions are going to be tested by two or three orthogonal assays. And not only that, but do that under conditions where you have a positive and negative gold standard," says Vidal, adding that the high-scoring interactions can then serve as hypotheses for researchers to test.

M. TYERS



Graphical representation of the current budding-yeast interaction network.

Whether or not these efforts and standards will lead researchers to rely more on large-scale data sets and mine them more deeply will only be known in time. For some, even with confidence measures, large-scale data sets lack information often found in smaller studies. "This is one of the paradoxes that we find when people talk about systems biology. With technology it is very easy to generate spreadsheets of interaction data, but that alone does not represent any knowledge," says Webster.

But for Greenblatt and others, large-scale data sets represent a starting point for further research efforts. "To me, high-throughput studies are just like the conventional literature," he says, "providing a gold mine for people to dig into." N.B.

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
Systems biology			
Assay Designs	Antibodies; ELISA; cell-signalling reagents; pathway maps	Ann Arbor, Michigan	www.assaydesigns.com
Aviva Systems Biology	Antibodies for the exploration of various biological pathways	San Diego, California	www.avivasysbio.com
Cellumen	Reagents, assays and analysis tools for exploring models of cellular disease	Pittsburgh, Pennsylvania	www.cellumen.com
GeneGo	Software for analysis of cell-signalling networks; interaction databases	St Joseph, Michigan	www.genego.com
Genstruct	Computational approach to understanding systems biology and mechanism-based drug discovery	Cambridge, Massachusetts	www.genstruct.com
Optimata	Computational tools for patient physiology and pathology; dynamics of drug interactions	Ramat Gan, Israel	www.optimata.com
Physiomics	Systems-biology approach to development of software and models of cancer	Oxford, UK	www.physiomics-plc.com
Plectix BioSystems	Software called Cellucidate for the analysis of cell-signalling pathways	Somerville, Massachusetts	www.plectix.com
RayBiotech	Protein and antibody arrays; ELISA; tissue microarrays	Norcross, Georgia	www.raybiotech.com
R&D Systems	Protein arrays, along with other molecular- and cell-biology reagents and kits	Minneapolis, Minnesota	www.rndsystems.com
Chromatography			
Advanced Separation Technologies	Analytical and preparative chromatography products; HPLC and chiral chromatography	Bellefonte, Pennsylvania	www.astecusa.com
Alcott Chromatography	HPLC equipment and software, ion chromatography, hydrochromatic chromatographs	Norcross, Georgia	www.alcottchromatography.com
Biotage	Chromatography columns and accessories; FLASH chromatography systems	Uppsala, Sweden	www.biotage.com
Capital HPLC	HPLC and CE columns and accessories	West Lothian, UK	www.capital-hplc.co.uk
Cecil Instruments	Modular HPLC instruments, ion chromatography, UV/visible spectrophotometers	Cambridge, UK	www.cecilinstruments.com
DataApex	Chromatography software	Prague, Czech Republic	www.dataapex.com
GE Healthcare	AKTAExpress automated liquid chromatography platform	Little Chalfont, UK	www.gehealthcare.com
GeneTel	Antibodies for proteomics; antibody services	Madison, Wisconsin	www.genetel-lab.com
Gilson	Solid-phase extraction systems, liquid chromatography	Middleton, Wisconsin	www.gilson.com
Hamilton Company	Automated solid-phase extraction systems	Reno, Nevada	www.hamiltoncompany.com
Phenomenex	Chromatography columns and accessories	Torrence, California	www.phenomenex.com
Tosoh Bioscience	Liquid chromatography columns, products and accessories	Stuttgart, Germany	www.tosohbioscience.com
Mass spectrometers			
Agilent Technologies	Mass-spectrometry instruments, sample preparation and software	Santa Clara, California	www.agilent.com
Applied Biosystems	Mass-spectrometry instruments, reagents and software; ion sources	Foster City, California	www.appliedbiosystems.com
Bruker Daltonics	Mass spectrometers for different applications	Billerica, Massachusetts and Bremen, Germany	www.bdal.de
Cecil Instruments	Spectroscopy and chromatography instruments and systems	Cambridge, UK	www.cecilinstruments.com
Hitachi High Technologies	Liquid-chromatography systems; LC/MS systems	Tokyo, Japan	www.hitachi-hitec.com
Leco	Mass-spectrometry instrumentation and analysis tools	St Joseph, Michigan	www.leco.com
MDS Sciex	Mass-spectrometry systems and technology; HPLC systems	Sunnyvale, California	www.mdssciex.com
Shimadzu	Laboratory instruments including mass spectrometers and data-management systems	Kyoto, Japan	www.shimadzu.com
Thermo Fisher Scientific	High-end analytical instruments, including mass spectrometers, as well as laboratory equipment, software, services, consumables and reagents	Waltham, Massachusetts	www.thermo.com
Varian	Range of analytical instruments including nuclear magnetic-resonance and magnetic-resonance imaging systems, mass-spectrometry instruments, Fourier-transform infrared and X-ray crystallography	Palo Alto, California	www.varianinc.com
Waters	Liquid-chromatography systems; chromatography columns and chemistry products; mass-spectrometry systems; laboratory informatics solutions	Milford, Massachusetts	www.waters.com

