6 March 2009 | $10

Science

AAAS

EDITORIAL

NEWS OF THE WEEK

NEWS FOCUS

ORIGINS

LETTERS

BOOKS *ET AL.*

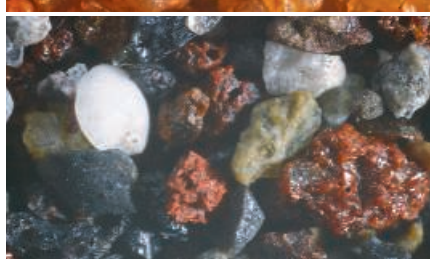
POLICY FORUM

PERSPECTIVES

REVIEW

CONTENTS continued >>

page 1282



page 1291

COVER

Gray and black wolf pups howling in Wolf Park in Battle Ground, Indiana. In North American gray wolves (*Canis lupus*), a mutation that alters coat color from brownish-gray to black, which derives from past hybridization with domestic dogs, is increasing in forested habitats because of positive selection. See page 1339.

Image: Monty Sloan

DEPARTMENTS

Published by AAAS

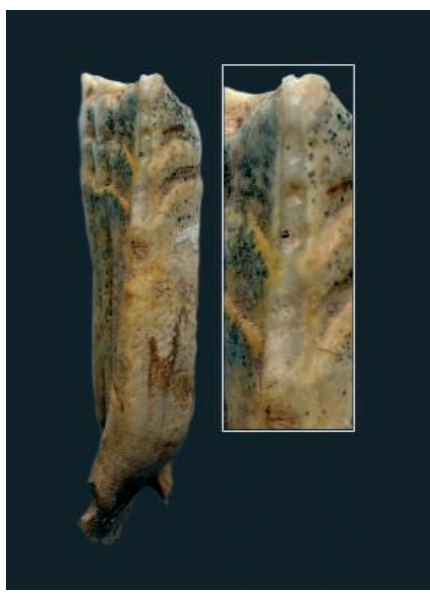
BREVIA

RESEARCH ARTICLES

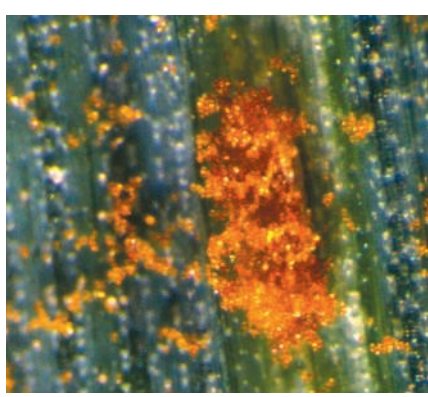
REPORTS

CONTENTS continued >>

page 1302



page 1332



pages 1301, 1357, & 1360

SCIENCE**ONLINE**

SCIENCE**XPRESS**

www.sciencexpress.org

Leucine-Rich Repeat Protein Complex Activates Mosquito Complement in Defense Against *Plasmodium* Parasites

M. Povelones et al.

A family of molecules, apparently unique to mosquitoes, binds to invading parasites and initiates innate immune responses.
10.1126/science.1171400

Rare Variants of *IFIH1*, a Gene Implicated in Antiviral Responses, Protect Against Type 1 Diabetes

S. Nejentsev et al.

Deep resequencing revealed rare alleles that were significantly associated with protection from diabetes.
10.1126/science.1167728

Exomic Sequencing Identifies *PALB2* as a Pancreatic Cancer Susceptibility Gene

S. Jones et al.

Mutations in a gene previously implicated in breast cancer are a contributing factor in hereditary pancreatic cancer.
10.1126/science.1171202

CD24 and Siglec-10 Selectively Repress Tissue Damage–Induced Immune Responses

G.-Y. Chen et al.

A signaling pathway involving an immune protein protects cells against the potentially fatal immune response induced by tissue damage.
10.1126/science.1168988

SCIENCE**NOW**

www.sciencenow.org

Highlights From Our Daily News Coverage

A Better Way to Make Embryonic-Like Stem Cells

"Jumping genes" may improve efficiency and safety of creating therapeutic cells.

Why Nicotine Prefers Brains Over Brawn

Single amino acid change helps explain the first step in smoking addiction.

Drink Up, Energy Hogs

Bottled water guzzles up to 2000 times more energy than tap water.

SCIENCE**SIGNALING**

www.sciencesignaling.org

The Signal Transduction Knowledge Environment

RESEARCH ARTICLE: Odor Coding by a Mammalian Receptor Repertoire

H. Saito et al.

Identification of mammalian olfactory receptor agonists enables development of a predictive model of receptor activation.

PERSPECTIVE: How Perfect Can Protein Interactomes Be?

E. D. Levy et al.

Evolutionary noise in protein-protein interactions implies that not all interactions must be functional.

PODCAST

J. Schlessinger and A. M. VanHook

Autophosphorylation events that control the activity of the FGF receptor are kinetically controlled and disrupted by a mutation associated with glioblastoma.

GLOSSARY

Find out what DARC, SFK, and QTL mean in the world of cell signaling.

SCIENCE**CAREERS**

www.sciencecareers.org/career_magazine

Free Career Resources for Scientists

Redefining Tenure at Medical Schools

C. Wald

Fewer faculty jobs are tenure-track in academic medical research.

Taken for Granted: Shovel-Ready Science

B. L. Benderly

What does the stimulus windfall mean for the scientific labor market?

Can You Be Too Ready for Research Independence?

S. Carpenter

The National Institutes of Health's K99/R00 award is a great success—maybe too great a success.

March 2009 Funding News

J. Fernández

Learn about the latest in research funding, scholarships, fellowships, and internships.

SCIENCE**PODCAST**

www.sciencemag.org/multimedia/podcast

Free Weekly Show

Download the 6 March *Science* Podcast to hear about promoting intellectual discovery through markets, the increasing demands of data-intensive science, news about the U.S. science budget, and more.

ORIGINS**BLOG**

blogs.sciencemag.org/origins

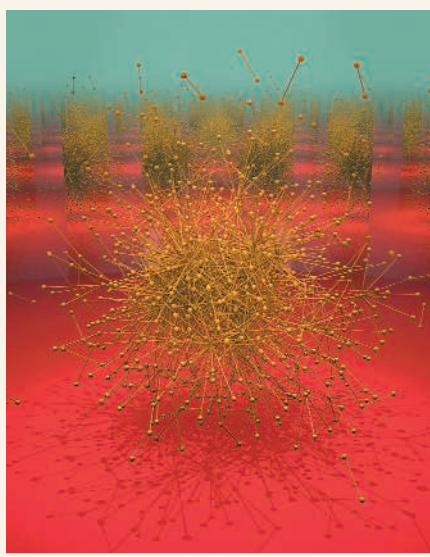
A History of Beginnings

SCIENCE**INSIDER**

blogs.sciencemag.org/scienceinsider

Science Policy News and Analysis

SCIENCE**NOW**
Purely expensive.



SCIENCE**SIGNALING**
Protein interactomes.

CREDITS: (*SCIENCE NOW*) SAMI SARKIS/PHOTOGRAPHER'S CHOICE/GETTY IMAGES; (*SCIENCE SIGNALING*) EMMANUEL D. LEVY/UNIVERSITÉ DE MONTRÉAL

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): $146 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $835; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**

Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $10.00 current issue, $15.00 back issue prepaid includes surface postage; bulk rates on request. **Authorization to photocopy** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $20.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

EDITED BY STELLA HURTLEY

A Lady Bug's Life

The responses of species to environmental change are both ecological and evolutionary. **Harmon *et al.*** (p. 1347; see the Perspective by **Tylianakis**) investigated how species interactions change in an invertebrate predator-prey system and how resistance to heat shock evolves in the prey. During heat-shock disturbance, which reduces aphid abundance, one ladybeetle predator of pea aphids maintained constant predation pressure, while a second ladybeetle species decreased predation pressure. Decreased predation pressure from the second ladybeetle ameliorated the impact of heat shock on aphid abundance. Tolerance to increasing frequency of heat shock evolved rapidly in aphid strains, which was mediated by *Buchnera* endosymbiotic bacteria. Thus, changes in population abundance depend on disturbance-induced changes in the strength of interactions with other species in a food web.

Approaches to HIV Eradication

The disappointing outcomes of recent HIV vaccine trials have led to a call for new approaches to treat and cure HIV. **Richman *et al.*** (p. 1304) review one such option: eradication of HIV infection by antilatency therapies. Antiretroviral therapies are very effective at suppressing viral replication and lowering viral loads; however, a latent reservoir of virus still persists. Therapies targeting latently infected cells aim to eliminate viral infection. The authors describe potential therapeutic targets; discuss the best experimental systems to test antilatency therapies; and suggest strategies to help researchers, clinicians, funding sources, government agencies, and industry to accomplish these goals.

AMPA Receptor Modulator

AMPA receptors are the most commonly found neurotransmitter receptors in the nervous system. They mediate fast glutamatergic signaling in many parts of the brain, and are thought to coassemble with regulatory proteins, of which the TARP (transmembrane AMPA receptor regulatory protein) family is perhaps best known. Using proteomic analysis, immunohistochemistry, and electrophysiology of native AMPA receptors from rat and mouse brain, **Schwenk *et al.*** (p. 1313; see the Perspective by **Tigaret and Choquet**) found that two members of the cornichon family, CNIH-2 and CNIH-3, tightly coassembled with the pore-forming subunits of the AMPA receptors. Indeed, about 70% of the AMPA receptor complexes in the mammalian brain assemble with the cornichon proteins, which promote surface expression of the AMPA receptor complexes and modulate their gating kinetics.

Testing Times

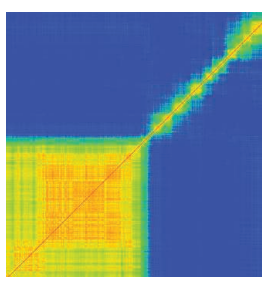
The double-pulsar system J0737-3039A/B has been shown to provide the most precise tests of general relativity. **Deller *et al.*** (p. 1327, published online 5 February; see the Perspective by **D'Amico**) determined the distance to this system by measuring its annual geometric parallax with the Australian Long Baseline Array of radio telescopes. Using this distance, the authors determined that with another 10 years of pulsar-timing observations, PSR J0737-3039A/B will be able to test the validity of general relativity and other theories of gravitation with an accuracy of 0.01%.

Brighter Emission sans Oxygen

Although single-walled carbon nanotubes (SWNTs) emit characteristic emissions based, for example, on their diameters and chirality, the emission for dispersed SWNTs in organic solvents is often quite weak, which is a drawback for many applications, such as sensing. Two potential causes of weak emission can both be caused by poor surface packing of surfactants: The exposure of the SWNT surface can cause reaggregation, facilitating excitation energy transfer between SWNTs, or oxygen doping, which can open up nonradiative decay pathways. **Ju *et al.*** (p. 1319) show that a flavin mononucleotide, modified with a long alkyl chain, packs tightly around SWNTs and, for dispersion in toluene, leads to quantum yields from SWNTs approaching 20%.

Connecting Glass Transitions and Phase Transitions

The glass transition, in which a liquid converts to an amorphous solid, differs from thermodynamic phase transitions in many ways; for example, the temperature at which the transition occurs depends on the rate of heating or cooling. **Hedges *et al.*** (p. 1309, published online 5 February) have performed numerical simulations of a two-component system that interact through a Lennard-Jones potential. They show that unlike an equilibrium phase transition, a first-order transition occurs that is controlled in trajectory space by variables that drive the system out of equilibrium. Two phases coexist heterogeneously: a mobile phase that is relaxed in finite time and an immobile phase.



X-ray Probing of Magnetism

Correlated electron systems can give rise to a number of exotic electronic and magnetic phases. While probing of the magnetic structure has usually been done with neutron scattering, recent work has shown that x-rays can also be used to probe the magnetism. **Kim *et al.*** (p. 1329) use resonant x-ray scattering to determine fully the magnetic structure of the transition metal oxide, Sr_2IrO_4, and unravel the quantum phase of the material. The technique should be applicable to probe the ground states of many other complex electronic systems.

Continued on page 1263

CREDITS (TOP TO BOTTOM): ANTHONY R. IVES; HEDGES *ET AL.*

Published by AAAS

Continued from page 1261

Ancient Domestication of Horses

The domestication of horses changed the course of human history. The Botai Culture, present on the Eurasian Steppe from 5700 to 5100 years ago, may have been a location of at least one domestication event. The Botai Culture is known to have interacted with local horses in Northern Kazakhstan but whether the animals were husbanded or hunted has been unclear. **Outram *et al.*** (p. 1332) provide three lines of evidence suggesting that the horses were indeed domesticated: First, isotope data from fragments of pottery indicates that the horses were milked during the summer months; second, identification of bit and bridle wear suggests that horses at Botai were harnessed and ridden; and third, the Botai horses resemble later, more slender domestic Bronze Age horses, and not the sturdier wild horses in the region.

Black Wolf of the Family

Black coat color is fairly common in domestic dogs but relatively rare in their close relatives, wolves and coyotes. **Anderson *et al*.** (p. 1339, published online 5 February; see the cover) analyzed the *K* locus associated with black coat color in dogs, wolves, and coyotes and document the introgression of a domesticated allele into a wild species. The K^B allele leads to a dominant black coat in dogs, wolves, and coyotes. Against the common flow of genes from wild to domesticated animals, the K^B mutation originally occurred in dogs and later introgressed into wolves and coyotes by hybridization. Furthermore, the relatively high frequency of black wolves in Yellowstone probably reflects positive selection for the K^B allele in the wild.

Rot-Eat-Wheat World

Fungal diseases have a large impact on the yield of wheat crops. Some strains of wheat carry genes that make them naturally resistant. Breeders have introduced these genes into other wheat strains to share the valuable disease-resistance traits. Two papers now identify the molecular basis of some of these traits (see the Perspective by **Kliebenstein and Rowe**). **Krattinger *et al.*** (p. 1360, published online 19 February) used comparisons with other grass genomes to clone the gene *Lr34*, which resembles adenosine triphosphate–binding cassette transporters. **Fu *et al.*** (p. 1357, published online 19 February) similarly cloned the gene *Yr36*, which resembles a kinase. Each gene provides, to a greater or lesser extent, resistance to certain fungal pathogens that the fungi cannot overcome by genetic variation. Such durable resistance has tremendous agricultural value.

Divisive Tension

Chromosome attachments to the mitotic spindle are regulated by tension across the centromere, which ensures accurate chromosome segregation. Spindle microtubules can bind chromosomes at the kinetochore in different configurations, but only bi-oriented attachments can support accurate chromosome segregation. Tension across the centromere distinguishes correct and incorrect attachments, but how tension is sensed to regulate attachments is unclear. The mitotic kinase Aurora B is a key regulator of kinetochore-microtubule attachments, and phosphorylation of Aurora B substrates at kinetochores reduces microtubule affinity. **Liu *et al.*** (p. 1350, published online 15 January) tested to see if a tension-sensitive mechanism regulates Aurora B activity. Instead their results suggest that Aurora B kinase activity is not directly regulated by tension, but that phosphorylation of an Aurora B substrate at the kinetochore depends on its distance from the kinase at the inner centromere, and that it is this distance that is affected by tension.

Form → Function

How do proteins localize to particular sites within the bacterial cell? In many systems, biochemical features of the proteins and their target membranes specify particular localization signals. Now **Ramamurthi *et al.*** (p. 1354) report that a geometric cue—positive (convex) membrane curvature—dictates the localization of a small peripheral membrane protein during spore formation in *Bacillus subtilis*. Membrane curvature may represent a general cue for protein localization in bacteria in the absence of biochemically identifiable signals.

CREDIT: LYNN EPSTEIN

Published by AAAS

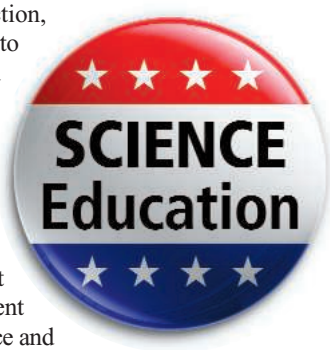
Leon M. Lederman is a resident scholar at the Illinois Mathematics and Science Academy, Aurora, IL.

Shirley M. Malcom is director of Education and Human Resources, AAAS, Washington DC.

The Next Campaign

THE U.S. PRESIDENTIAL ELECTION IS OVER; THE INAUGURATION HAS COME AND GONE. Now come the challenges of governing. President Obama, his leadership team, and the 111th Congress face daunting problems. Among the most stubborn in the United States are those related to providing a world-class education to all children, updating the knowledge and skills of their parents, and preparing all to face the threats and opportunities of the 21st century. The science community must take advantage of growing public dissatisfaction with the current education system and ask how science teaching and learning can be transformed. In short, scientists must mount the next campaign.

First, scientists must help the adult public develop a clear understanding of what science is and what science education should be: a way of knowing about the world based on evidence and logical analysis. A consensus is emerging across science, philanthropic, and political communities about what our goals must be. Surveys from the nonprofit organization Public Agenda indicate that adults realize something is amiss in science instruction, though not necessarily for their children. That unease can provide license to introduce a common core of science education standards across the United States; to invest as a nation in assessments that measure the science understandings and abilities that every child needs for success in today's global economy; and to mount vigorous efforts to recruit, train, and retain highly effective science teachers.



If the United States is to address its many challenges—including developing a "green economy" and alternative sources of energy that lessen climate impacts and address global climate change as perhaps the greatest threat we face as a planet—science education must move to center stage. President Obama has recognized the challenges of recruiting and compensating science and mathematics teachers and of making science, as in post-Sputnik years, a more integral, inspirational part of national culture. Now we must turn such far-reaching national vision into grassroots reality.