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
General			
Active Motif	Tools for studying epigenetic regulation; antibodies; transfection reagents; bioinformatics tools	Carlsbad, California	www.activemotif.com
Alexis Biochemicals	Reagents for molecular- and cell-biology research	Lausanne, Switzerland	www.alexis-corp.com
Attagene	Transcription-factor profiling system; software	Research Triangle Park, North Carolina	www.attagene.com
BD Biosciences	Research reagents, bioimaging systems, instruments	San Diego, California	www.bd.com
BioGenes	Antibody and immunoassay services	Berlin, Germany	www.biogenes.de
Biomol	Services for chemical synthesis, cell culture and antibody production	Hamburg, Germany	www.biomol.de
Bio-Rad	Products, instruments and software for life-sciences research	Hercules, California	www.bio-rad.com
BMG Labtech	Microplate and array readers and handling systems	Offenburg, Germany	www.bmglabtech.com
Cole-Parmer	Instruments and reagents	Vernon Hills, Illinois	www.coleparmer.com
CombiMatrix	Influenza detection systems and microarrays	Mukiteo, Washington	www.combimatrix.com
EMD Biosciences	Calbiochem, Novabiochem and Novagen product lines	San Diego, California	www.emdbiosciences.com
Enzo Life Sciences	Consumables and assays for molecular biology, gene expression and genomic analysis	New York, New York	www.enzo.com
Geneservice	Genomic and proteomic resources; contract services for DNA sequencing, microarray analysis and SNP genotyping	Cambridge, UK	www.geneservice.co.uk
Genovac	Gene synthesis, plasmid purification services; antibody purification and processing	Freiburg, Germany	www.genovac.com
Hamilton Robotics	Automated liquid-handling equipment, automated robotic platforms	Reno, Nevada	www.hamiltonrobotics.com
Harvard Apparatus	Instruments and equipment for electrophysiology and cell biology	Holliston, Massachusetts	www.harvardapparatus.com
Horiba Scientific	Spectroscopy systems and accessories including Raman, atomic emission and UV spectroscopy	Edison, New Jersey	www.horiba.com
Integra Biosciences	Equipment for sterilization, liquid handling, cell culture and sample storage	Baar, Switzerland	www.integra-biosciences.com
Invitrogen	Kits and reagents for molecular-biology, genomics, and cell-biology research	Carlsbad, California	www.invitrogen.com
Irvine Scientific	Defined media for cell-culture applications; custom media services	Santa Ana, California	www.irvinesci.com
Lonza	Molecular-biology reagents and systems; advanced chemical synthesis	Basel, Switzerland	www.lonza.com
Merck	Chemicals, kits and reagents for chemistry, molecular- and cell-biology-related research	Darmstadt, Germany	www.merck.de
Metrohm	Laboratory instrument supplier; consumables	Westbury, New York	www.metrohmusa.com
Molecular Devices	Liquid-handling and microplate-processing equipment; imaging instruments	Sunnyvale, California	www.moleculardevices.com
MP Biomedicals	Reagents and chemicals for research	Aurora, Ohio	www.mpbio.com
Nalgene Nunc	Labware	Rochester, New York	www.nalgenunc.com
New England Biolabs	Molecular-biology-related reagents, kits and enzymes	Ipswich, Massachusetts	www.neb.com ●
Oligoengine	RNAi-synthesis design software; RNAi vectors and clones	Seattle, Washington	www.oligoengine.com
PARC	Software for the analysis of glycan mass-spectrometry data sets	Palo Alto, California	www.parc.com
PerkinElmer	Instruments, reagents and kits for life sciences	Waltham, Massachusetts	las.perkinelmer.com
Pierce Protein	Protein assays, purification, Western blotting	Rockford, Illinois	www.piercenet.com
Premier Biosoft	Software for life-sciences research; SimGlycan analysis tool	Palo Alto, California	www.premierbiosoft.com
Princeton Separations	DNA-purification columns and reagents, fluorescent protein-labelling kits,	Adelphia, New Jersey	www.prinsep.com
Promega	Chemicals for mass-spectrometry analysis	Madison, Wisconsin	www.promega.com
Stratagene	Tools and reagents for molecular biology, genomics and proteomics	La Jolla, California	www.stratagene.com
Takara Bio	Reagents, kits and consumables for molecular biology	Shiga, Japan	www.takara-bio.com
USB	Chemicals and reagents for molecular biology	Cleveland, Ohio	www.usbweb.com
Wako USA	Speciality chemicals supplier; clinical diagnostic reagents	Richmond, Virginia	www.wakousa.com

● see advertisement

Q&A

Taxonomist **Douglas Yanega** has been appointed to the International Commission on Zoological Nomenclature (ICZN) in London.

How did you become interested in taxonomy?
I was always curious about living things. During my senior year of college, I showed my entomology professor specimens of sweat bees nesting in my driveway. Their biology was unknown and my professor encouraged me to document it as an undergraduate thesis project, which evolved into a PhD dissertation.

Were you surprised by the appointment to the ICZN?
Yes. I'm an opinionated individual, and my opinions don't go down too well with large segments of the taxonomic community.

I think the present system by which we name species is not policed effectively and has loopholes and ambiguities. For example, scientific names can be published in journals without peer review. Although that freedom is fine, the reality effectively permits taxonomic vandals to plagiarize others or publish without scientific merit.

The ICZN, a 28-member commission, is now trying to deal with this and other messy issues. I advocate substantial changes to the code, such as provisions for online-only publication, but others prefer the status quo, which makes for a contentious debate.

Have you had a career-defining moment?
I can't pick one moment. The thrill of making new discoveries drives my work. When you work with insects, virtually everything you do is a discovery because so few species have been studied in depth.

I have personally discovered at least 200 species, which have all been gratifying. In my current position I oversee one of the largest collections of insects in North America at the Entomology Research Museum at the University of California, Riverside.

What has been your biggest career challenge?
Entomology as a discipline doesn't get much respect or support. Unfortunately, no one seems to care about the estimated 10 million undocumented life forms on this planet. However, as a career choice, it is phenomenal. I help bring attention to these species. Unfortunately the current California budget turmoil has put my career at risk because I'm considered a staff member rather than a faculty member.

Do you think your ICZN appointment will help your career survive lay-offs?
As it is not a paid position, serving as a commissioner on the ICZN is ultimately a matter of prestige. But my standing at the university depends entirely on the interest of administrators overseeing my position. So I hope they will be sympathetic when assessing the nature and impact of my work. But I worry that entomology doesn't garner much respect.

What do you value most about the scientific process?
That in the long run, reason does win out. As a reasonable person, you like to see the proper outcome. ■

Interview by Virginia Gewin

POSTDOC JOURNAL

Wealth of knowledge?

Several years ago, while still a student at Cornell University in Ithaca, New York, I told a friend that I would be spending six months at the University of Calgary in Alberta. He joked that I was a "mercenary for science". That joke has become a depressing reality.

In the past six years, I've lived in five locations across the United States and Canada. I feel as though I lack roots. I am constantly moving, living out of boxes, leaving friends and adjusting to a new town. It is a solitary, nomadic life.

In those six years I have accrued an enviable hoard of scientific wealth. While a graduate student enrolled at Cornell in a department famous for nonlinear dynamics, I plundered the University of California, Berkeley, for knowledge of metabolic biochemistry and the University of Calgary for knowledge of muscle mechanics. As a postdoc, I raided the University of Vermont in Burlington, escaping with invaluable experience in experimental single-molecule biophysics. Now at Johns Hopkins University in Baltimore, Maryland, I am greedily absorbing the very latest in theoretical biophysics. But even with all this wealth, I just want a place to call home.

In a few months, I begin the next round of faculty applications, and I think I need to change my mindset. Maybe I should stop thinking about how to maximize my knowledge and instead start thinking about where I want to live. I just hope I've amassed enough scientific wealth to afford the rent. ■

Sam Walcott is a postdoc in theoretical biophysics at Johns Hopkins University in Baltimore, Maryland.