Although the need for creating a talent base for careers and science-savvy citizenry for the 21st century is national (indeed, global), most of the action to accomplish it is local. Messaging is needed to explain past scientific contributions to economic growth and science's traditional role as an engine of change. Campaign workers must be recruited, trained, and deployed from the ranks of academic scientists, business and industry, patient advocacy groups, and unions. Resources must be marshaled from local philanthropy, understanding the need to build public awareness and support for an agenda for change. Endorsements must be enlisted from the media, but also from community groups, parent-teacher associations, and senior citizens of all political persuasions. The science community itself has much to offer here. Think of expanding by orders of magnitude the numbers of retired scientists and engineers currently working with teachers and students in schools or in museums and science centers as docents; scientists serving on statewide education commissions and councils being convened by many governors; and in local communities, scientists advocating for science education to mayors, school boards, and superintendents and supporting implementation with principals, teachers, and students.

In this campaign, scientists will need to assess assets nationally and locally, noting what has worked elsewhere, such as in Nordic countries where science literacy is high. Unlike some countries, the United States has no overarching education ministry. Instead, a strategy for transforming science education must unite the interests and actions of 50 states, 15,000 local districts, 3500 colleges and universities, and countless informal science organizations that constitute the U.S. educational system. Federal incentives must foster collaboration in gathering and sharing evidence from experimentation informed by research, guided by core national science education standards. States could begin with decades-old lessons learned by the American Association for the Advancement of Science (AAAS) and the National Academies, and by states such as Massachusetts or countries such as Singapore, both of which have performed well on international assessments.

Although scientists are generally more comfortable presenting facts than going on the stump, the next campaign demands that they do both. **– Leon M. Lederman and Shirley M. Malcom**

10.1126/science.1169820

Published by AAAS

PHYSIOLOGY

Mad Dogs and Englishmen

It is well established that organisms respond to climate change by adapting, by shifting their geographic distribution, or—in the unluckier cases—by becoming extinct. Most models of responses to climate change focus on changes in distribution, usually based on "climatic envelope" concepts: that is, the range of climatic conditions that a species can endure. For cold-blooded animals such as reptiles and insects, a missing element from such models has been the ability of these organisms to regulate their body temperature by behavioral means, such as not going out in the midday sun, which has the potential to buffer their geographic response to changing climate. By including such thermoregulatory behavior in biophysical models of temperature responses of Australian ectotherms, Kearney *et al.* show that the challenge for many such species in a warming world will be to stay cool. If moving with the climate is not an option, which it will not be for many tropical organisms, survival will depend on factors such as the availability of shade and the ability of ectotherm species to alter their seasonal or daily patterns of activity. — AMS

Proc. Natl. Acad. Sci. U.S.A. **106**, 10.1073/pnas.0808913106 (2009).

PALEONTOLOGY

Slipping Through the Cracks

The fossil record shows that, at least in the marine realm, unusually small taxa predominate in the aftermath of mass extinctions. In extreme cases such as after the end-Cretaceous extinction, 65 million years ago, it took several millions of years for diversity to recover. This pattern poses a potential bias in assessing the impact of the extinction, as small species tend to be more difficult to preserve in the fossil record. Sessa *et al.* evaluate this bias by comparing fossils in lithified and unlithified sediments across the Cretaceous-Paleocene boundary from thick sections in the Gulf of Mexico. Their data show that small fossils are indeed lost (perhaps by dissolution) from the fossil record as sediments are compacted and form rocks—by a factor of up to 2.4. This process, however is not systematic in time, and lithified sediments tend to predominate in the Paleocene after the mass extinction. Thus, the pattern of a delayed recovery may be partly exaggerated by the sediment record, as might enhanced diversity before the extinction. This bias decreases further back in time, as unlithified sediments become scarce, but illustrates the inherent selection of the fossil record. — BH

Geology **37**, 115 (2009).

CHEMISTRY

Zeolites Left and Right

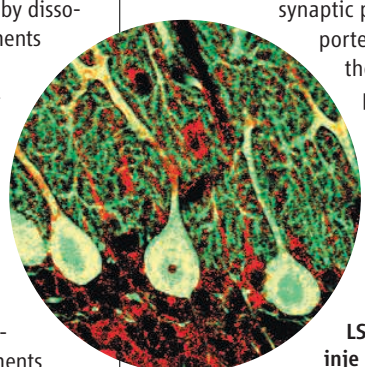
Formally, a chiral substance is one that changes its appearance upon rotation about any axis, followed by reflection in a perpendicular plane. This property can arise from arrangements as simple as four elementally distinct atoms bonding to a central carbon, or as complex as hundreds of atoms mutually adopting slight positional shifts in an extended crystal lattice. From a practical standpoint, however, most chirally selective processes in chemical syntheses and separations rely on the influence of purified small molecules, whether free in solution or tethered to an achiral solid support. Dryzun *et al.* have now analyzed the parameters of previously characterized solid zeolite structures and uncovered 20 bulk crystal lattices that manifest extended chirality. Moreover, calorimetry revealed that several such zeolites distinguish between the enantiomers (mirror-image isomers) of the amino acid histidine. Given the widespread role of zeolites in catalysis, the authors envision potential for enantioselective extensions of their current applications. — JSY

J. Mater. Chem. **19**, 10.1039/b817497k (2009).

NEUROSCIENCE

Putting mRNA in Its Place

Many cell types, including epithelial cells and neurons, have structurally and functionally separate cytoplasmic compartments, which enable them to perform distinct functions. This cellular polarization can be established by blocking the immediate translation of some mRNAs, allocating them instead to specific locations in the cell. Translation can then be initiated as needed, thus ensuring that the proteins are synthesized only in the selected location at the required time. In neurons, mRNAs coding for synaptic proteins are transported to dendrites, where they are translated into proteins upon synaptic stimulation. Di Penta *et al.* have found that the factor LSm1, which has previously been shown to participate in the degrada-



LSm1 (yellow) in Purkinje neuron dendrites.

*Helen Pickersgill is a locum editor in *Science*'s editorial department.

CREDITS (TOP TO BOTTOM): GERRY PEARCE/ALAMY; DI PENTA *ET AL.*, *J. CELL BIOL.* **184**, 423 (2009)

tion of mRNAs, is involved. They identified a protein-mRNA complex containing LSm1 and the nuclear cap-binding protein CBP80, which recognizes the modified nucleotide that caps the 5′ end of the mRNA. The authors propose that this complex assembles in the nucleus, which would indicate that the mechanisms that regulate localized protein synthesis come into play soon after transcription, and in neurons this protein-mRNA complex was targeted to dendritic spines. These alternative functions for an mRNA degradation factor and a nuclear cap-binding protein in mRNA transport reveal molecular links between apparently diverse cellular processes. — HP*

J. Cell Biol. **184**, 423 (2009).

ATMOSPHERIC SCIENCE

Ozone in the Gulf

Ozone is a secondary pollutant formed in the interactions of reactive carbon compounds, nitrogen oxides, and ultraviolet sunlight, with large potential impacts on the health of both animals and plants. Tropospheric ozone is a naturally occurring

Dubai.

atmospheric species, but the highest concentrations are found in sunny areas with abundant air pollution, such as the Los Angeles basin in California. Lelieveld *et al.* analyze another area in which ozone pollution is severe—the Persian Gulf region. The locale possesses all of the ingredients for ozone production in abundance: intense air pollution (arriving from long distances as well as originating from strong local anthropogenic emissions), unusually vigorous stratospheric-tropospheric exchange, sparse deep convective mixing and precipitation, and copious sunlight. Using an atmospheric chemistry model, air-quality measurements, and support from satellite data, the authors predict and observe intense ozone pollution during the period from 1996 to 2006. Reducing the levels of ozone pollution in this region will depend largely on both decreasing the amounts of long-range pollution advected to the region and reducing local sources of nitrogen oxides. — HJS

Atmos. Chem. Phys. **9**, 1393 (2009).

CREDIT: 1APIX/ALAMY

PSYCHOLOGY

Seeking a Second Opinion

A marketplace—whether for ideas or goods and services—provides a remarkably efficient means of information exchange, and prediction markets yield astonishingly accurate forecasts of election outcomes, Oscar winners, and so forth. But how might a single individual, lacking access to the wisdom of the crowds, attempt to improve his or her best guess? Herzog and Hertwig offer one prescription based on the Hegelian dialectic: After making the first estimate, consider the reasons and assumptions underpinning that estimate (and how they might be off target), and then formulate a new, second estimate. They tested the efficacy of this method by asking university students to date a collection of 40 historical events covering the past four centuries and tabulating the average of repeated guesses from the same individual, which is analogous to standard reliability sampling, to the average (synthesis) of the dialectical guesses (thesis and antithesis). They found that the improvement in accuracy (in years) over the first estimate was twice as large for the dialectical average than for the repeat average, although averaging the first estimates from two random individuals worked better still. — GJC

Psychol. Sci. **20**, 231 (2009).

NEUROSCIENCE

Getting to the Right Place

The six-subunit protein complex known as Elongator is known for its role in the acetylation of histone H3 and its association with actively transcribed regions of the genome. However, how this function fits with observations linking genetic disruption of the scaffold subunit Elp1 in humans to a defective development of autonomic and sensory neurons has been puzzling. Creppe *et al.* show that Elongator—particularly the acetyltransferase subunit Elp3—participates in the development of neuronal cells. In the brain, the proliferation of precursor cells generates neuronal cells, but these neurons find their place and adopt their shape only after they have finished their last cell division. For the accurate development and integration of cortical projection neurons, the newly born cells must migrate to the cortex and then form extensive networks of connections with other neurons. The authors suggest that the Elongator complex supports neuronal migration and branching by acetylating tubulin, a component of the cytoskeletal network. — PJH

Cell **136**, 551 (2009).

www.sciencemag.org

Science

1200 New York Avenue, NW
Washington, DC 20005
Editorial: 202-326-6550, FAX 202-289-7562
News: 202-326-6581, FAX 202-371-9227
Bateman House, 82-88 Hills Road
Cambridge, UK CB2 1LQ
+44 (0) 1223 326500, FAX +44 (0) 1223 326501

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

Institutional Site Licenses please call 202-326-6755 for any questions or information

Reprints: Author Inquiries 800-635-7181
Commercial Inquiries 803-359-4578

Permissions 202-326-7074, FAX 202-682-0816

Member Benefits AAAS/Barnes&Noble.com bookstore www.aaas.org/bn; AAAS Online Store www.apisource.com/aaas/ code MKB6; AAAS Travels: Betchart Expeditions 800-252-4910; Apple Store www.apple/eppstore/aaas; Bank of America MasterCard 1-800-833-6262 priority code FAA3YU; Cold Spring Harbor Laboratory Press Publications www.cshlpress.com/affiliates/aaas.htm; GEICO Auto Insurance www.geico.com/landingpage/go51.htm?logo=17624; Hertz 800-654-2200 CDP#343457; Office Depot https://bsd.officedepot.com/portalLogin.do; Seabury & Smith Life Insurance 800-424-9883; Subaru VIP Program 202-326-6417; VIP Moving Services www.vipmayflower.com/domestic/index.html; Other Benefits: AAAS Member Services 202-326-6417 or www.aaasmember.org.

science_editors@aaas.org	(for general editorial queries)
science_letters@aaas.org	(for queries about letters)
science_reviews@aaas.org	(for returning manuscript reviews)
science_bookrevs@aaas.org	(for book review queries)

Published by the American Association for the Advancement of Science (AAAS), *Science* serves its readers as a forum for the presentation and discussion of important issues related to the advancement of science, including the presentation of minority or conflicting points of view, rather than by publishing only material on which a consensus has been reached. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

AAAS was founded in 1848 and incorporated in 1874. Its mission is to advance science, engineering, and innovation throughout the world for the benefit of all people. The goals of the association are to: enhance communication among scientists, engineers, and the public; promote and defend the integrity of science and its use; strengthen support for the science and technology enterprise; provide a voice for science on societal issues; promote the responsible use of science in public policy; strengthen and diversify the science and technology workforce; foster education in science and technology for everyone; increase public engagement with science and technology; and advance international cooperation in science.

Information for Authors
See pages 807 and 808 of the 6 February 2009 issue or access www.sciencemag.org/about/authors

CREDITS (TOP TO BOTTOM): STEPHAN TZOTZIS; SOURCE: GLEICK AND COOLEY/CALIFORNIA ENERGY COMMISSION; FREDERICK DELFIN/MAX PLANCK INSTITUTE FOR EVOLUTIONARY ANTHROPOLOGY

Black Death Diagnosis

Bubonic plague, which first broke out in Europe in 1347, killed some 75 million Europeans before it was brought under control in the early 19th century.

Archaeologists continue to unearth mass graves. Up till now, scientists have looked for the DNA of the plague organism, *Yersinia pestis*, to determine if people died of the plague. But the method is unreliable and in medical testing, it's recently been replaced by a cheaper, faster, and more accurate rapid diagnostic test (RDT) that reveals the presence of a protein specific to *Y. pestis*.

Anthropologist Raffaella Bianucci, working at Turin University in Italy, decided to apply the RDT to the remains of four 17th century nuns and two priests from different parts of France who, records show, took care of the poor and sick. Bianucci confirms in the March issue of the *Journal of Archaeological Science* that all were victims of the Black Death.

"This is very important research," says mummy expert Bob Brier of Long Island University in Brookville, New York. There are conflicting views of how widespread the plague was in Europe. "Now we don't have to rely just on historical sources" to trace its pathways, he says. Bianucci plans to use the test at more sites. Next up will be Lazzaretto Vecchio, the plague infirmary founded on an island in Venice, Italy, in 1403, where 1500 skeletons await their checkup.

Back to the Tap

Bottled water can use up to 2000 times as much energy as tap water, according to a study in *Environmental Research Letters*.

Despite rising concerns about "green" consumption, Americans are drinking 70% more bottled water than they did in 2001, according to Peter Gleick and Heather Cooley of the Pacific Institute in Oakland, California. The pair have added up the energy use in every step of the process. Bottle manufacture alone uses 50 million barrels of oil a year globally, they estimate. Transport is the other big cost. Depending on the distance traveled, U.S. bottled water consumption in 2007 required the equivalent of 32 million to 54 million barrels of oil. In effect, says Gleick, each bottle is 15% to 25% oil.

U.S. SALES, 2006 (millions of liters)	
Carbonated soft drinks	57,169
Bottled Water	31,238
Beer	24,489
Milk	21,476

Hyung-Chul Kim, an industrial ecologist at Columbia University, says the analysis didn't include savings from recycling bottles. Gleick counters that most bottles end up not in new bottles but in carpet, clothing, or toys from China.

Swept Away

Geologist William Dickinson thinks he's worked out when and how coral atolls became habitable abodes for Pacific Islanders. Based on that understanding, he says most may go under during this century.

In the March issue of *GSA Today*, Dickinson, of the University of Arizona, Tucson, explains how swings of sea level formed more than 175 atolls strewn across 9000 kilometers of the tropical Pacific. About 2000 years ago, he says, falling seas allowed waves to build stable islets of sediment on old reefs where 180,000 people now live.

But as global warming drives sea level back up, the process will reverse, notes Dickinson. Once the sea rises higher than the solid atoll foundation, the water "starts to chew into the flanks long before the islets are overtopped," he says. So well before livable areas are flooded, he predicts they'll be washed away "some time between a few decades from now and a century from now." Geologically, "it's a nice story," says geologist Robert Ginsburg of the University of Miami in Florida. Culturally, not so good.

Saliva's Secrets

There's a world in your mouth, and it bears little relation to the world about you.

Mark Stoneking, a molecular anthropologist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, wondered whether mouth bacteria could tell a tale of evolution and human migration analogous to what has been revealed by different strains of the stomach bug *Helicobacter pylori.*

Stoneking and colleagues collected saliva from 10 volunteers at each of a dozen locations around the world and sequenced parts of a key bacterial gene in each sample. They found 101 known bacterial genera, 39 of them newly seen in the mouth, and about 64 novel types. Individuals had between six and 30 kinds of mouth bacteria. The big surprise was that there were no geographical patterns in species' presence, despite presumed regional commonalities in diet and environment, the team reported 26 February in *Genome Research*. People in each location shared about 60% of the genera; individuals from different places had about 50% in common.

The study adds to the growing awareness of the personalized world of human-associated microbes, an underappreciated aspect of human health and disease until recently, says Ruth Ley, a microbial ecologist at Cornell University: "These communities are very complex, and the degree to which they vary between people continues to surprise [us]."

Donating saliva samples in the Philippines.

Published by AAAS

EDITED BY YUDHIJIT BHATTACHARJEE

Three Q's >>

Cancer biologist **Frank Torti**, 61, took a leave of absence from his post at Wake Forest University School of Medicine last May to join the U.S. Food and Drug Administration (FDA) in the newly created role of chief scientist. On 20 January, he was appointed acting director of the agency. While the Obama Administration hunts for a permanent head, Torti has been busy elevating the role of science at an agency criticized for not taking it seriously enough.

Q: How are things these days at the agency?

The FDA has faced some real challenges, and it hasn't solved all of those challenges yet—I think it's poised to do so. There are so many crosscutting issues in science that need to be anticipated, understood … [to regulate] these products, which are very different than traditional small molecules.

Q: What have you been doing to bolster science at FDA?

Just tons of stuff. But we're not done. … At the FDA until recently, science has been largely invisible. … I asked each center [at FDA] what their overarching [scientific] priorities for the center are. … They [include] rapid detection, sensitive high-throughput methodologies, adverse-event analysis, biomarkers of both toxicity and efficacy, personalized medicine, and nutrition. … For each, we have projects that the centers can tackle.

Q: And do you have the budget to do this?

Right now, we don't have the budget to do it all. We'll show [Congress] what the projects are and what they cost, and what we can do now and what we [can't afford]. It won't be that we'll have our hands out for funding; [we will talk about] priorities that everyone can endorse.

CREDITS (TOP TO BOTTOM): FDA; SCRIPPS INSTITUTION OF OCEANOGRAPHY/UC SAN DIEGO; SOURCE: RICHARD ALLEY; SONJA FLEMMING/CBS

Got a tip for this page? E-mail people@aaas.org

DATA POINT

MISSION ACCOMPLISHED? In his speech to Congress last week, President Barack Obama asked the country to help him improve U.S. education. The goal, he said, should be that "by 2020, America will once again have the highest proportion of college graduates in the world."