IN BRIEF

Energy bill to create jobs?

Calls to create clean-energy jobs were a common refrain at a 7 July US Senate hearing for a bill pending in Congress. The Waxman–Markey bill aims to create jobs, reduce global warming and encourage the use and creation of alternative sources of energy. The House of Representatives passed the bill on 26 June and the Senate is likely to introduce its version this autumn.

Senator Benjamin Cardin (Democrat, Maryland) argued that the priority should be to foster green jobs. "This bill is about keeping jobs and expanding jobs," he said. Senator Kit Bond (Republican, Missouri) countered that the bill would "impose new energy taxes", reduce jobs and create a "bureaucratic nightmare to implement a carbon cap-and-trade programme".

Californian budget woes

More than 300 scientists from the University of California system are protesting proposed state budget cuts that they warn will endanger California's science and technology enterprise.

In a 6 July letter to Republican governor Arnold Schwarzenegger, the scientists caution that the cuts "are likely to destroy" the university's status as "the leading public university in the United States". The cuts would lead to salary reductions, resignations and would jeopardize the university's ability to attract new faculty members, the document says.

A spokeswoman for the governor says the state's US$26-billion budget deficit has created a "worst-case scenario" in which the cuts are unavoidable.

Physicists wanted

A boost in funding and a shift in focus to photon science have prompted expansion at the SLAC National Accelerator Laboratory in Menlo Park, California.

Lisa Mongetta, the lab's human-resources manager, says the lab is seeking more than half-a-dozen physicists, research scientists and engineers, although she warned that the number could change rapidly. In recent weeks SLAC has filled dozens of other positions. The expansion is a reversal of fortunes for the lab, which in January 2008 announced it would shut down one of its colliders as a result of budget cuts (see *Nature* 451, 235; 2008).

Down but not out

Indian outsourcing has been hit badly by the global downturn, but signs of growth remain. **K. S. Jayaraman** investigates.

An economic tsunami is assailing the Indian shores. The country has been hit by the global crisis, financial scandal in the nation, terror attacks in Mumbai and US President Barack Obama's campaign promise to end tax breaks for US companies that send their work overseas. The result: the outsourcing industry, a source of Indian economic prosperity and national pride since the mid-1990s, has been hit hard. Clinical research, aerospace, information technology and 'business process outsourcing' (tasks such as human resources, customer service and accounting duties) have all been affected. But they all have the promise to weather the storm.

Bombings in Mumbai in November 2008 drove away potential investors, and the uncovering of a 71-billion-rupee (US$1.5-billion) fraud operation at Hyderabad-based Satyam Computers, India's fourth-biggest software and service exporter, raised severe credibility issues. Satyam's chairman, Ramalinga Raju, who is now in jail for his part in the offence, confessed to having falsified Satyam's accounts for years to make the company's performance seem better. Meanwhile, tax disincentives introduced by Obama in May mean trouble at Indian outsourcing hot spots, including Gurgaon, Mumbai and Pune in the north and Bangalore, Hyderabad and Chennai in the south.

The information technology (IT) and business-process outsourcing sector may be the biggest victim. On 2 June, nearly 15,000 workers from this sector sent a petition to their trade body, the National Association of Software and Service Companies (NASSCOM), urging employers to explore creative cost-cutting options rather than laying people off.

"India's main attraction for the drug industry — its huge diseased population — will not go away."
— Vasudeo Ginde

Between January 2008 and February 2009 — after the fraud scandal broke — Satyam's workforce shrank by 9,400 people to 41,600, according to Karthik Shekhar, secretary-general of the Union for IT-Enabled Services Professionals. The Delhi and Bangalore offices of IT firm HCL Technologies laid off 450 of their nearly 53,000 employees, Shekhar says. Two other outsourcing companies, Tata Consultancy Services, based in Mumbai, and Wipro Technologies in Bangalore, stopped hiring this year. Wipro chairman Azim Premji told an industry meeting in Bangalore on 13 June that Indian companies may not be able to hire more than one-third of the 35,000 engineers expected to graduate in India next year. Shekhar predicts that at least 50,000 people in the sector will lose their jobs by the end of the year. "Instead of firing them, the companies should try to enhance their skills to enable them to shoulder new tasks," says Shekhar.