A stretch? Hardly. Turns out the United States is already there. A recent report by the U.S. National Science Foundation (www.nsf.gov/statistics/seind08/) shows the United States leading the world, with roughly 30% of its adult population holding 4-year college degrees.

But Obama is right to be worried, according to Thomas Snyder of the U.S. National Center for Education Statistics. Over the past 2 decades, countries across Europe and Asia have poured money into their universities on the assumption that a well-educated population is essential for long-term economic and national security. As a result, several industrialized nations now top the United States in the percentage of younger adults ages 25 to 34 with college degrees. "The concern is with younger people," says Snyder.

AWARDS

CLIMATE BIGWIGS. Atmospheric scientist V. Ramanathan and glaciologist Richard Alley share this year's $200,000 Tyler Prize for Environmental Achievement for their work on climate change.

Ramanathan, of the Scripps Institution of Oceanography at the University of California, San Diego, says the prize will help him get "back to my roots." As a child in India, he watched his grandmother cook over a smoky fire of dung and wood. As a climate researcher, he documented the "brown cloud" blowing off India from such fires and burning fossil fuels. He then showed how the cloud warmed the upper atmosphere, reduced monsoon rainfall and rice harvests, and led to the retreat of Himalayan glaciers supplying drinking water to billions of people. Now he is organizing a project in north India to show how much using more efficient stoves and cleaner fuels could thin the brown cloud.

Alley, of Pennsylvania State University, University Park, has retrieved records of climate's wild gyrations from ice cores and investigated how glaciers might rush to the sea under global warming. His communication style has been described as combining Woody Allen and Carl Sagan, with a bit of "Weird Al" Yankovic thrown in. In one YouTube video, a guitar-strumming Alley warns of global warming in a folksy parody of "Proud Mary."

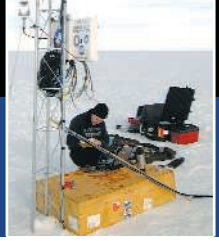
Two Cultures

STELLAR ROLE. What do a Nobel Prize and a television sitcom have in common? George Smoot.

Next week, the astrophysicist at the University of California, Berkeley, who was honored in 2006 for confirming the big bang theory, appears as himself on the show of the same name. The cast was starstruck by Smoot's presence. Jim Parsons, who plays the arrogant genius Sheldon, said, "You['ve] actually got *a scientist* to talk to on this set!" Producers Chuck Lorre and Bill Prady glowed as Smoot autographed gift copies of his book *Wrinkles in Time*. "We've peaked with Dr. Smoot," said Lorre.

"On the show, I was a little bit harsh," Smoot says about his reaction to meeting an overconfident Sheldon. In reality, says Smoot, "I try very hard to be encouraging to young scientists, even the ones that are way off base." He even ad-libbed a few positive comments during rehearsal. After the filming, he received a memento: a director's chair with his name on it.

Published by AAAS



The climate data gap
1276

UCS at 40
1279

U.S. BUDGET

Amid the Gloom, Researchers Prepare for a Boom in Funding

Your retirement account may be taking a beating, but if you have a grant application pending at a U.S. science agency, there's an upside to the global financial meltdown: Your chances of being funded have never been better. And if your application isn't already in the pipeline, don't despair. The competition for funds should be eased significantly next year.

This improved research outlook comes partly from the $787 billion stimulus package signed by President Barack Obama last month. It provides an additional $22.5 billion across several research agencies, including $10.4 billion for the National Institutes of Health (NIH), $3 billion for the National Science Foundation (NSF), and $1.6 billion for the Department of Energy's (DOE's) Office of Science. Agency officials are under orders from White House budget czar Peter Orszag to spend the money "quickly and wisely," and they are now working out procedures for getting it spent.

The stimulus money is in addition to agencies' regular budgets, which were expected to be finalized this week as part of a $410 billion spending bill covering the rest of the 2009 fiscal year. The version passed 25 February by the House of Representatives contains a 20% boost for DOE's Office of Science, a major supporter of basic research across the physical sciences, and a 6.5% increase for NSF. NIH would receive a $937 million bump. (The Senate was due to act after *Science* went to press.)

And that's not all. On 26 February, Obama delivered a 2010 budget request to Congress that would start to make good on his campaign promise to double the federal investment in basic research over the next 10 years. If Congress goes along, the $3.6 trillion proposal would boost NIH spending on cancer research by 21% during the fiscal year that begins on 1 October, raise NSF's budget by another 8.5%, to $7.04 billion, and give "similarly large increases" for DOE science and for research at the National Institute of Standards and Technology (NIST). A detailed breakdown for each agency won't be available until April.

That trifecta of spending will mean tremendous opportunities for the U.S. research community. It puts NSF, DOE science, and NIST back on the doubling track that President George W. Bush proposed in 2006 and that is enshrined in a 2007 law filled with good intentions but no cash. "I'm very pleased that Congress and the White House have provided us with an unprecedented level of resources for this year," gushed NSF Director Arden Bement at a briefing last week for the National Science Board, NSF's oversight body. Adding the stimulus package to NSF's regular appropriation, Bement explained, gives NSF a total budget of $9.5 billion in FY 2009. That's nearly 60% more than it received in 2008. And Obama's first budget request, Bement adds, "contains a very good number for NSF."

It's also welcome news for biomedical researchers, who have felt the squeeze from an NIH budget that has remained essentially flat since 2004 after a 5-year doubling. "We are very happy that we are getting out of the doldrums," says cell biologist Richard Marchase, president of the Federation of American Societies for Experimental Biology. Marchase expects that many high-scoring grant proposals sidelined in 2008 will now get approved, adding that the 2010 budget framework holds out hope for steady and sustained annual growth of roughly 7%.

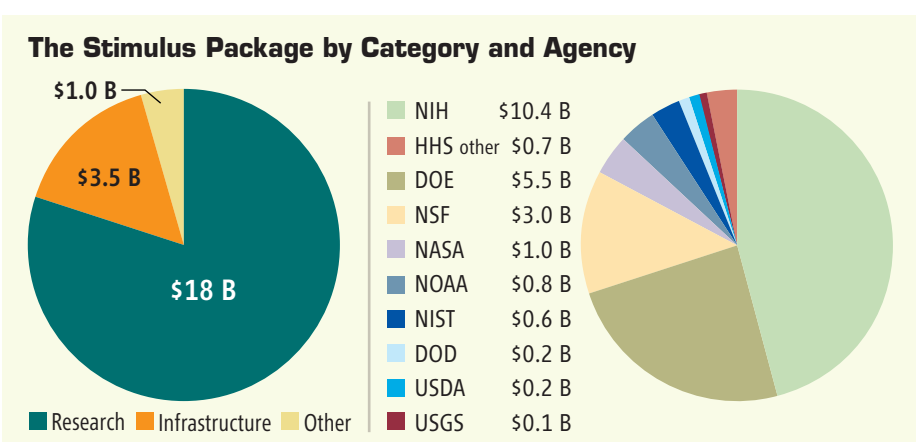
Online
sciencemag.org
Podcast interview with author Jeffrey Mervis.

Opening words. Peter Orszag presents the president's 2010 budget framework.

Shovel-ready science

For research agencies, the immediate challenge is how to spend their unexpected wealth from the stimulus package. The Obama Administration has assured Congress and the public that these funds will go out the door as quickly as possible without lowering standards, and that their impact on the economy—in particular, on the number of jobs created—will be monitored closely. Although each agency is setting out its own guidelines for how to spend the money and what information grantees will need to provide, the rules drawn up by NSF and NIH appear typical.

Both agencies plan to dip into the existing pool of applicants for the bulk of the new



Sharing the wealth. Ten agencies receive research and construction funds from the massive recovery plan.

CREDIT (TOP TO BOTTOM): GREG MATHIESON/MAI/LANDOV; SOURCE: AAAS

Biosecurity after Ivins
1282

The "ant idea man"
1284

awards. For NIH, that includes proposals that didn't make the merit review cutoff in 2008; for NSF, it means proposals submitted since last fall that seek funding in the current fiscal year. These peer-reviewed projects could be called shovel-ready science. To avoid commitments beyond 2010, most NSF awards will be in the form of standard grants, in which the full 3-year total is funded up front.

NIH will give top priority to its bread-and-butter R01 grants, but most other categories are also eligible for stimulus money. Unlike at NSF, NIH is also preparing a solicitation for 2-year challenge grants, using up to $200 million from an $800 million pot given to the NIH director. The agency hopes to fund challenge grants in 14 areas, with an application deadline of 27 April.

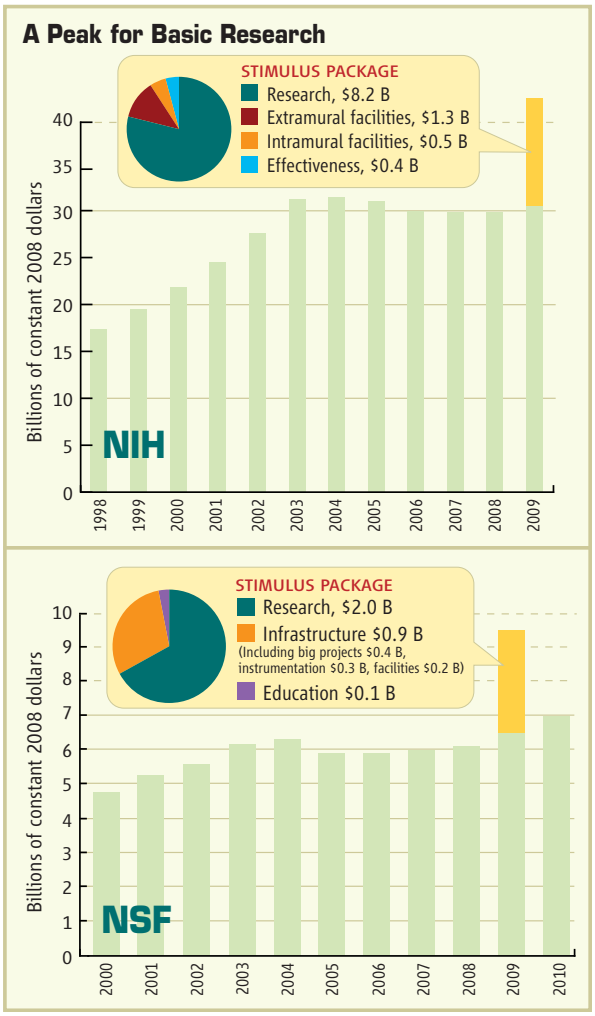
Agency officials promise that speed won't compromise the quality of the reviews. NSF's Division of Mathematical Sciences, for example, is in the midst of running some 50 panels that are sifting through roughly 2500 proposals submitted last fall. "Our instructions to the panels haven't changed," says division director Peter Marsh. "We're still looking for the best science." But Marsh, whose 2008 budget was $212 million, expects that the additional funds (which haven't yet been allocated among NSF's six research directorates) will result in a healthy rise over last year's 31% success rate.

That could be good news for mathematician Andrew Belmonte of Pennsylvania State University, University Park. Belmonte is awaiting word on a grant proposal he submitted in November to NSF's applied math program after taking an unsuccessful shot the previous year. His work on the transformation of materials during fluid flows is expensive—costing 10 times the normal math project, he estimates—because it requires a well-equipped wet lab. And if funded, he'll need to hire a postdoc and several undergraduates, a priority for the stimulus package. Belmonte says he's ready for any additional paperwork: "It's the government's money, and if they want more reporting, I'm happy to do it."

The emphasis on funding what's already in the pipeline obviously favors those lucky enough to have applied in the current funding cycle. But Marsh says that because NSF will be funding so many more grants this year, there should be fewer applications competing for funds next year. At the same time, the recipients of this year's spending spree are likely to show up 3 years hence in the competition for renewals. "So the stimulus will have both a ripple and an echo effect," he says.

–JEFFREY MERVIS

With reporting by Eliot Marshall.



A big payday. The stimulus money provides a major boost to recent sluggish budgets at NIH (*top*) and NSF (*above*).

SOURCES: AAAS; (INSETS, PIECHART DATA) AMERICAN RECOVERY AND REINVESTMENT ACT, 2009

Budgets in Brief

Some highlights from the stimulus package, the 2009 budget before Congress, and President Barack Obama's 2010 request:

■ **ENERGY:** DOE's Office of Science gets $4.77 billion in 2009, up 19%. The ITER project in France gets $124 million, $90 million less than requested. DOE's national labs and other facilities divide the $1.6 billion stimulus money, and the department gets $400 million to create the Advanced Research Projects Agency–Energy. No 2010 research numbers are available.

■ **NSF:** The $394 million boost for 2009, to $6.49 billion, is on top of $3 billion in stimulus money for research, infrastructure, and education. The agency's six research directorates would grow to $5.18 billion, and education to $845 million. The $11 million Robert Noyce Scholarship program for prospective teachers gets its second straight $40 million bump, and a program to help researchers in states that struggle to win NSF grants would grow by $20 million, to $133 million.

■ **NIH:** The agency's 3.2% increase this year, to $30.3 billion, is supplemented by $10.4 billion from the stimulus package. The 2010 budget mentions only "over $6 billion within the National Institutes of Health to support cancer research." The National Cancer Institute's budget this year is $5 billion.

■ **NASA:** The space agency will receive $17.8 billion in 2009, $380 million more than in 2008. A $200 million drop in the agency's science programs, to $4.5 billion, is remedied in its $1 billion stimulus package. The $18.7 billion request for 2010 bolsters earth sciences and robotic probes to visit other planets. The new launcher to send humans to the moon wins funding, but the White House may review the target of a 2020 lunar landing.

■ **NIST:** A $63 million jump in 2009, to $819 million, maintains the $65 million Technology Innovation Program. Obama would raise that to $70 million in 2010. The stimulus funds add $220 million to a $470 million research budget, $180 million to a $172 million lab-construction program, and provide $180 million for a facilities grants competition.

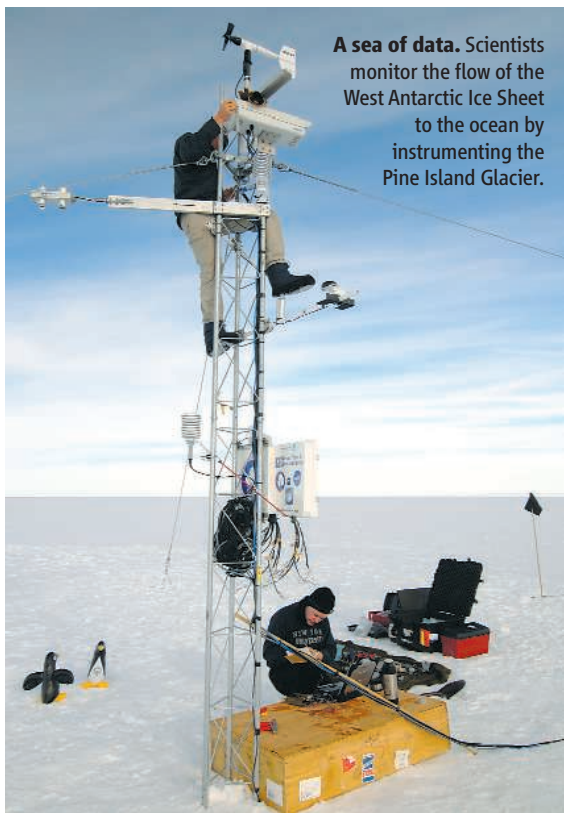
–NEWS STAFF

Published by AAAS

EARTH MONITORING

Loss of Carbon Observatory Highlights Gaps in Data

With rising temperatures altering a variety of ecological and weather systems on Earth, the current patchwork of sensors can't answer all the questions that scientists are asking. Land-based sensors have provided a conclusive picture of rising CO_2 levels worldwide, for example, but researchers don't fully understand where all the carbon that humans and natural sources are pouring into the atmosphere ends up. How much is being absorbed and where?



A sea of data. Scientists monitor the flow of the West Antarctic Ice Sheet to the ocean by instrumenting the Pine Island Glacier.

CREDIT: CLIFF LEIGHT/NSF

NASA's Orbiting Carbon Observatory (OCO) was supposed to provide some answers about the nature of carbon sinks on land and in the oceans. But on 24 February, the rocket carrying the $278 million satellite crashed shortly after takeoff, the victim of a failure related to the nose cone. "Bang—it's gone. It was absolutely terrible," says Pieter Tans, an atmospheric scientist at the National Oceanic and Atmospheric Administration's laboratory in Boulder, Colorado. Tans was hoping to improve his four-dimensional maps of CO_2 flows—useful to close the global carbon budget—by using OCO's unique ability to detect tiny carbon fluxes.

The loss of OCO is part of a bigger problem of data gaps related to climate. Some gaps, like OCO, rest upon relatively new measurements. In other cases, climate scientists worry that satellites well beyond their operational lifetimes will fail before their replacements are in orbit. Sometimes, as with the monitoring of forest biomass, polar ice, and sea levels, the records go back for decades. "The need for a systematic and comprehensive approach to collecting climate observations has taken on new urgency," concluded a panel from the National Academies in a report issued last week. No single agency or person has the overall authority for the multibillion-dollar challenge of observing Earth's climate, it noted. Researchers are hopeful that a $400 million boost for space-based Earth sensors included in the stimulus package (see p. 1274) will make a difference. But a 2007 academy panel estimated it would cost $6 billion through 2020 to fix a system it said was "at risk of collapse" in 2005.

This backlog exists at a time when better observations are urgently needed, especially at Earth's poles, say researchers. The 2007 report from the Intergovernmental Panel on Climate Change identified basic observations to better understand ice sheet physics as a key requirement for more reliable predictions of sea-level rise. POLENET, an effort to install seismic monitors all over the Antarctic continent, will provide measurements of the temperature of the bedrock on which ice sheets sit, a major factor that affects the speed at which ice slides toward the coast. But bad weather, compounded by rising fuel costs that forced the National Science Foundation to trim spending on polar research, allowed scientists to place only one new station in the field out of 16 originally planned; most operational stations are clustered near McMurdo Station, the main U.S. base on the continent.

The speed at which ice sheets are declining is also governed by the interaction of glaciers with water at the ocean shore. A robotic submarine has recently collected data on the Pine Island Glacier, the fastest moving glacier in Antarctica. Scientists would like a long-term picture of shifting dynamics at the edge of the ice sheet as well as this snapshot. But a project to set up stations on the glacier's edge to monitor the region below the ice was delayed by logistical hurdles: The ice was too rough to allow planes to land on skis, and arranging helicopter facilities wouldn't be possible until the 2011–12 season. "Logistics in Antarctica proceed slowly, [and] we only get to make progress for about 3 months in any year," says NASA's Robert Bindschadler.

Modelers say more data could also clarify the role of atmospheric aerosols in global warming. Aerosols help form clouds, which can both warm and cool the atmosphere. The main sensor on the $14 billion National Polar-orbiting Operational Environmental Satellite System (NPOESS), scheduled for a 2010 launch, will provide aerosol data that will be inferior to data gathered by a sensor on an existing NASA craft called Terra, already 2 years beyond its operational life. In 2007, the National Research Council's (NRC's) decadal study for the field suggested that NASA launch a replacement between 2013 and 2016, and Michael Freilich, NASA's earth sciences chief, says his staff has begun to scope out the mission's requirements. "We're committed to the decadal," he says.

Even Earth-observing mainstays like Landsat are at risk. Landsat 5 and Landsat 7 are both well beyond their nominal design life, regularly missing swaths of images during malfunctions. Scientists who rely on the crafts to study forests' contributions to the carbon cycle say a gap in land imagery is virtually assured because the next satellite in the series, the Landsat Data Continuity Mission, won't be launched before 2012. "We may lose one or both of these satellites before then," says Compton Tucker of the White House Climate Change Science Program (CCSP).

Another problem area is ocean color, a variable that scientists use to measure photosynthesis in the ocean—a giant and poorly understood contributor to the global carbon cycle. NASA's orbiting SeaWiFS sensor has been steadily recording data since 1997. But the instrument is 7 years beyond its design life, and the sensor meant to measure ocean color on the first NPOESS mission will likely be unable to provide sufficient data because of technical problems (*Science*, 15 February 2008, p. 886). That has spurred some managers to push for a "gap filler" mission sooner, ▶

Published by AAAS

says Tucker. But Freilich is hesitant to commit.

CCSP is leading an interagency effort to finalize a plan that would for the first time provide a set of government priorities for Earth observations, says Tucker. Scientists say that's the first step needed to fill the gaps.

In some cases, international partners might help, he says. Japan's recently launched Greenhouse Gases Observing Satellite, for example, detects CO_2 starting at the ground level, like OCO. Although its resolution is lower, it well complements an existing NASA satellite that only picks up CO_2 readings at an altitude of 5 km. "We're going to see what we can continue to do with our Japanese partners to fill the gap," says NASA's Stacey Boland.

The effort to better coordinate climate-related monitoring will no doubt be strengthened by the NRC report released last week and presented on 26 February to presidential science adviser John Holdren. Holdren has extolled the virtues of "observation and scientific study of the condition of our home planet's land, vegetation, oceans, and atmosphere." Climate researchers are hoping those words, delivered at his confirmation hearing to be director of the White House Office of Science and Technology Policy, will soon translate into support for filling the holes in the nation's current Earth-monitoring system.

–ELI KINTISCH

EDUCATION

Study Questions Value of School Software for Students

U.S. students using educational software do no better learning primary school math and first-year algebra than their counterparts who follow a traditional curriculum. That's the conclusion of a new federally funded study that is loaded with caveats about what it means for students, educators, and the companies that make the software.

The $14.5 million study, funded by the U.S. Department of Education and conducted by Mathematica Policy Research Inc. in Princeton, New Jersey, was designed to find out whether students are benefiting from the growing use of educational software. Preliminary results released in April 2007 (ies.ed.gov/ncee/pubs/20074005/index.asp) suggested that the answer, for first- and fourth-grade students in reading and for sixth- and ninth-grade students in math, was no. But those results were widely criticized by educators and software makers for lumping together the outcomes from many different products and for testing their impact on student achievement too early, in the first year the teacher had used the material. The second-year results, posted 17 February (ies.ed.gov/ncee/pubs/20094041/index.asp), address those concerns by reporting results for individual software packages and by testing a new cohort of students whose teachers already had a year of using the software under their belts.

What it found is that none of the four math products tested produced a statistically significant difference in achievement between control and treatment groups over the 2-year study, which began in the fall of 2004. (One of the six reading packages registered a meaningful jump in test scores by fourth graders.) Students of teachers using Cognitive Tutor, a computer-based curriculum for algebra I students developed by Carnegie Mellon University researchers (*Science*, 2 January, p. 64), for a second year showed a meaningful improvement in test scores. But the software had no overall impact on student achievement when data from both student cohorts were combined.

Education research is rarely definitive, however, and this study is no exception. For one thing, the students and teachers were not randomly assigned to the products, explains Mathematica's Mark Dynarski. That rules out comparing one software package with another. It also means that the results shouldn't be generalized to different student populations. "Everyone began at different starting points. So we can only say how this product works at a particular school," he says.

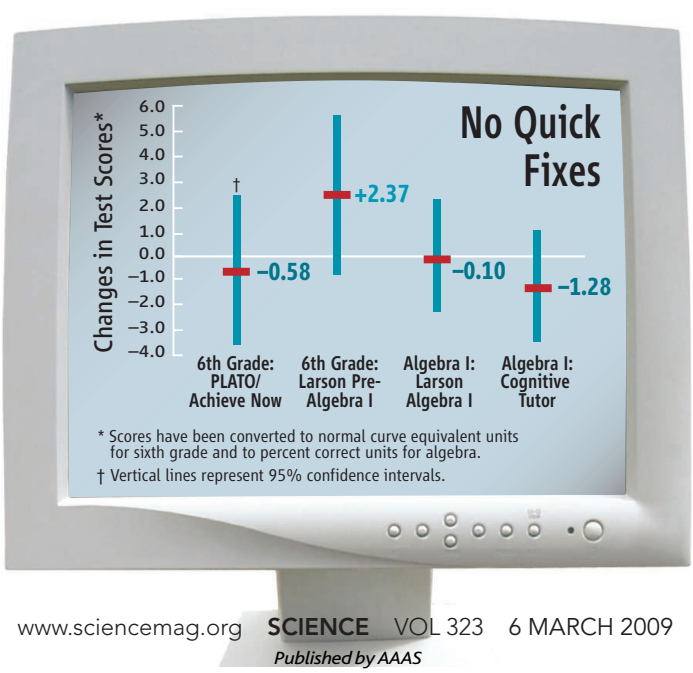
Another confounding factor is that researchers didn't have enough money to continue observing teachers in their classrooms or surveying them about how they used the product, something that was done in the first cohort. That makes it difficult, for example, to interpret large fluctuations from one year to the next in the amount of time spent using particular software.

Does an increase for algebra I students, for example, mean that teachers found the product more helpful once they became comfortable with it? That seems a logical explanation to Cognitive Tutor's Steve Ritter, who points to his company's emphasis on supporting teachers. "For us, usage went up, and second-year teachers achieved statistically significant improvements," Ritter says. "That's worth talking about."

In contrast, does a decline in usage mean that teachers found that it wasn't right for their students, who are disproportionately from large, poor, and urban schools? Or was it simply due to a dearth of computer facilities, inadequate tech support, or scheduling problems? And what effect does an extremely mobile teacher corps—only one in four taught the same grade at the same school for both years—have on student outcomes?

Dynarski cautions against overinterpretation of the data. He notes that Congress wanted to know whether education software improves student learning, not whether individual products can help certain populations taking certain subjects. "There might be a lot of positive things going on in the classroom, such as greater fluency with computers," he says. "But it's not leading to the kind of results that people want, which are better test scores."

–JEFFREY MERVIS



Subpar scores. Software didn't help sixth-grade math or first-year algebra students do better.

SOURCE: EFFECTIVENESS OF READING AND MATHEMATICS SOFTWARE PRODUCTS, NCES 2009-4042; (IMAGE) PHOTOS.COM

Published by AAAS

EPIDEMIOLOGY

Author of Iraqi Deaths Study Sanctioned

The lead author of a controversial study that concluded that the U.S.–led invasion of Iraq in 2003 caused a high number of violent deaths—more than 600,000 Iraqis—has been sanctioned for a lapse in ethics. The Bloomberg School of Public Health at Johns Hopkins University in Baltimore, Maryland, found last week that epidemiologist Gilbert Burnham had committed "violations of the Bloomberg School's policies regarding human subjects research" by failing to fully protect the confidentiality of interviewees. It ordered a 5-year suspension of "Burnham's privileges to serve as a principal investigator on projects involving human subjects research," the school announced in a press release.

"The [Bloomberg] School has asked me to make no comments at the present," Burnham said to *Science*.

A university investigation discovered that the full names of Iraqi people interviewed appeared on some of the 1800 data-collection forms filled out during the house-to-house survey designed by Burnham. The university gave Burnham approval for the study with the understanding that no unique identifiers would be recorded. The survey was carried out by a team of Iraqi researchers who met with Burnham in Jordan to receive instructions and hand over data (*Science*, 20 October 2006, p. 396). According to a Hopkins press release, the investigation found "no evidence that the violations caused harm to any individuals involved in the study" because the data sheets were "never out of the possession of the research team." As principal investigator, Burnham was held responsible.

"The punishment is very severe, and making it public is unusual," says Gary King, a statistician at Harvard University.

Les Roberts, a co-author of the 2006 *Lancet* study who is now at Columbia University, acknowledged in an e-mail exchange with *Science* that a "serious" breach of confidentiality occurred in the survey forms but said that Burnham was not initially aware of it. "When the forms arrived in Jordan all filled out, [Burnham] asked and was assured that these words at the top of the forms were not complete names or unique identifiers," Roberts wrote, and "most are written in Arabic, which [Burnham] does not read."

The debate is likely to continue. The sanction "has nothing to do with the scientific merits or defects of the study or its results," notes Debarati Guha-Sapir, an epidemiologist at the World Health Organization's Collaborating Centre for Research on the Epidemiology of Disasters in Brussels. Statistician Seppo Laaksonen of the University of Helsinki notes that Hopkins didn't evaluate the study's sampling methodology or statistical approach, which he cares about more than the punishment. But Roberts dismissed criticism of the study. In an e-mail, he wrote: "There is a way to verify [our] findings: replicate. The posh academics who spend their lives in offices are offended by that, but in this arena, rigor is more about ground activities than software manipulations."

According to the Hopkins press release, "an erratum will be submitted to *The Lancet*" concerning the unauthorized recording of names. **–JOHN BOHANNON**

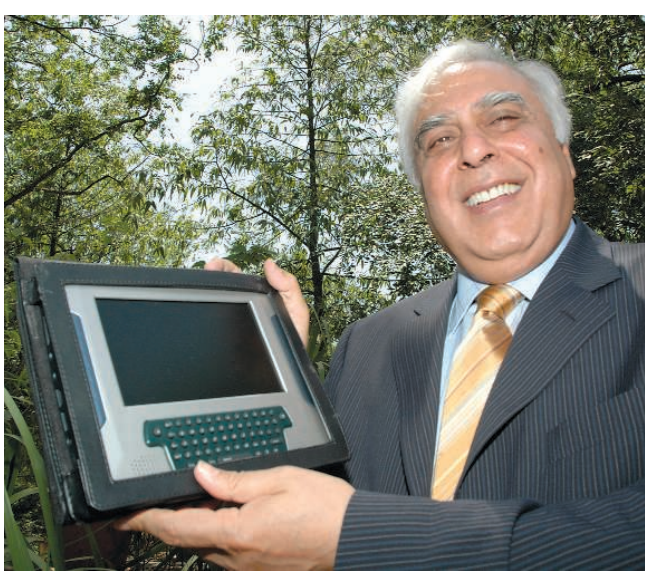
SCIENCE AND BUSINESS

India Allows Government Scientists to Own Companies

NEW DELHI—In 2001, Swami Manohar and three colleagues at the Indian Institute of Science (IISc) in Bangalore invented the Simputer, a simple and cheap hand-held computer. But as civil servants, the computer scientists by law could not commercialize their invention. "I had no choice but to resign," says Manohar, who is now chief of intellectual property and strategy at Geodesic Limited, a telecom firm that bought the company Manohar and his colleagues founded after leaving IISc in 2001.

Indian scientists will no longer be forced to make such a stark choice. On 24 February, the Department of Scientific and Industrial Research issued regulations that permit researchers at government-funded institutions to hold equity stakes in scientific enterprises and spinoff companies. The "historic decision," says Science Minister Kapil Sibal, will "unleash the latent entrepreneurial potential of Indian scientists."

The policy shift is expected to have a profound impact in India: Some 400,000 scientists, about three-quarters of the scientific work force, are employed at public institutions. By bringing India in line with the United States and other Western nations, the new rules should create an attractive environment for talented expatriate scholars to return to India, says Samir Brahmachari, director general of the Council of Scientific and Industrial Research in New Delhi who helped shepherd the regulation through 18 ministries over 18 months.



More freedom. Inventors like Swami Manohar, co-inventor of the Simputer, will gain from a regulatory change.

CREDIT: PALLAVA BAGLA

The new rules also permit research institutes to hold equity stakes in commercial enterprises. To facilitate this process, the government will encourage the lateral mobility of researchers between institutes and industry. "Cross-fertilization between the academics and industry is very much necessary," says Sibal. Although the regulations came too late for Manohar to keep his post at IISc, he applauds what he sees as a long-overdue change. "Scientists need the freedom to flourish, and now they have gotten it," he says.

–PALLAVA BAGLA

Published by AAAS

UNION OF CONCERNED SCIENTISTS

From Protest to Power: An Advocacy Group Turns 40

On 4 March 1969, some of the most prominent scientists at the Massachusetts Institute of Technology (MIT) issued a declaration of political dissent and scientific self-criticism. Stirred into action by student protests against the war in Vietnam, the professors convened a campuswide meeting and declared that the "misuse of scientific and technical knowledge presents a major threat to the existence of mankind." The statement bore the name of a previously unknown organization: the Union of Concerned Scientists (UCS).

Forty years later, UCS has 80,000 members and a staff of 130 working in four cities. The organization campaigns to shrink nuclear arsenals, fight global warming, and reduce agriculture's harm to the environment. For 8 years, it was a thorn in the side of the Bush Administration, criticizing White House policies and zinging its appointees for "politicizing science." That aggressive approach raised its visibility and helped triple its budget this decade, to almost $20 million.

But just as some MIT faculty members of yesteryear ignored the teach-ins and went ahead with their normal duties, some scientists believe that UCS cannot claim to be the conscience of the scientific community. "Many of its statements and conclusions … were perverse oversimplifications of complex issues," says John Marburger III, science adviser to President George W. Bush and a frequent target of UCS attacks. "I think it's hard, maybe impossible, for an advocacy organization to be entirely science-based."

Although its launch attracted wide attention, UCS almost succumbed to the status quo soon afterward. "Many of us were quite disturbed that we'd spent several months playing hooky," recalls physicist Kurt Gottfried, one of the primary organizers of the MIT event. As students and faculty members returned to their normal routines, Gottfried says, "this hotbed of activity sort of faded away."

That it did not is due largely to MIT physicist Henry Kendall, a respected scientist—he won the Nobel Prize in physics in 1990—and heir to an industrial fortune. Kendall, who died in 1999, once said he learned a valuable lesson about advocacy during the 1960s as a member of the JASON group that provided confidential technical advice to the Department of Defense. "Being inside [the government] was not an effective way to change national policy," he told the American Institute of Physics in 1986, because the government "would not take advice it didn't like."

"For the first couple of years, Henry Kendall pulled some money out of his own pocket" to cover expenses, recalls Daniel Ford, a young economist who became UCS's first executive director. Kendall and Ford teamed up on studies that were critical of reactor safety, bringing UCS to national prominence. Perhaps their biggest media coup took place in 1976, when Ford arranged for Robert Pollard, a safety official at the U.S. Nuclear Regulatory Commission (NRC), to announce on *60 Minutes* that he was resigning in protest and going to work for UCS. "In the 1970s, the organization was 90% focused on nuclear energy," says Ford.

An abiding concern. UCS arose in the antiwar movement, was kept afloat by Henry Kendall and doubts about the safety of nuclear reactors, and now helps governors promote cars using alternative fuels.

Yet nuclear power was not a simple issue for the organization. "A lot of physicists were strong supporters of nuclear power; they'd been involved in its invention," says Gottfried, who is stepping down this year as chair of UCS. He will be replaced by James McCarthy, past president of AAAS (which publishes *Science*). By calling for tighter safety regulations while insisting that it was not "antinuclear," UCS walked a fine line. It tried to strike a similar balance when criticizing other technologies, such as missile defense or genetically engineered crops, that some scientists strongly support.

Its opponents accuse it of giving mere lip service to scientific objectivity. George Keyworth, science adviser to President Ronald Reagan, says that "their political agenda has been manifest for their entire career. … I just never take them very seriously." But John Ahearne, a former NRC chair who now directs Sigma Xi's ethics programs, says UCS has been "a good counterweight to industry."

Along with successfully pushing for tighter regulation of nuclear power in the 1970s and 1980s, UCS played a prominent role in opposing deployment of antisatellite weapons and missile defenses during the 1980s. In recent years, its attention has shifted toward finding ways to reduce greenhouse gas emissions.

As it marks its 40th birthday, UCS has never been closer to the center of political power. Issues on which UCS has campaigned—energy efficiency, clean cars, and big cuts in greenhouse gas emissions—have been embraced by the Obama Administration. Yet political prosperity may also leave it weakened. UCS was a child of the 1960s antiwar movement, and the organization has prospered during periods when it strongly opposed government policies. In contrast, membership and donations fell during the 1990s when President Bill Clinton was in office.