Yet some see signs of hope. NASSCOM chairman Pramod Bhasin told the June NASSCOM business-process outsourcing strategy summit in Bangalore that the industry will start to bounce back in early 2010, thanks to a huge domestic market, government-related outsourcing and expanded international reach. Bhasin suggests that companies explore new sectors and regions, focus more on the domestic market and rely less on the United States and Europe. Despite the slowdown, the industry will continue to hire more people than it loses this year; it will have grown by 15% rather than the 30% typical of earlier years, says Bhasin.

A massive case of fraud at Satyam Computers has hit the information-technology industry hard.

N. SEELAM/AFP/GETTY IMAGES

Seeking alternatives

The sagging economy has encouraged young scientists and engineers to consider safer options, such as government jobs. The Indian Space Research Organisation (ISRO), based in Bangalore, received more than 130,000 applications from scientists and engineers this year for just 300 advertised posts, according to ISRO spokesman S. Satish. And positions at the Defence Research & Development Organisation are in high demand. The organization may hire as many as 60 engineering graduates — 12% of the people it recruits this year — from the prestigious Indian Institutes of Technology compared with one or two in a typical year, according to its chief controller of research and development, Prahlada. Traditionally, graduates from the institutions travel abroad or opt for the high salaries of big companies.

But outsourcing companies such as Ananth Technologies in Hyderabad are trying to turn the adversity into an opportunity. Ananth applies geographic information system (GIS) tools to spatial data from satellites — whether providing data to telephone companies, about the best locations for their transmitters or to oil companies about the best route for laying pipelines.

"India is a huge marketplace for GIS, and it would be quite easy for the industry to absorb all its GIS professionals," says Subba Rao Pavuluri, founder and managing director of Ananth. His firm also provides engineering design services for construction companies in the United States. "This is a software-driven service but the job requires structural and mechanical engineers," he says. The company, which had just 40 people on its staff three years ago, now employs 1,100 geologists, geomorphologists and engineers, among others. The firm also provides services to the aerospace industry and expects to benefit from a policy established in 2006 stating that foreign companies selling defence or space systems such as aircraft or satellites to India must buy components and software from the nation worth 30% of the contract. "This is going to be a big boost to service providers like us," says Pavuluri. Many aviation firms have plans to expand their operations in India, which means great opportunity for designing and engineering outsourcing.

Even as they trim costs at home, some US companies continue to invest in India. In May, the industrial supply firm Honeywell International in Morristown, New Jersey, which employs 6,000 engineers at its Bangalore research-and-development site, opened a second, $50-million, facility in Bangalore that will employ 3,000 people in sectors such as aerospace and defence. Still, the future of Indian outsourcing could be with nations such as Japan. In a report entitled *Global 1000 R&D Spenders: India Opportunity Identified*, Zinnov, a management-consulting company based in Bangalore says that Japan will potentially send a lot of its research and development in areas such as electronic equipment and automotive research to India.

ANANTH TECHNOLOGIES LTD.

"The space-components policy is going to be a big boost to service providers like us." — Subba Rao Pavuluri

Clinical research

Contract manufacturing and clinical-services research have become mainstays in India. But the economic downturn has rattled this sector, too, prompting professionals to discuss the sector's status at the Pharmaceutical Outsourcing Congress in Mumbai on 28–30 July. Nearly 80 India-based contract research organizations employ about 22,000 workers, according to Venkat Jasti, an executive committee member of the Association of Contract Research Organizations.

A. V. Rama Rao, the founder, chairman and managing director of Avra Laboratories, a contract research and manufacturing services organization in Hyderabad, says that many other such organizations have had to cease operations owing to a lack of contracts. "Two years ago organic chemists were in great demand," he says. "Today there are no jobs." Hyderabad-based GVK Biosciences reports that it is not hiring at its usual pace. The company currently employs about 800 chemists (150 of whom have PhDs), some 300 informatics specialists and 200 clinical pharmacists. "Contract research organizations that depend on a single client or a single project are the ones most affected," says Jang Bahadur Gupta, vice-president for collaborative research at GVK.

Some such organizations, however, continue to thrive. Vasudeo Ginde, founding director of DiagnoSearch Life Sciences in Mumbai, says that his company has hired 30% more people this year, to make a total of more than 100. Ginde thinks that turnaround is inevitable. "India's main attraction for the drug industry — its huge diseased population — will not go away," he says. Nor, presumably, will the sizable cost savings for offshore companies. Vis Niranjan, president of RxMD, a leading clinical research and development outsourcing company in Chennai, says that the recession in the United States has helped him hire highly skilled physician–scientists of Indian origin from the United States. "I am unable to employ physicians coming out of Indian medical schools," he adds. "They may provide good patient care but not the good science that our industry needs."

For global contract research organizations such as Quintiles Transnational, meanwhile, it's business as usual. "We expect to continue to add jobs because we are experiencing double-digit growth in India," says the firm's spokeswoman Mahrukh Engineer. Quintiles, which employs 23,000 people all over the world, has more than 1,300 employees at four sites in India — Ahmedabad, Bangalore, Mumbai and New Delhi — and will open a fifth in Hyderabad next year. According to Engineer, the firm has not had any contracts cancelled or significant delays in starting up clinical studies, a challenge that other contract-research organizations have faced this year.

K. S. JAYARAMAN

Indian radiologists such as Arjun Kalyanpur will interpret medical images sent in from abroad.

Some medical-outsourcing niches have seen little fall-off. Radiologists at several companies have started reading computed-tomography scans, magnetic-resonance images and other medical images sent by hospitals abroad. A team of radiologists at Teleradiology Solutions in Bangalore, for instance, provides radiology services for more than 70 hospitals in the United States and 10 hospitals in Singapore. The firm also plans to connect to hospitals in Africa. "We report nearly 1,000 scans a day and plan for doubling the number," says Arjun Kalyanpur, a radiologist trained in the United States, who, with his paediatrician wife, founded the company 7 years ago.

More clinical trials in stem-cell therapy could be outsourced after the Indian government proposed new guidelines for researchers and clinicians. Soon, Parliament is expected to adopt the proposed legislation. Once companies have registered their stem-cell trials, researchers would be permitted to run them anywhere in India, irrespective of where the stem cells had been obtained from. So far, the government has cleared only two clinical trials, both by the firm Stempeutics in Bangalore. Its president, Balu Manohar, says that this may open up clinical-trial collaborations with foreign researchers.

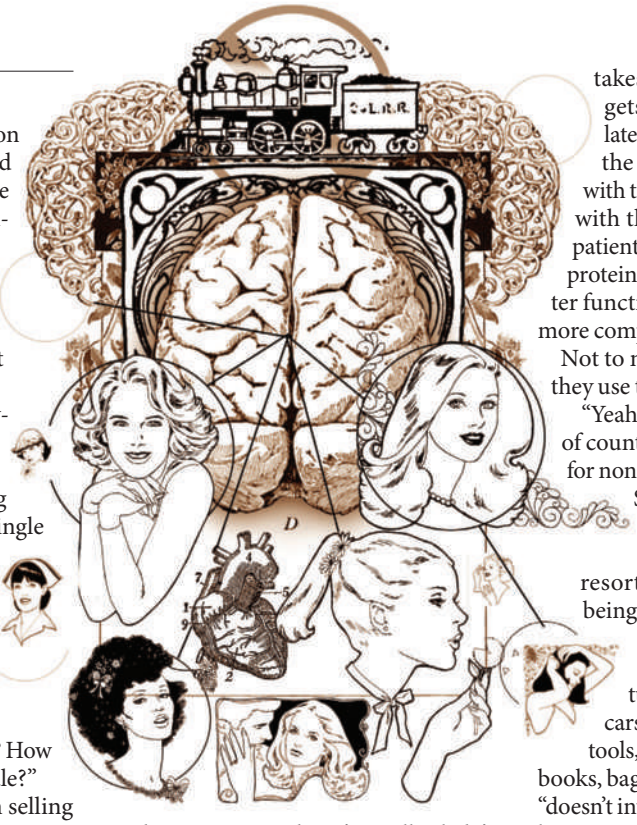
Such trends suggest that outsourcing in India's medical sector still holds ample potential for growth. "We will catch up in no time," says Jasti. "That is what our numbers say." IT professionals, meanwhile, may have a bigger, more challenging recovery ahead. ■

K. S. Jayaraman writes for *Nature* from Bangalore.

Makeover

Off on the wrong track.

James L. Cambias



JACEY

I stopped by Hub Hobbies down on Broad Street, and literally bumped into Mike LaSalle on the way in. He put out an arm to keep me from falling over, then broke into a grin.

"Stephen! How've you been?"