Kevin Knobloch, a former journalist and congressional aide who has been president of UCS since 2003, sees no signs of waning support now that another Democrat occupies the White House. "We have this tremendous window of opportunity to win adoption of farsighted policies," he says. "This is our moment." **–DAN CHARLES**

CREDITS (TOP TO BOTTOM): COURTESY MIT MUSEUM; BETTMANN/CORBIS; MARIO ANZUONI/REUTERS/LANDOV

Published by AAAS

SCIENTIFIC MISCONDUCT

Retractions Put Spotlight on China's Part-Time Professor System

BEIJING—Traditional Chinese medicine (TCM) is a slippery subject: In herbal concoctions, it's often difficult to discern active compounds from fillers and contaminants. Sorting wheat from chaff is all the harder when data themselves are concocted. In the latest scandal to grip Chinese academia, Zhejiang University (Zheda) in Hangzhou last November fired an associate professor in a recently established TCM research lab after finding him guilty of scientific misconduct, the university said in a statement. Three journal articles have been retracted so far out of eight that the university has charged contain plagiarized or fabricated data. Since then, more questionable papers have come to light.

The case, which came to national attention in February, could have widespread fallout. Some observers criticize Zheda for firing the junior researcher, He Haibo, before the investigation is complete. "I don't think Zhejiang University handled the case properly," says Zhang Haixia, a professor at Peking (Beijing) University who led a misconduct inquiry a few years back. "What about the responsibility of the corresponding author?" asks Zhang. The handling of the case has sparked an outcry in Chinese media and on the Internet focused on the senior scientist who established the TCM lab and, more broadly, on a system that critics say gives power but little accountability to academicians.

In 2004, Zheda recruited Li Lianda, who was elected the previous year as an academician of the Chinese Academy of Engineering, to be part-time dean of its College of Pharmaceutical Sciences. Li, an expert on TCM modernization, also set up a lab at Zheda and has supported it with a 5-year, $430,000 grant from the Ministry of Science and Technology to study TCM drugs for cardiovascular disease. But Li spends most of his time in his main lab in Beijing, where official meetings require his presence, as he has explained to Chinese media.

He Haibo joined the Hangzhou lab as its first postdoc in July 2006 after receiving a Ph.D. from China Pharmaceutical University in Nanjing 1 month earlier. Judging from his publication record, He was a dynamo. Between April and November 2007, he submitted eight manuscripts on which he was first author to international journals; all appeared in print in 2007 or 2008. According to a Chinese reporter who tracked down a former grad student in the lab, He was well-liked because he helped students fulfill the college's requirement that Ph.D. candidates publish one first-authored and at least a second paper in a journal in the Science Citation Index database. (Since 2006, Zheda has published more indexed papers than has any other Chinese university.) As a reward for He's remarkable productivity, Zheda hired him as an associate professor last July.

Things quickly unraveled for He. In October, an editor at the *International Journal of Cardiology* came across a paper in *Phytotherapy Research* with figures and tables similar to those in a manuscript *IJC* was about to publish. The manuscript reported a study on two molecules in a signaling pathway involved in irregular heart palpitations caused by inflamed cardio muscles in rats, whereas the published paper is about the same pathway during a heart attack. On 11 October, the editor contacted the manuscript's senior author, pharmacologist Dai De-Zai of China Pharmaceutical University, for an explanation, according to Dai's blog. Dai—who was He's Ph.D. adviser—wrote in his blog that he was surprised to see that his former student had, as Dai alleged, plagiarized graphs, tables, and texts from his lab. In his blog, Dai says he got in touch with the executive vice dean of Zheda's pharmacology school and contacted the journal that published He's paper. Contacted by e-mail, Dai declined to provide further comment.

According to a Zheda statement received by *Science* on 13 February, the college confronted He on 16 October; administrators asked him to retract the fraudulent paper, reflect on his mistakes, and apologize to his former adviser. Ten days later, according to the statement, He submitted a self-criticism essay in which he cleared Li of any knowledge of his wrongdoing and then left the college. To protect privacy, the university says, it is not making He's essay public. Zheda President Yang Wei told Chinese media that He, in his self-criticism, said he committed misconduct because he was desperate for a faculty position, which would enable him to remain at Zheda and earn a living. After ascertaining that He had e-mailed the requisite retraction to *Phytotherapy Research*, Zheda says, the university fired him on 13 November. He Haibo could not be reached by telephone and did not respond to e-mail messages.

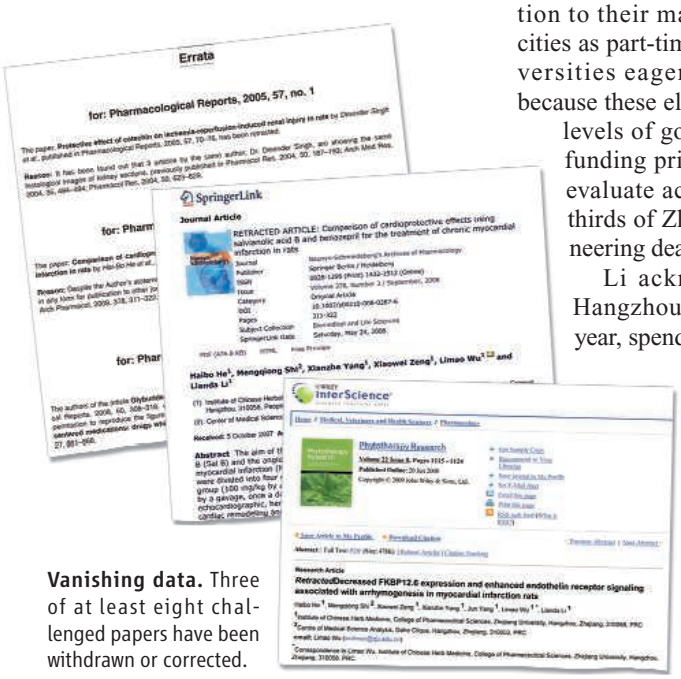
The story did not end there. Li and his lab came under fire on the Internet. On 23 October, an anonymous posting on New Threads (www.xysforum.org) alleged that Li's Zheda lab published duplicate articles in *Pharmacological Research* and *Naunyn-Schmiedeberg's Archives of Pharmacology*. Apparently, before he disappeared, He e-mailed both journals to retract the papers, giving "duplicate publication" as the reason, according to e-mails provided to *Science*. The journal editors alerted Zheda to the retractions. Three days later, another anonymous posting on New Threads alleged that the duplicate papers, along with a third in the *Journal of Ethnopharmacology*, contain fabricated data.

In the crossfire. Academician Li Lianda remotely oversees a lab that has been shaken by allegations of plagiarism.

CREDIT: HE TAO/*SCIENCE NEWS*

Published by AAAS

The postings caught the eye of Zhu Guo-Guang, a TCM practitioner in Oulu, Finland. Zhu downloaded the three papers for a closer look and says he found fabricated data "written in black and white." To Zhu, the alleged misconduct is a big blow to TCM research. "Fake research worsens the reputation of traditional Chinese medicine," he says. As a vice chair of the Pan European Federation of TCM Societies, Zhu says he felt a responsibility to report his findings to the three journals and to Zheda. Zhu assembled the evidence and mailed hard copies to the journal editors and to Yang Wei in mid-November. As Chinese netizens turned up more questionable journal articles from Li's Hangzhou lab, Zhu set up a blog to post his analysis and provide updates.



Vanishing data. Three of at least eight challenged papers have been withdrawn or corrected.

Zhu's allegations grabbed national attention after they were reported in a 3 February article in the Chinese newspaper *21st Century Business Herald*. Zheda issued a statement the next day, laying the blame on He and stating that Li's name had been given as a co-author of the fraudulent papers without Li's authorization. Yang Wei says he is now personally sending out retraction requests.

Meanwhile, more anonymous postings appeared on New Threads; Zheda now says that 19 publications allegedly have problems, including papers from Li's lab at the China Academy of Traditional Chinese Medicine in Beijing and by other members of the Hangzhou lab that did not include He as an author.

A team led by Harold Garner of University of Texas Southwestern Medical Center in Dallas has analyzed several papers and added them to the Déjà vu database of extremely similar publications (spore.swmed.edu/dejavu/duplicate/74404/); also see entries 75181 through 75184. "We try to provide the most unbiased data possible which an appropriate body, such as an editorial board or a university ethics board, can use to make their own evaluations," says Garner.

The scandal has prompted fresh debate about moonlighting by academicians. Many, like Li, hold two or more positions in addition to their main job, often in different cities as part-time deans or lab chiefs. Universities eagerly recruit academicians because these elite scientists advise various levels of government on research and funding priorities, review grants, and evaluate achievements. At least two-thirds of Zheda's 15 science and engineering deans are such part-timers.

Li acknowledges that he visits Hangzhou only a half-dozen times a year, spends 3 to 5 days in the college, and has only 1 day for his five or six grad students in the TCM lab. Wu Limao, who received his Ph.D. from Li's Beijing lab in 2004, ran the lab's day-to-day operations until August 2008 when Wu left to spend 2 years as a visiting researcher at Yale University. Li says that, as a victim of He's fraud, he has become a subject of investigation.

In the meantime, Zheda has expanded its probe and has asked Wu to return to Hangzhou to answer some questions. Wu is corresponding author on eight papers that Zheda has found to be fraudulent and first or corresponding author of other papers called into question. According to a preliminary investigation report received by *Science* on 25 February, the university faults Wu for not reviewing the papers before publication. Wu declined to comment for this article.

The saga is inflicting collateral damage on researchers who have collaborated with Li's labs. *Phytotherapy Research* Editor-in-Chief Elizabeth Williamson says the journal is preparing to retract more articles from Li's labs and has held up all manuscripts with authors whose names appear on suspicious papers, pending Zheda's further investigation. "I don't want to penalize honest people, but it's getting difficult to know who to trust these days," she says.

–HAO XIN

*Science*Insider

From the *Science* Policy Blog

From his "State of the Union"–like speech to Congress last week, to his new budget, to his seemingly never-ending selection of advisers and Cabinet members, the U.S. president commanded the science policy spotlight over the past week. Here are some highlights from *Science*Insider:

The big news has been **Obama's 2010 budget.** Mirroring the enthusiasm Congress has shown for science, the president asked for big boosts for the National Science Foundation, NASA, and the Environmental Protection Agency. The EPA budget includes a $19 million increase for the agency to create a greenhouse gas emission inventory—an important step in preparing the United States for a cap-and-trade system. Also seeing green is autism research. Obama requested $211 million as part of the Department of Health and Human Services budget for research into causes and new treatments. But the president isn't showing everyone the money. Fulfilling a campaign promise to oppose plans to build a storage vault for nuclear waste at Yucca Mountain in Nevada, Obama's budget scales back funding for the project.

To many scientists, the president's budget gets a lot right. But in **Obama's speech to Congress** a couple of days earlier, he got at least one thing wrong. The president promised to ensure that "by 2020, America will once again have the highest proportion of college graduates in the world." The United States is already there, however. The country leads the world, with roughly 30% of its adult population holding 4-year college degrees.

In **non-U.S. news**, U.K. Prime Minister Gordon Brown assured scientists that U.K. research will not be "a victim of the recession." Brown's address highlighted his vision for science as a driver of the United Kingdom's economic future, a desire that has triggered much debate within the scientific community. Meanwhile in Japan, the incoming president of the University of Tokyo dismissed the importance of school rankings. "University research should be not to satisfy individual egos but to enrich the entire society and humanity," he said.

For the full postings and more, go to **blogs.sciencemag.org/scienceinsider**.

The Danger Within

The suicide of Army researcher Bruce Ivins will inevitably mean changes in biosecurity policy, but government and academic scientists are divided on how stringent the new procedures should be

On the morning of 2 August 2008, 3 days after Army researcher Bruce Ivins committed suicide, the Biodefense Policy Coordination Committee assembled in a conference room of the Eisenhower Executive Office Building, a stone's throw from the White House. The mood around the table was somber when Robert Kadlec, committee chair and member of the Homeland Security Council, spelled out the meeting's agenda, according to one of the participants. If Ivins had indeed carried out the 2001 anthrax letter attacks, as the Federal Bureau of Investigation (FBI) claimed, what could be done to prevent another bioterrorist act by an insider at a government or academic lab?

The insider threat has never loomed so large in policy discussions on bioterrorism. The massive expansion of biodefense research after the anthrax mailings in 2001, which killed five people and sickened 17, was largely predicated on the concern that a terrorist group could launch an attack using biological weapons it had developed or stolen from a military facility. The FBI's implication of Ivins in the mailings—although still questioned and untested in a court of law—sparked the unsettling realization among researchers and biosecurity experts that "the insider threat is a lot bigger problem than we ever thought," says a former scientist at the U.S. Army Medical Research Institute of Infectious Diseases in Frederick, Maryland, where Ivins worked.

In recent months, that concern has echoed through the halls of the U.S. government. Soon after the 2 August meeting, the White House asked the National Science Advisory Board for Biosecurity (NSABB)—a panel set up in 2005 to address the potential security risks from life sciences research—to consider how best to ensure that academic and industry researchers working with select agents are trustworthy. It is expected to come up with recommendations next month. Independently, an interagency working group, co-chaired by the secretaries of Defense and Health and Human Services, is looking at ways to strengthen lab biosecurity.

Some change in biosecurity procedures is inevitable post-Ivins, experts agree. But how stringent any new steps should be is being hotly debated. Since NSABB got its charge from the White House, board members have been nervously discussing the costs and benefits of measures such as tougher background investigations, psychological screenings, drug tests, and credit checks. Other possibilities include video surveillance of labs and a two-person rule prohibiting researchers from working alone.

Insider. Bruce Ivins worked at USAMRIID.

Some of those measures are standard practice at the Department of Defense (DOD), the Department of Energy (DOE), and other agencies where scientists work with nuclear, chemical, or biological agents. But some academics worry that extending such measures to academia would undermine the open culture of life sciences research and impede biodefense research. A better approach, some say, would be to trust lab managers to keep a watchful eye on employees for signs of troubling behavior. But even advocates of self-policing concede that that approach may not allay fears of the insider threat.

Trying to foil an insider attack

The U.S. government already vets researchers working with select agents through rules that were put in place a year after the anthrax attacks. Under those rules, institutions and individuals who wish to work with any of 80 biological entities, including the Ebola virus and botulinum toxin, must go through a Security Risk Assessment (SRA).

The SRA process involves a basic FBI background check. Disqualifying factors include a criminal history, a substance-abuse or mental health problem, a dishonorable discharge from the military, or being a citizen of a country that the United States deems a sponsor of terrorism. The

CREDITS: CHRISTOPHER T. ASSAF/MCT/LANDOV; REUTERS/LANDOV

Published by AAAS

Eye of the storm. Reporters flocked to Ivins's home in Frederick, Maryland, after news of his suicide.

FBI runs an applicant's fingerprints through criminal and terrorist databases but generally doesn't investigate whether an applicant has honestly answered questions about substance abuse and mental illness. An SRA can take up to 45 days and is valid for 5 years.

Although many scientists complain that the SRA is onerous and impedes research, experts point out it is a fairly limited form of vetting. "The SRA catches the people who are dirty already," FBI's Kristine Beardsley said at a 12 December 2008 NSABB meeting during which the board discussed personnel reliability. By contrast, a clearance looks at an individual's entire background—not just law enforcement records.

Personnel reliability procedures at federal agencies are already much stricter than those prescribed by the select agent rules. At DOD, employees whose work involves access to select agents and toxins have to undergo a secret-level Personnel Security Investigation that involves, among other things, credit checks and interviews with friends and acquaintances. (Like other defense researchers, Ivins had this clearance.) Once admitted into the lab, employees are monitored through drug tests and periodic medical evaluations. Although there is no formal psychological testing, those who display mentally unstable behavior—such as threatening colleagues—can be removed from the lab. The monitoring measures have been formally in place since 2004.

Since Ivins was implicated, DOD has considered making the system even more stringent. An agency task force has recommended background checks equivalent to those required for a "top secret" clearance. The prospect worries even those used to DOD's security-conscious culture. "They'll hunt down your ex-wife, your landlord, anybody who knows you," Kenneth Cole, a senior Pentagon official, said at the 12 December meeting.

DOE's safeguards could also be a model for NSABB. At DOE's Lawrence Livermore National Laboratory (LLNL), where a security clearance is mandatory for most employees, the dozen or so researchers who work with select agents are further vetted through extensive conversations with past managers. Eric Gard, select agent manager at LLNL, says he has turned down individuals "who have trouble managing their temper, for instance," assigning them instead to work with less hazardous pathogens in a biosafety level 1 lab. Once they are cleared to work with select agents, employees are given a psychological evaluation at least once a year. Occasional credit checks are done "to determine whether somebody might be tempted by their poor financial circumstances to do something desperate," Gard explains.

These measures come at a cost: Academic researchers often suffer a rude shock when they come to work at Livermore, and the intense scrutiny there makes it difficult to attract and retain talent, Gard says.

Peering into the mind

At the 12 December meeting, some NSABB members urged caution before embracing measures such as psychological screenings, which they worry would drive academic scientists away from biodefense projects, slowing vaccine development and other biomedical advances.

Richard Ebright, a chemist at Rutgers University, New Brunswick, doesn't buy the argument. A proponent of stronger regulations, Ebright points out that for all the grumbling, researchers entered the biodefense field in droves after the select-agent rules came into effect.

Reliable? Researchers at biocontainment labs may now be subject to increased scrutiny and monitoring.

At the meeting, some board members wondered aloud whether psychological monitoring would yield too many false alarms without adding much security. Many scientists take "pride in their eccentricity," said Stanley Lemon, a microbiologist at the University of Texas Medical Branch at Galveston.

Gerald Epstein, a biosecurity expert at the Center for Strategic and International Studies in Washington, D.C., agrees. He also questions whether the art of "peering inside people's psyches" to predict their future actions has been perfected yet. "If Ivins was the anthrax mailer, what could we have seen ahead of time in his behavior to stop him?" he asks.

In the summer of 2007, the Office of the Director of National Intelligence (ODNI) commissioned an exercise by scientists and security experts to answer a question along those lines. Group members were divided into red teams and blue teams, says ODNI's Lawrence Kerr, who coordinated the study. The red teams came up with three scenarios of bioterrorist attacks, while the blue teams identified points in the scenarios at which the offender could be captured and the attack foiled. "It was like reading a Michael Crichton book, except that the scenarios contained actual recipes for attacks," says Kerr, a former academic microbiologist. (The study is classified.)

In one of the scenarios, a trained biologist plotted an attack from within a sophisticated lab. "The key judgment we arrived at was that if such an individual was truly intent on carrying out an attack, they would succeed," Kerr says. "It scared the crap out of us."

The exercise led Kerr and his colleagues to devise three categories of indicators that could help detect an insider threat: human, technological, and material. In all three, the study found, the lab's principal investigator, other researchers, and administrators were in the best position to spot red flags such as odd behavior or anomalous purchases of equipment and materials. The take-home message was that researchers need to be sensitized to security concerns. At the same time, institutions need to develop procedures for reporting potentially suspicious activities to higher-ups without fear of embarrassment or recrimination. Kerr is working with scientific societies including the National Academies and AAAS, the publisher of *Science*, to educate life scientists about indicators to watch for.

Even the idea of keeping an eye on colleagues makes some academics nervous. "The default position should be of mindful trust, not of distrust," NSABB chair Dennis Kasper said at the meeting.

NSABB is likely to emphasize "local judgment" in its recommendations, says the board's Susan Ehrlich. But mere self-governance is likely to be a hard sell, board members admit. "We are in a classic double bind," says Kasper. "If we don't implement a program that is very serious, we are going to get criticized. If we do and research shuts down, we'll get criticized."

Kerr says once all the reviews are complete, the government is likely to prescribe a combination of grassroots vigilance and increased top-down oversight. But no matter what measures are taken, it may be impossible to guarantee that a insider attack won't happen. The reason, Kasper says, is "inherent imperfection in people."

–YUDHIJIT BHATTACHARJEE

CREDIT: AP PHOTO/TIMOTHY JACOBSEN

Ant man. Nigel Franks's studies of individual ant behavior help unlock the secrets of the colony.

PROFILE: NIGEL FRANKS

Watching as Ants Go Marching—And Deciding—One by One

The "Ant Idea Man" challenges established views by using paint and radio tags to see each ant as an individual

BRISTOL, UNITED KINGDOM—"She looks so purposeful," says Nigel Franks, pointing at an ant about the size of a printed colon (:) and laughing. "It's absolutely anthropomorphizing, but sometimes an ant seems to get a buzz out of doing something successfully." The ant in question, a member of a *Temnothorax albipennis* colony, had just carried one of her sisters from their old home to a new one and was already speeding back to fetch another ant. Franks, a behavioral ecologist and head of the Ant Lab at the University of Bristol, shook his head in amusement at her tenacity, but his eyes never left the tiny cinnamon-brown figure as she rushed ahead on her task. Other ants were also engaged in moving the colony to its new nest—an urgent undertaking that Franks and graduate student Elizabeth Franklin had set in motion by removing the top of the old one. For the ants, it was as though a tornado had suddenly ripped off the roof of their home, and they immediately sprang into action. Another experiment in Franks's Ant Lab was under way.

Over the past decade, Franks's mix of experimentation and ant-watching has produced a steady stream of research, both controversial and groundbreaking, about how *T. albipennis* ants make decisions. Decision-making in groups, whether ant or human, is increasingly a hot research topic in everything from computer design to election committees to armies. In a special issue on the topic last month in the *Philosophical Transactions of the Royal Society B*, Franks and his team described how ants implement quick decisions in the best possible way even in a time of crisis, such as when their home has just been destroyed.

"Most people think that ants are stupid, but they're not," says Franks. "They have very sophisticated behaviors, although they don't have language or theory of mind [the ability to understand what another individual is thinking]. And that's what intrigues me: this behavior that looks like thinking, yet we can reduce it to an algorithm."

Franks is probably best known—some would say infamous—for showing what he called teaching in ants, as foragers led naïve nestmates on searches for food. It was the first time this talent was shown in a species other than humans. Franks's assertion outraged many, however, and prompted changes in the criteria for teaching as set out by animal cognition researchers. Franks's experimental work was "outstanding," says bee cognition expert Lars Chittka of Queen Mary, University of London. "But what the ants are doing is not 'teaching' in the human sense."

Critics and supporters alike agree that Franks has a unique talent for asking questions and designing ingenious experiments. "You can tell which students have worked in Nigel's lab," says his former mentor Bert Hölldobler, who was nevertheless dismissive of the teaching finding. "You can see his handwriting in just the way they ask questions." Franks "always has a new angle," adds Thomas Seeley, an insect sociobiologist at Cornell University. "In ant biology, he's known as the 'Idea Man.'"

Tracking ant opinion polls

Franks has been watching ants since he was a child in Yorkshire, when he kept a colony on top of his wardrobe closet. "Even then, I really wanted to understand how the individuals cooperated to make a functioning whole." His interest took him to the famous ant lab of E. O. Wilson at Harvard University, where he worked with Hölldobler, now at Arizona State University in Tempe. Hölldobler and Michael Möglich had recently discovered a new ant communication signal in a *Temnothorax* species (*Science*, 13 December 1974, p. 1046), which they labeled "tandem calling." When scouting for a new nest, individual ants led other scouts to the site using a mix of chemical signals and physical touching.

In the 1980s, on a research trip to Bulgaria, Franks encountered another *Temnothorax* species, and the proverbial light bulb went off: He'd found a good experimental species. *Temnothorax* colonies often live in cracks in rocks, so their habitat could be easily mimicked in a lab, and they tend to have between 40 and 400 individuals. "I could hold an entire colony in the palm of my hand," Franks says.

CREDITS: COURTESY OF NIGEL FRANKS

Fashionistas. A few daubs of bright paint turn identical ants into individuals.

Published by AAAS

He and his students collect colonies from rocks along the English Channel each year for the Ant Lab. The ants set up their colonies inside a cardboard layer sandwiched between two microscope slides, the whole colony about 3 to 4 millimeters high and set into a petri dish the size and shape of a CD jewel case. The lab's counters are stacked with the dishes, each one a world unto itself with a separate colony. In each dish, "I have everyone in their society," Franks explains. That means, he argues, that he has a better understanding of their social behavior than do colleagues studying apes or birds in labs.

Most of Franks's experiments are based on the ants' readiness to move from one nest to another. He and his students have discovered things such as how the ants evaluate a new nest: They assess its interior size and shape by keeping track of how often they cross their own trails. The ants agree on which nest to move to via "opinion polls." When enough scouts agree on a new nest, they stop leading other ants to it and switch to the faster method of simply carrying the remaining members of the colony there. "They basically vote with their feet," says Franks.

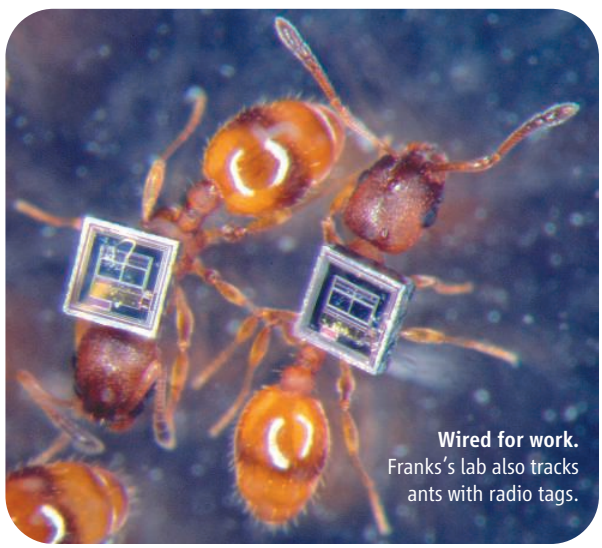
In a current paper, Franks and his team investigate a problem faced by ants and many human decision-makers, too: how a speedy and so possibly poor decision can lead to an even greater disaster (think of Hurricane Katrina). In the ants' case, in a crisis, a handful of scouts might locate a poor but acceptable nest and begin to emigrate while some of their fellows continue the search. Without the usual number of scouts to lead the emigrants, that would slow the emigration and leave the colony dangerously exposed for a longer time. To speed things up, these scouts lead "some of their undecided scout colleagues" from the new nest to the old one. This technique apparently turns them into advocates for the new home, because they immediately switch their behavior from searching to emigrating, saving the colony.

Ants as individuals

To understand how a colony makes decisions, Franks first separates it into its parts. He usually paints each ant by carefully wedging it headfirst into a piece of foam and touching its gaster with four hues of paint. It can take 10 hours—"a horrible job," Franks says. But then the ants are easily identified as individuals. "An ant colony is an information-processing unit, a little brain," says Seeley. "Nigel is dissecting it, then putting it back together again." With his interdisciplinary team, Franks reduces the decisions of the ants to mathematical models, revealing how simple rules followed by individual ants produce the complexity of the colony.

Recently, Franks's team took another step in tracking ants, reporting online in this month's *Behavioral Ecology and Sociobiology* the first successful experiment with extremely small radio-frequency identification tags (RFID) glued to the backs of each ant. "It's a fabulous technical advance," says Anna Dornhaus, a former postdoc in Franks's lab, now at the University of Arizona, Tucson. "It will make many things possible, such as monitoring the ants for 24 hours and finding out who is doing what."

Franks and lead author and postdoc Elva Robinson used it to investigate whether social information or individual experience is most important in an ant's decision to leave the safety of its nest. *T. albipennis* ants forage after receiving information from a scout about a good food source or because they are hungry. To find out which trigger is stronger, Robinson prevented scouts from returning to the nest, then used the RFID tags to track when the remaining ants decided to go foraging and which ones did so.



Wired for work. Franks's lab also tracks ants with radio tags.

Even ants without previous foraging experience ventured out to seek food if they were especially lean, the team showed. That finding adds to previous studies in Franks's lab showing that an individual ant's behavior has some flexibility. Such an assertion runs counter to previous studies by Wilson and others that correlate an ant's age with its given task, suggesting that tasks are hard-wired. "Our studies show the opposite, and they are frankly extremely controversial," Franks acknowledges. "But we followed individual marked ants for 6 months, photographing each colony several times a day," amassing data that he asserts no one else has gathered.

"When Wilson says, 'It's this way,' many people won't investigate further," says Dornhaus. "But Franks is different, which is why he has such original papers; he studies things other people think are already settled. That's why I asked to join his lab." Still, Hölldobler warns that conclusions based on Franks's chosen ant species, with its small colony size, may not be a good model for ants in general.

Franks's creative defiance is apparent in his decision to test whether ants can teach. "There was a definition that everyone accepted," he explained, "and so we looked at the ants' behavior to see if an ant 'modifies its behavior in the presence of another, at a cost to itself, so another individual can learn more quickly.' " He and grad student Thomas Richardson watched hours of videotaped "tandem runs" of foraging ants leading naïve ants (all painted) from the nest to a new source of food. These tandem runs are carefully orchestrated, with the lead ant walking slowly and waiting for her follower to tap her antennae on the leader's legs and abdomen before continuing, and "fulfill the criteria" of teaching, the two wrote in *Nature*. Not everyone was impressed. "This might be a charming metaphor," Hölldobler and Wilson wrote in their new book, *The Superorganism*, "but it adds little, if anything, to our understanding." Franks insists that the behavior is teaching and says that the lead ant in effect evaluates her pupil, a behavior he and Richardson describe further in a 2007 *Current Biology* paper.

Back at Franklin's experiment, the scout ants were ferrying the last of the brood into the new nest. But these scouts were young juveniles, or "callows," because Franklin had carefully brushed away the older ants. Were the callows up to the task? "There's one who has two ants following her," Franks pointed out. "But she's moving too fast. … Oh, she's lost them now." The lost ants milled around, looking … lost. What would they decide to do? Which elements of a tandem run are innate; which are learned? Franks fired off question after question. The Idea Man leaned in close, his eyes tracking the ants' every move: A new experiment was surely in the works.

–VIRGINIA MORELL

CREDITS (TOP TO BOTTOM): TOM RICHARDSON; NIGEL FRANKS

On the Origin of Photosynthesis

Try to picture the world without photosynthesis. Obviously, you'd have to strip away the greenery—not just the redwoods and sunflowers, but also the humble algae and the light-capturing bacteria that nourish many of the world's ecosystems. Gone, too, would be everything that depends on photosynthetic organisms, directly or indirectly, for sustenance—from leaf-munching beetles to meat-eating lions. Even corals, which play host to algal partners, would lose their main food source.

Photosynthesis makes Earth congenial for life in other ways, too. Early photosynthesizers pumped up atmospheric oxygen concentrations, making way for complex multicellular life, including us. And water-dwellers were able to colonize the land only because the oxygen helped create the ozone layer that shields against the sun's ultraviolet radiation. Oxygen-producing, or oxygenic, photosynthesis "was the last of the great inventions of microbial metabolism, and it changed the planetary environment forever," says geobiologist Paul Falkowski of Rutgers University in New Brunswick, New Jersey.

Given its importance in making and keeping Earth lush, photosynthesis ranks high on the top-10 list of evolutionary milestones. By delving into ancient rocks and poring over DNA sequences, researchers are now trying to piece together how and when organisms first began to harness light's energy. Although most modern photosynthesizers make oxygen from water, the earliest solar-powered bacteria relied on different ingredients, perhaps hydrogen sulfide. Over time, the photosynthetic machinery became more sophisticated, eventually leading to the green, well-oxygenated world that surrounds us today. In the lab, some biochemists are recapitulating the chemical steps that led to this increased complexity. Other researchers are locked in debates over just when this transition happened, 2.4 billion years ago or much earlier.

Looking so far into the past is difficult. The geological record for that time is skimpy and tricky to interpret. Eons of evolution have blurred the molecular vestiges of the early events that remain in living organisms. But "it's a terribly important problem," says biochemist Carl Bauer of Indiana University, Bloomington, one well worth the travails.

To catch a photon

Over more than 200 years, researchers have ironed out most of the molecular details of how organisms turn carbon dioxide and water into food. Chlorophyll pigment and about 100 other proteins team up to put light to work. Plants, some protists, and cyanobacteria embed their chlorophyll in two large protein clusters, photosystem I and photosystem II. And they need both systems to use water as an electron source. Light jump-starts an electrical circuit in which electrons flow from the photosystems through protein chains that make the energy-rich molecules ATP and NADPH. These molecules then power the synthesis of the sugars that organisms depend on to grow and multiply. Photosystem II—the strongest naturally occurring oxidant—regains its lost electrons by swiping them from water, generating oxygen as a waste product.

However, some bacteria don't rely on water as an electron source, using hydrogen sulfide or other alternatives. These nonconformists, which today live in habitats such as scalding hot springs, don't generate oxygen. Their photosynthetic proteins huddle in relatively simple "reaction centers" that may have been the predecessors of the two photosystems.

Envisioning the steps that led to this complex biochemistry is mind-boggling. Similarities between proteins in photosynthetic and nonphotosynthetic bacteria suggest that early microbes co-opted some photosynthesis genes from other metabolic pathways. But protophotosynthesizers might also have helped each other piece these pathways together by swapping genes. Biochemist Robert Blankenship of Washington University in St. Louis, Missouri, and colleagues say they've uncovered traces of these lateral gene transfers by comparing complete bacterial genomes. For example, their 2002 study of more than 60 photosynthetic and nonphotosynthetic bacteria (*Science*, 22 November 2002, p. 1616) suggested that bugs had passed around several photosynthesis genes, including some involved in synthesizing the bacterial version of chlorophyll.

Gene-sharing might also explain the puzzling distribution of the photosystems, Blankenship says. A cell needs both photosystems to carry out oxygenic photosynthesis. Yet modern nonoxygenic bacteria have the presumptive predecessor either of photosystem I or of photosystem II, never both. To explain how the two protein complexes wound up together, Blankenship favors "a large-scale lateral [gene] transfer" or even a fusion of organisms carrying each photosystem. However, other researchers remain skeptical, arguing that one photosystem evolved from the other, possibly through the duplication of genes, creating an ancient cell with both. No one knows for sure.

The electron thief

Either way, it took some fancy fiddling to convert the primitive reaction centers to oxygen-generating photosystems. Oxygenic photosynthesis was a huge upgrade, leading to a land of plenty, says biochemist John Allen of Queen Mary, University of London. "Water is everywhere, so the organisms never ran out of electrons. They were unstoppable."

But water clings to its electrons. With its oxidizing power, photosystem II can wrench them away, but the reaction centers in nonoxygenic photosynthesizers cannot. Biochemists James

THE YEAR OF DARWIN

This essay is the third in a monthly series. More on evolution online at **blogs.sciencemag.org/origins**.

CREDIT: K. SUTLIFF/*SCIENCE*

Published by AAAS

Allen (no relation to John Allen) and JoAnn Williams of Arizona State University, Tempe, and colleagues are working out how a bacterial reaction center could have evolved photosystem II's appetite for electrons.

Taking a hands-on approach, they have been tinkering with the reaction center of the purple bacterium *Rhodobacter sphaeroides* to determine if they can make it more like photosystem II. First they targeted bacteriochlorophyll, the bacterial version of chlorophyll that's at the core of the reaction center, and altered the number of hydrogen bonds. Adding hydrogen bonds hiked the molecule's greed for electrons, they found.

The water-cleaving portion of photosystem II sports four manganese atoms that become oxidized, or lose electrons. So the team equipped the bacterial reaction center with one atom of the metal. In this modified version, the added manganese also underwent oxidation, the researchers reported in 2005. James Allen says that their creations aren't powerful enough to split water. But eventually, they hope to engineer a reaction center that can oxidize less possessive molecules, such as hydrogen peroxide, that would have been present on the early Earth. Even if the researchers never replicate photosystem II, "if we define the intermediate stages, we've accomplished a lot," he says.

Catching rays. Long before plants got in on the act, photosynthetic cyanobacteria living in pools like this one in Yellowstone National Park were changing the composition of the atmosphere.