"Oh, fine, fine. I'm here to drool over some of the rolling stock. Did you see that really sweet brass Allegheny they just got in?"

"Nope. I'm selling today, not buying."

"Really? What are you getting rid of?" Mike had some amazing stuff — being both a lawyer and single he could buy pretty much any model he wanted without having to live on ramen and tuna, the way I did.

"The whole thing. I just worked out a consignment deal with the store."

"Everything? Your whole layout? How come? Are you switching to N-scale?"

"I'm cured. Giving it all up. I'm selling the house and getting a condo in the Warehouse District."

"What? How? Why?"

It was starting to rain so we moved inside and stood between the theft detector posts while he told me.

"I signed up with MatchScan. Are you on that?"

"No, I met Stevie before that got big. I never used dating services anyway."

"It's more than a dating service. They actually read your brain. You put your head in this gadget called an fMRI, and then they show you a bunch of images and video clips. It reads how your brain reacts and builds up a personality profile. Tells what kind of person you're attracted to. Then they just match up people with compatible scans. Lots better than filling out questionnaires."

"No lying."

"Exactly. They've scanned thousands of couples to find out what personality types match up. I know some people who got together through MatchScan — they said it was love at first sight."

"But what does that have to do with you selling your trains? Is it expensive?"

He waved his hand dismissively. "I can afford it. No, the trains were the problem. The MatchScan people said it made me an 'outlier'. Any woman I find attractive would hate being with a model railroader."

There are times when I'm really glad I'm gay. Stevie and I both like trains. He keeps the locomotives running and handles the electronics, so I can spend my time painting rolling stock and building terrain. A match made in Man Heaven.

Something puzzled me. "Wait a second — getting rid of the trains won't help. It's all based on that brain scan, right? Even if you don't have a single inch of track in your house that MRI will still tell you like them."

"That's right." Mike looked smug. "That's why I signed up for the advanced programme."

"Which is?"

"It's a step beyond profiling. They actually modify your brain to make your profile more appealing."

"What, they put you in a chair and show you newsreels with your eyelids clamped open?"

Mike's not a movie buff so he looked puzzled. "No, it's not like that. They use NMT. Neural Marker Therapy. Doctors use it to treat people with addictions."

I dimly remembered Stevie talking about it. They give the patient a marker protein which binds to brain cells, then they expose him to material linked to the undesirable trait — alcohol, nicotine, child porn, whatever. I guess for Mike it was playing with trains. A couple of hours later the patient takes an antagonist drug, and then gets exposed to a bunch of unrelated stimuli. Once it's all done, only the part(s) of the brain concerned with the undesirable habit are marked with the protein. Then they give the patient a drug that binds to the marker protein and suppresses neurotransmitter function and blood flow. I expect it's more complicated than it sounds.

Not to mention kind of creepy. "Don't they use that in places like Syria to — "

"Yeah, and Burma, too. Probably a lot of countries. The FDA won't approve it for non-medical purposes here. MatchScan has a referral deal with a clinic in the Bahamas. It's a nice place; there's a little beach resort you can stay at while you're being treated."

"Did it work?"

"Absolutely. All this," he gestured at the shelves of model rail cars, miniature building kits, track, tools, paint, power supplies, reference books, bags of green sawdust, balsa wood, "doesn't interest me at all any more. It's like being in a vegetarian restaurant. Or a jewellery store. Some of it's kind of neat, but I just don't *care*."

"Jesus, Mike. You used to love running trains. You said once you'd been in love with model trains since you were four years old. Is it worth it?"

"Yes," he said, and I was surprised at how intense he sounded. "Yes, it really is worth it. I'm tired of being alone. Of six-month relationships. Of seeing really nice girls going around with these complete *jerks* and wondering why they didn't like me. Now I've got a list of potential matches two pages long. It is absolutely worth it."

There's not much one can say in a situation like that. I wished him luck and turned to go pick through the boxes of train loot he'd left behind. He went out and almost bumped into another customer coming in. She stepped back to let him pass. He nodded absently at her, then headed for his car.

She looked a couple of years younger than Mike, a little chunky but with an intelligent smile. Her eyes lit up when she saw the gear he'd left on the counter. ■

James L. Cambias writes science fiction and designs games. His fiction has appeared in *Fantasy & Science Fiction* and anthologies. He is a partner in Zygote Games, and blogs at Science Made Cool (www.sciencemadecool.com).