> Oxygenic photosynthesis "was the last of the great inventions of microbial metabolism, and it changed the planetary environment forever."
>
> —Paul Falkowski, Rutgers University

Something in the air

How the photosystems got their start is crucial for understanding the origin of photosynthesis. But the question that's drawn the most attention—and provoked the most wrangling—is when photosynthesis began. Most researchers accept that nonoxygenic photosynthesis arose first, probably shortly after life originated more than 3.8 billion years ago. "Life needs an energy source, and the sun is the only ubiquitous and reliable energy source," says Blankenship.

The sharpest disputes revolve around when organisms shifted to oxygenic photosynthesis. At issue is how to interpret a watershed in the fossil record known as the great oxidation event (GOE). In rocks from about 2.4 billion years ago, geologists see the first unmistakable signs of significant, sustained levels of atmospheric oxygen. These signs include red beds, or layers tinged by oxidized iron, i.e., rust. Further support that the GOE marks an atmospheric revolution comes from a technique that detects skewed abundances of sulfur isotopes that occur if the air lacks oxygen. These imbalances persisted until the GOE, when they vanished.

Hard-liners construe these data to mean that oxygenic photosynthesis could not have emerged until shortly before the GOE. But other scientists disagree. "We are finding more and more hints that oxygenic photosynthesis goes deeper into the fossil record," says astrobiologist Roger Buick of the University of Washington, Seattle. These hints could push the origin back 600 million years or more.

One line of evidence is oil biomarkers that researchers think are the remains of cyanobacteria. They've turned up in rocks that are up to 2.7 billion years old. And in western Australia, thick shale deposits that are 3.2 billion years old hold rich bacterial remains but no traces of sulfur or other possible electron sources, suggesting that the microbes were using water to make energy.

Geologist Euan Nisbet of Royal Holloway, University of London, and colleagues found additional support for an early origin when they went searching for traces of RuBisCO, a key photosynthetic enzyme. RuBisCO feeds carbon dioxide into the reactions that yield sugars. The enzyme version found in oxygenic photosynthesizers plays favorites: It prefers carbon dioxide that contains the carbon-12 isotope over the bulkier carbon-13. In 2007, Nisbet and his colleagues found disproportionately low carbon-13 values indicative of RuBisCO activity when they analyzed organic matter in rocks from three sites in Zimbabwe and Canada that are between 2.7 billion and 2.9 billion years old. Nisbet concludes that oxygen-making photosynthesis began at least 2.9 billion years ago.

The early-origin case isn't ironclad. For example, a 2008 paper that has some researchers fuming claims that the oil biomarkers are contaminants that seeped in from younger rocks. Advocates also have to explain why it took hundreds of millions of years for oxygen to build up in the air.

Although the last word on the origins of oxygen-making photosynthesis isn't in, researchers say they are making progress. One thing is for certain, however: Without this innovation, Earth would look a lot like Mars.

–MITCH LESLIE

LETTERS

edited by Jennifer Sills

For Science, No Escape from Politics

IN HER LETTER "*SCIENCE* SHOULD STICK TO SCIENCE" (9 JANUARY, p. 208), A. M. Thro urges scientific journals to refrain from commenting on politics. Every one of the scientific researchers that I know relies on U.S. federal government support. The choice of a president is therefore very much a scientific issue. Rather than encouraging scientists to stick to science, perhaps it would be more useful to say that science is stuck to politics. This being the case, it is entirely appropriate and, indeed, responsible for scientific journals to comment on politics.

Science is a social activity, and scientists should never work alone. Science is not, and should never become, separated from the real world that it tries to explain.

TIMOTHY CLAIR

5540 Eaglebeak Row, Columbia, MD 21045, USA. E-mail: obscurechemist@gmail.com

Research Rankings: Does Size Matter?

THE NEWS OF THE WEEK STORY ABOUT THE U.K. Research Assessment Exercise (RAE) ("U.K. university research ranked; funding impacts to follow," J. Travis, 2 January, p. 24) used a table of "research power" from the journal *Research Fortnight* to show university rankings. This table is flawed. The rankings are derived by multiplying the GPA (or "Quality Index") from the RAE by the number of staff submitted by the institution. The Quality Index varied across 130 institutions from 17.7 to 60.3 (a range of less than 4), whereas staff numbers varied from 12.5 to 2246 (a range of 180). Consequently, the latter dominates the final metric and the table published in the News of the Week story is in effect a table of size of university: The university ranking almost exactly matches the staff number ranking. University performance tables are increasingly prominent and, if we are to have them, they should reflect quality, not size.

ALASTAIR FITTER

Department of Biology, University of York, York YO10 5DD, UK. E-mail: ahf1@york.ac.uk

Response

THE *RESEARCH FORTNIGHT* RESEARCH ASSESSment Exercise (RAE) analysis indeed factors in staff size, but that does not make it flawed; they and others argue that since the number of staff submitted to RAE will factor into the money allocation, this type of ranking will more accurately reflect the ultimate funding decisions than a simple ranking by quality index would have. Indeed, if *Research Fortnight* or *Science* had simply run an unadjusted quality index, the big schools, who submitted the work of thousands of researchers, would have claimed we were using a flawed system. The *Times Higher Education* league table, for example, had the Institute of Cancer Research ranked first, despite the fact that it submitted the work of only 97 staff in just 2 of the 67 RAE research categories.

JOHN TRAVIS

Certificates Should Be Retired

IN THEIR THOUGHTFUL POLICY FORUM ON research ethics, "Certificates of Confidentiality and compelled disclosure of data" (14 November 2008, p. 1054), L. M. Beskow, L. Dame, and E. J. Costello describe a case where research records, despite having this Certificate, were not fully protected from a court-ordered subpoena. Considering this case and our Institutional Review Board (IRB) experiences, we wonder whether the Certificate, unevenly required by IRBs and time-consuming to obtain, provides any additional, substantive protections to research subjects.

Currently, Informed Consent and HIPAA authorization documents explicitly state how research data will be secured, with whom data may be shared, and what monitoring agencies may have access. Furthermore, consent documents often state which information (e.g., child or elder abuse) will be disclosed to authorities. These explicit data-sharing permissions and protections appropriately cover all research records, regardless of their supposed "sensitive" or "nonsensitive" content.

Obtaining a Certificate of Confidentiality from NIH places burdens on investigators and lengthens informed consent documents already laden with confusing legal language. If time is critical, investigators may choose to remove sensitive questions or instruments to achieve more rapid IRB approval. Hence, the value of the research may be compromised.

With the goals of enhancing subject recruitment and obtaining honest research data on sensitive topics, one must weigh the purported benefits of the Certificate against their potential impedance of established legal and judicial system processes. As imperfect as those processes may be, we believe that the effectiveness of legal arguments and the privacy rights provided to all research subjects

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 3 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Published by AAAS

Ecosystems under climate change

1300

Bracing for extreme events

1302

by covered entities should govern final judicial decisions, and not a Certificate selectively issued by NIH.

JOHN A. HERMOS* AND AVRON SPIRO III

VA Boston Healthcare System, Boston, MA 02130, USA.

*To whom correspondence should be addressed. E-mail: john.hermos@va.gov

Note

1. The opinions expressed here are those of the authors and do not represent those of the Department of Veterans Affairs.

Certificates Should Be Strengthened

AS ATTORNEYS PRACTICING IN THE AREA OF bioethics, and members of the RAND Corporation's Institutional Review Board, we read with interest the Policy Forum by L. M. Beskow *et al.* ("Certificates of Confidentiality and compelled disclosure of data," 14 November 2008, p. 1054). In our view, however, the authors overstated the implications of a recent appellate case involving Duke University Health System (DUHS), *State v. Bradley*, 179 N.C. App. 551 (2006). On the basis of *Bradley*, the authors concluded that "the protection Certificates [of Confidentiality] offer [to research data] is uncertain" and further suggest that "caution is warranted when representing the impact of a Certificate to potential research participants." We disagree.

The appellate court in *Bradley* did not address the scope of protection for research data provided by Certificates of Confidentiality. The court found it unnecessary to consider such protections because the defendant failed to show that the data were relevant and material to his defense. Because *Bradley* did not specifically address the protections offered by Certificates of Confidentiality, the decision establishes no precedent (one way or the other) on the issue. The only precedent we are aware of is that established by *People v. Newman*, 32 N.Y.2d 379 (1973), in which the New York Supreme Court upheld the authority of Certificates of Confidentiality. Notably, the *Bradley* decision never cited *Newman*.

In addition, we question the authors' suggestion that researchers should qualify promises of confidentiality to potential research participants. Doing so may actually serve to weaken a Certificate's protections. A qualification that confidential data will not be disclosed "except as required by law" may provide a subpoenaing party with grounds to argue that compelled disclosure was anticipated by the subjects and that enforcement of the subpoena, therefore, is consistent with the limited assurances given the subject under the Certificate (*1*). Guidance issued by the National Institutes of Health is in accord. NIH's Web site, under "Frequently Asked Questions on Certificates of Confidentiality," recommends that researchers "eliminate provisions in consent form templates that may be inconsistent with the Certificate protections (such as references to disclosures required by law, since the Certificate enables researchers to resist disclosures that would otherwise be compelled by law)" (*2*). Given the limited case law bearing on Certificates of Confidentiality, we also worry that, by qualifying confidentiality assurances in consent forms, researchers are potentially creating factual scenarios that, if ultimately reviewed by appellate courts, are more likely to result in unhelpful precedent affording only limited protection to Certificate holders.

PATRICK P. GUNN* AND SCOTT D. JOINER

Cooley, Godward, Kronish LLP, 101 California Street, 5th Floor, San Francisco, CA 94111–5800, USA.

*To whom correspondence should be addressed. E-mail: pgunn@cooley.com

References

1. M. Traynor, *Law Contemp. Prob.* **59**, 119 (1996).
2. National Institutes of Health, Office of Extramural Research, "Frequently asked questions about Certificates of Confidentiality" (http://grants1.nih.gov/grants/policy/coc/faqs.htm).

Response

EMPIRICAL EVIDENCE CONCERNING THE USE of Certificates and the strength of the protections they afford is scant, but available studies support several of Hermos and Spiro's points: Some investigators perceive the process of obtaining a Certificate as lengthy and bureaucratic, and at least one report suggests that investigators sometimes alter their research because of the time it takes to explain a Certificate during the consent process (*1*). Prospective participants may well find consent form language describing Certificates confusing or even alarming, rather than reassuring (*2*).

On the other hand, there are anecdotal reports of situations where a Certificate did work as one might expect (*1*). More generally, when a court is asked to decide whether to order disclosure of relevant material, the presence of a Certificate could tip the balance in favor of nondisclosure—or at least persuade the judge to limit the extent of disclosure, as occurred in the case we described.

Thus, despite the concerns identified in our Policy Forum and raised by Hermos and Spiro, we believe it is premature to call for Certificates' demise. Some of the problems might be addressed in ways other than avoiding Certificates: Streamlined application procedures could reduce burdens associated with obtaining a Certificate, and improved model consent form language could reduce research participants' confusion. The nature of such reforms, however, should be based on a systematic analysis of investigators' experiences with Certificates.

In contrast to Hermos and Spiro, Gunn and Joiner suggest that our concerns about Certificates go too far. We agree that Certificates could be an important tool to protect research data and that *North Carolina v. Bradley* (*3*) should not be interpreted too broadly. As Gunn and Joiner correctly note, the case does not provide legal precedent one way or the other on the protections a Certificate offers. It does, however, provide a number of important lessons about the real-world workings of a Certificate.

For example, we believe that forced disclosure of identifiable research data in order to judge its materiality would surprise many Certificate holders. Furthermore, the case suggests a potentially critical limitation in the protections Certificates offer, which is the clash with constitutional rights that may arise in a criminal case. *Bradley* differs substantially from *People v. Newman* (*4*) in that *Newman* involved the State seeking information for use in a criminal prosecution, whereas *Bradley* involved a criminal defendant asserting constitutional rights to access information allegedly material to his defense. Although the court resolved the dispute without deciding this constitutional question, other published court opinions demonstrate that no statute can guarantee the secrecy of documents from a defendant because the

Sixth Amendment right to a fair trial trumps a statutory privilege (*5*). Thus, we believe that recommended consent language overstates the protection a Certificate can offer.

We would like to correct Gunn and Joiner's impression that we suggested specific qualifications to recommended consent form language. We are concerned about how a Certificate's protections are described, but we believe that it is premature to suggest particular changes without more data on how Certificates actually perform.

LAURA M. BESKOW,* LAUREN DAME, E. JANE COSTELLO

Duke Institute for Genome Sciences and Policy, Center for Genome Ethics, Law, and Policy, Duke University, Durham, NC 27708, USA.

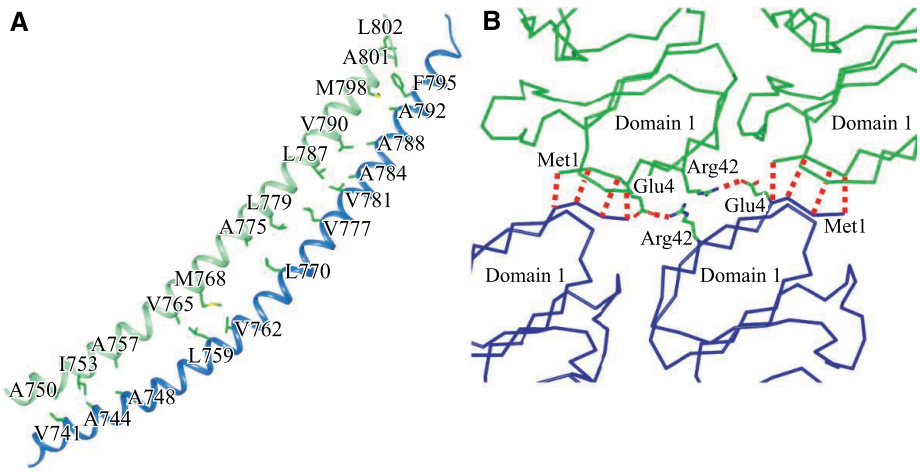
*To whom correspondence should be addressed. E-mail: laura.beskow@duke.edu

References and Notes

1. L. E. Wolf, J. Zandecki, *IRB* **28**, 1 (2006).
2. J. A. Catania *et al.*, *J. Empir. Res. Hum. Res. Ethics* **2**, 53 (2007).
3. *State of North Carolina v. Bradley*, 179 NC App 551, 634 SE2d 258 (2006).
4. *People v. Newman*, 32 NY2d 379, 298 NE2d 651, 345 NYS2d 502 (1973), *cert. denied*, 414 US 1163 (1973).
5. *Commonwealth v. Craig Neumyer*, 432 Mass 23 (2000).
6. The authors thank L. H. Glantz for his valuable input.

CORRECTIONS AND CLARIFICATIONS

Reports: "The structure of rat liver vault at 3.5 angstrom resolution" by H. Tanaka *et al*. (16 January, p. 384). In the print version, the bottom half of Fig. 4 lacked color. The figure is correct in the HTML and PDF online versions. The correct figure is also shown here.



Reports: "Biomass, size, and trophic status of top predators in the Pacific Ocean" by J. Sibert *et al.* (15 December 2006, p. 1773). In Table 1, three of the four values for estimates of fishery impact on the eastern Pacific Ocean (EPO) yellowfin tuna population are incorrect. The correct values are as follows: exploited/unexploited ratio total, 0.49; exploited/unexploited ratio adult, 0.38; biomass ratio total, 0.45; biomass ratio adult, 0.38. The corrected estimates change the conclusions with respect to the EPO yellowfin population but not the more general conclusion that the exploited/unexploited ratio is a more honest expression of the impact of the fishery than comparison with an arbitrary date in the past.

Published by AAAS

GEOSCIENCE

Many Worlds in Tiny Grains

Rob Holman

Sand is one of the most common materials on Earth, completely familiar like grass or trees, yet largely unknown. Asked what color sand is, most will respond "tan"; asked what sand is made of, some may say quartz but many will answer "sand." As with many things in nature, we often don't know what we don't know.

It is this large gap that Michael Welland tries to fill with *Sand*. Starting with a quote from the 2004 film *Eternal Sunshine of the Spotless Mind* that "Sand is overrated—it's just tiny little rocks," Welland (a London-based consulting geologist) sets out to demonstrate exactly the opposite. With obvious enthusiasm, he fills the reader with information and stories, anecdotes and details, and facts and folklore about sand. His accessible account ranges from grains to mountains, from the journey of a single particle to huge wind-blown dune fields on other planets.

Sand
The Never-Ending Story
by Michael Welland
University of California Press, Berkeley, 2009. 374 pp. $24.95, £17.95. ISBN 9780520254374.

There are many interesting stops along the way. The composition of our "tan" sand is found to include not just quartz, the hardiest of the minerals, but also an array of other grain types from sea shell fragments to volcanic minerals whose presence serves as a fingerprint for the source region (the geological provenance) of collected sands. We are introduced to sedimentary forensics. Many are familiar with the geological interpretation of sand and sedimentary patterns to learn about the history of Earth, for instance through the interpretation of bedding patterns to constrain the size of the early Atlantic Ocean or the role of uranium decay in zircon grains in constraining the age of the earliest presence of flowing water on our planet. But how many know that the launching point of Japanese helium balloon bombs during World War II was determined from the characteristics of the ballast sand or that sand and soil databases often play an important part in criminal investigations?

Although the strength of the book lies in its breadth, the weakness is its sometimes shallow depth. Welland's account can be frustrating for the specialist or curious scientist who seeks a 101-level understanding of processes. For example, the discussions of sediment transport and beach processes are typically light on details and sometimes misleading. Instead, this is an awareness book, opening readers' eyes to the many sides of sand, its importance to our quality of life, and its roles in history and even art. To aid those having deeper interests, the author provides an extensive bibliography.

The opening chapter, which describes the birth and characteristics of individual grains, is rich with enjoyable details about the formation and nature of sand particles. However, my clear favorite among the chapters is "Tribes: The Strange World of Granular Materials." It seems paradoxical that tiny sand grains, the epitome of simplicity, can often exhibit a collective behavior that is rich, complex, and frequently counterintuitive. Why do collections of sand grains abhor blandness, instead forming patterns of ripples, sand bars, and sand dunes under the action of wind or waves? Why is it likely that a mixture of large and small grains will unmix by the processes of movement (something known as the Brazil nut effect, reflecting the sorting of mixed nuts by vibration such as a rock concert)? Why would a meteorite remain on the surface of Saharan sand for thousands of years instead of gradually being buried? All of these phenomena are glimpses of the strange world of granular media, in turn glimpses into the newly discovered world of complex system behavior.

I would be remiss to not note Welland's extensive references to the links between sand and history and between sand and literature as well as to the broader connections of sand and society. But my focus was more on geology. The book wanders somewhat anecdotally into the role of sand in anthropology with the discovery of missing links like Dikika and Lucy, fossils preserved due to sudden sand burial. We learn about the importance of catastrophic events in Earth's history revealed through the sedimentary record and the role of tidal sedimentary deposits in revealing the slow lengthening of the day and the decreasing proximity of the Moon. We are told about the curious

Sand samples: Top to bottom: Rodeo Beach, Marin County, California; Agate Beach, Oregon; Daytona Beach, Florida; Bermuda; Santorini, Greece; Ayers Rock, Australia; Sahara Desert, Mauritania; Old Course Beach, St. Andrews, Scotland.

CREDIT: DAVE REINERT/OREGON STATE UNIVERSITY

The reviewer is at the College of Oceanic and Atmospheric Sciences, Oregon State University, 104 Ocean Administration Building, Corvallis, OR 97331–5503, USA. E-mail: holman@coas.oregonstate.edu

Published by AAAS

need for just the right water content for carving sand sculptures. And one chapter includes a wide-ranging summary of the many roles of sand in everyday life, from its contribution to concrete through its provision of the primary ingredient of computer chips to its service as a catalyst in relaxation therapy. At times the connections to the core topic of sand are tenuous and the material reads like a Google listing, but there are gems throughout.

Overall, Welland's *Sand* offers a broad introduction to its subject and a useful roadmap for deeper reading by anyone with a curiosity about the world around them. And the book certainly shows that small-but-mighty sand is not just tiny little rocks.

10.1126/science.1170842

AUTOBIOGRAPHY

Portrait of the Artist

Thoru Pederson

Many great scientists leave only their work, without commentary on it or themselves. *Also sprach F = ma* and $E = mc^2$. Newton spent the productive phase of his life out of the limelight, and only 253 years after his death did a truly definitive biography (*1*) appear. Einstein became a beloved public figure late in life, but it was for Abraham Pais to fully capture him (*2*). And despite authoritative and engaging biographies of Lise Meitner (*3*) and Rosalind Franklin (*4*), we will always long to have known them in their own words. So additions to the canon of scientists' autobiographies [e.g., (*5*–*7*)] are worth celebrating.

The Art and Politics of Science

by Harold Varmus

Norton, New York, 2009. 337 pp. $24.95, C$27.50. ISBN 9780393061284.

It is challenging to write simultaneously both for peers and a broader readership. As we have seen (consider Carl Sagan), the guild can react with wrath to those who popularize too much, and yet a failure to reach the broader audience (if the intent) obviously mars the product. Virologist and president of Memorial Sloan-Kettering Cancer Center Harold Varmus gives us as handsome an example of the chimeric genre of scientific-philosophical autobiography as can be imagined.

The Art and Politics of Science is organized as a natural triad, coursing from the author's codiscovery of the proto-oncogene, his years as director of the U.S. National Institutes of Health, and his current outlook on global science policy. Each of these domains will richly inform readers less familiar with Varmus's career and be pleasurably informative to those who know more about him. And all readers will enjoy the life-changing transitions that the author recounts, typically with recollections of the innocence he felt at the time. When he decides to try medical school after a year doing graduate work in English literature (in which he had majored at college), the Harvard Medical School admissions officer suggests military training instead. This vignette, like others during his formative stages, conveys the difficulty some had in believing that his saltatory path was more than idle vacillation. As became manifestly clear, the young Varmus's testings of various fields (literature, medicine, and psychiatry before bacterial gene regulation and then virology) were more a reflection of his agile mind's passions than acts of rebellion. During a European trip the summer after college graduation, in pursuit of the historical halls of literary icons prior to starting graduate school, Varmus accompanied a former classmate to a biochemistry congress in Moscow, which turned out to be highly influential. Not brash, just profoundly curious. (This story especially appealed to me because I have always had a pipe dream of sneaking off to the annual meeting of the Modern Language Association.)

Subsequent catalytic events, e.g., listening to a Columbia biochemist's lecture on hemoglobinopathies, are portrayed with deep appreciation and yet none of the soppy "If I hadn't had that conversation with (insert Sydney Brenner, Peter Medawar, *et al.*), I would never have gotten anywhere in science" passages that scientists paint excessively in their memoirs. Varmus's style of writing is very much like his personal and oratorical demeanor—an economy of words, an abundance of insight, and a dose of self that is serene and balanced.

The author's early research years at NIH (working on cyclic adenosine monophosphate–regulated gene expression in *Escherichia coli*); how he came to do a postdoc with Michael Bishop at the University of California, San Francisco (and why that was such a special place at the time); and the scientific setting for their discovery of proto-oncogenes are luminously imparted. Without sounding grandiose, Varmus conveys their courage in questioning the existing paradigm and pushing ahead with open minds (science's greatest tool). The unique bond the author and Bishop formed, as co-equal partners in their achievements in the molecular biology of retroviruses, is well known to peers but will be fascinating reading for others, as will the question of whether the pair's extreme devotion to baseball ever came close to a clinical diagnosis and mental hospital admission.

The author's years as NIH director are a vivid memory for many scientists, at least in the biomedical field, because he wrote and spoke so eloquently (with his testimony before congressional committees often appearing on television). Varmus writes with verve yet humility about how he got the job (not exactly a straight path), the landscape he met on arrival, and what is at stake when testifying before Congress (viz, the public trust). In my opinion, Varmus earned the wide respect of Congress more quickly than his predecessors in the modern era, and he was a champion of getting his positions into the mainstream—perhaps his greatest achievement beyond the affairs of NIH intramural and extramural programs.

In the closing chapters, Varmus argues for what could be called "the habit of humane truth" about issues such as stem cell research and global healthcare. His key emphasis is on the use of politics—more in the literal "people" sense than in the sense of the legislative process—to illuminate the paths that will help the most people. The author sees transparency as crucial to scientific and medical progress, and he believes unwaveringly that making people believe that such progress is possible is an art. In the book, Varmus explains this art with penetrating wisdom and eloquence. As he continues to practice it, further advances are likely.

References

1. R. S. Westfall, *Never at Rest: A Biography of Isaac Newton* (Cambridge Univ. Press, Cambridge, 1980).
2. A. Pais, *Subtle Is the Lord: The Science and the Life of Albert Einstein* (Oxford Univ. Press, Oxford, 1982).
3. R. L. Sime, *Lise Meitner: A Life in Physics* (Univ. California Press, Berkeley, 1996).
4. B. Maddox, *Rosalind Franklin: The Dark Lady of DNA* (HarperCollins, London, 2002).
5. F. Jacob, *The Statue Within* (Basic, New York, 1988).
6. J. E. Murray, *Surgery of the Soul* (Science History, Canton, MA, 2001).
7. E. R. Kandel, *In Search of Memory* (Norton, New York, 2006).

The reviewer is at the Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 377 Plantation Street, Worcester, MA 01605, USA. E-mail: thoru.pederson@umassmed.edu

10.1126/science.1170525

CREDIT: RICK KOZAK

SCIENTIFIC INTEGRITY

Responding to Possible Plagiarism

Documenting reactions from authors and journal editors to plagiarism may help others address the problem.

Tara C. Long,[1] Mounir Errami,[2] Angela C. George,[1] Zhaohui Sun,[2] Harold R. Garner[1,2]*

The peer-review process is the best mechanism to ensure the high quality of scientific publications. However, recent studies have demonstrated that the lack of well-defined publication standards, compounded by publication process failures (*1*), has resulted in the inadvertent publication of several duplicated and plagiarized articles.

The increasing availability of scientific literature on the World Wide Web has proven to be a double-edged sword, allowing plagiarism to be more easily committed, while simultaneously enabling its simple detection through the use of automated software. Unsurprisingly, various publishing groups are now taking steps to reinforce their publication policies to counter the fraudulent acts of a few (*2*). There are now dozens of commercial and free tools available for the detection of plagiarism. Perhaps the most popular programs are iParadigm's "Ithenticate" (http://ithenticate.com/) and TurnItIn's originality checking (http://turnitin.com/), which recently partnered with CrossRef (http://www.crossref.org/) to create CrossCheck, a new service for verifying the originality of scholarly content. However, the content searched by this program spans only a small sampling of journals indexed by MEDLINE. Others include EVE2, OrCheck, CopyCheck, and WordCHECK, to name a few.

We recently introduced an automated process to identify highly similar citations in MEDLINE (*3*, *4*). Our detection of duplicates relies heavily on human inspection in conjunction with computational tools including eTBLAST (*5*, *6*) and Déjà vu, a publicly available database (*7*, *8*). As of 20 February 2009, there were 9120 entries in Déjà vu with high levels of citation similarity and no overlapping authors. Thus far, full-text analysis has led to the identification of 212 pairs of articles with signs of potential plagiarism. The average text similarity between an original article and its duplicate was 86.2%, and the average number of shared references was 73.1%. However, only 47 (22.2%) duplicates cited the original article as a reference. Further, 71.4% of the manuscript pairs shared at least one highly similar or identical table or figure. Of the 212 duplicates, 42% also contained incorrect calculations, data inconsistencies, and reproduced or manipulated photographs.

There has been a paucity of literature examining the reactions of stakeholders (both victims and perpetrators) when confronted with evidence of possible misconduct. Studying these reactions may help to illuminate the reasons for such misconduct and might provide a way for the scientific community to prevent such activity in the future. Therefore, we merged data from previous studies (*3*) with additional information based on our personal communications with authors and journal editors directly associated with 163 of these cases of potential plagiarism.

A questionnaire (see table S1) was composed, supplemented with annotated electronic copies of both manuscripts, and sent via e-mail to the authors and editors of the earlier and later manuscripts.

From the 163 sets of questionnaires sent, we received a reply in 144 cases (88.3%). Anonymity was guaranteed to all respondents. The reactions by the respondents were intense and diverse, and although it is difficult to quantify the various responses, a general picture can be painted. Before receiving the questionnaire, 93% of the original authors were not aware of the duplicate's existence. The majority of these responses were appreciative in nature. The responses from duplicate authors were more varied; of the 60 replies, 28% denied any wrongdoing, 35% admitted to having borrowed previously published material (and were generally apologetic for having done so), and 22% were from coauthors claiming no involvement in the writing of the manuscript. An additional 17% claimed they were unaware that their names appeared on the article in question. The journal editors primarily confirmed receipt and addressed issues involving policies and potential actions. Excerpts from statements made by authors and editors illustrate the many possible perspectives in response to evidence of possible plagiarism. Table 1 provides a sampling of these responses, with an expanded list available in tables S2 to S5.

Although the goal of the questionnaire was merely to solicit information, the very act of sending it appeared to trigger further action by journals in many cases. Editors have launched 83 internal investigations thus far, 46 of which have, according to the editors of the journals, led to eventual retraction of the duplicate article. It is unclear what defines a "retraction," however, because many editors only stated that a comment would be published in their journal, or that the article would simply be removed from the journal's Web site. Unfortunately, these actions do not propagate back to MEDLINE unless an explicit request is made by the journal; therefore, researchers and clinicians may never become aware of an article's retracted status.

To assess how articles of this nature affect the scientific community, we recorded the impact factors for each journal in which the 212 articles and their duplicates were published using the Thomson Scientific Journal Citation Reports feature (*9*). A large portion of the duplicates were published in low-profile journals; thus, impact factors were available for only 199 of the 285 different journals. The impact factors of journals publishing original articles were significantly higher ($P < 0.001$), averaging 3.87 and spanning 0.147 to 52.589, than those of the journals publishing duplicate articles, averaging 1.6 and spanning 0.272 to 6.25.

Utilizing the ISI Web of Knowledge to determine how many times each article had been cited (*10*), we found that original publications were cited 28 times on average, whereas their corresponding duplicates were cited only twice. Although the original articles are older and have thus had more exposure, in 10 of the pairs, the duplicate article was cited at least as often as the original publication. This may be because scientists rely heavily on finding information through PubMed searches which, by default, return more recent articles first, ensuring that a plagiarized article will always appear higher on a list of search

[1]McDermott Center for Human Growth and Development, The University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390–9185, USA. [2]Division of Translational Research, The University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390–9185, USA.

*Author for correspondence. E-mail: harold.garner@utsouthwestern.edu

results than its original counterpart. As a result, citations that would have otherwise gone to an original publication are instead diverted to a plagiarized one.

Of the 175 journal editors with whom we communicated, 11 admitted they had never personally dealt with a potentially plagiarized manuscript and were unsure how to proceed. The majority of these editors showed deep concern and were open to any helpful suggestions or recommendations we could offer, at which point we directed them to the Office of Research Integrity's guidance document for editors on *Managing Allegations of Scientific Misconduct* (*11*). In spite of this concern, nearly half of all the duplications brought to light by our questionnaires have received no action. In fact, on 12 separate occasions, editors specifically indicated that cases involving their journal would not be reviewed. This variation in feedback reveals a great deal about the attitudes and motivations of scientists around the globe, including why some journal editors do not pursue obvious cases of duplication. Some apparently do not want to deal with the added stress of conducting a thorough investigation. Others feel it may bring bad publicity or reflect poorly on their journal's review process.

Sampling of responses from authors and editors

Authors of earlier article

"I have been a research scientist for more than 50 years, and this is the first time I've ever experienced such a blatant case of plagiarism. It sure was an eye-opener!"

"I have no statement. I cannot prove that this is plagiarism. Even if it is, what can be done?"

"[My] major concern is that false data will lead to changes in surgical practice regarding procedures."

"We were very sorry and somewhat surprised when we found their article. I don't want to accept them as scientists."

Editors of journal publishing earlier article

"It's my understanding that copying someone else's description virtually word-for-word, as these authors have done, is considered a compliment to the person whose words were copied."

The two articles" are the same patients, the figures are the same, and the writing is blatant plagiarism. One of these papers is a false publication. We cannot let this one go unaddressed."

"We were not aware of this duplicate publication, and would not have given permission for this, as it clearly violates copyright."

"I have been Editor for 14+ years and this is the first time this issue has been raised."

"It is clear that the subsequent author frankly, fraudulently used identical data ... in writing the second article. There is no way under the stars that we could have picked that up ourselves."

Authors of later article

"I would like to offer my apology to the authors of the original paper for not seeking the permission for using some part of their paper. I was not aware of the fact I am required to take such permission."

"There are probably only 'x' amount of word combinations that could lead to 'y' amount of statements. … I have no idea why the pieces are similar, except that I am sure I do not have a good enough memory—and it is certainly not photographic—to have allowed me to have 'copied' his piece.... I did in fact review [the earlier article] for whatever journal it was published in."

"I know my careless mistake resulted in a severe ethical issue. I am really disappointed with myself as a researcher."

"It was a joke, a bad game, an unconscious bet between friends, 10 years ago that such things … happened. I deeply regret."

"I was not involved in this article. I have no idea why my name is included."

Editors of journal publishing later article

"Looks like [the author of the later article] did it again in 2001. This example is a bit more embarrassing because the author of the original paper is [the] editor of the journal where [the author of the later article] published the copied work. Looks like we will have to publish two retractions."

"Believe me, the data in any paper is the responsibility of the authors and not the journal."

"I really appreciate your work and your e-mail has promoted us to exercise more strict control over duplicate publication."

"There can be no doubt that this is willful and deliberate plagiarism. Like the chance of monkeys typing out the works of Shakespeare, it would be incredible that the similarities could arise by chance."

"The news has taken us by surprise and a sense of deep concern. We are calling an emergency meeting of the editorial board to discuss the matter. [Our journal] deeply condemns the act and we stand firm to take necessary actions against the authors."

While there will always be a need for authoritative oversight, the responsibility for research integrity ultimately lies in the hands of the scientific community. Educators and advisors must ensure that the students they mentor understand the importance of scientific integrity. Authors must all commit to both the novelty and accuracy of the work they report. Volunteers who agree to provide peer review must accept the responsibility of an informed, thorough, and conscientious review. Finally, journal editors, many of whom are distinguished scientists themselves, must not merely trust in, but also verify the originality of the manuscripts they publish.

References and Notes

1. L. Gollogly, H. Momen, *Rev. Saude Publica* **40**, 24 (2006).
2. C. White, *BMJ* **336**, 797 (2008).
3. M. Errami *et al*., *Bioinformatics* **24**, 243 (2008).
4. Materials and methods are available as supporting material on *Science* Online.
5. J. Lewis, S. Ossowski, J. Hicks, M. Errami, H. R. Garner, *Bioinformatics* **22**, 2298 (2006).
6. M. Errami, J. D. Wren, J. M. Hicks, H. R. Garner, *Nucleic Acids Res*. **35**, W12 (2007).
7. M. Errami, H. Garner, Nature 451, 397 (2008).
8. M. Errami, Z. Sun, T. C. Long, A. C. George, H. R. Garner, *Nucleic Acids Res.* **37**, D921 (2009).
9. Journal Citation Reports, ISI Web of Knowledge (Thomson Reuters, Philadelphia, 2008); http://isiwebofknowledge.com/products_tools/analytical/jcr/.
10. ISI Web of Knowledge (Thomson Reuters, Philadelphia, 2008); http://isiknowledge.com.
11. Office of Research Integrity (ORI), *Managing Allegations of Scientific Misconduct: A Guidance Document for Editors* (ORI, U.S. Department of Health and Human Services, Rockville, MD, 2000); http://ori.dhhs.gov/documents/masm_2000.pdf.

12. We thank D. Trusty for computer administrative support; J. Loadsman as a substantial contributing curator; W. Fisher for useful comments, discussions, and manuscript editing; D. Wu and W. Fisher for assistance in obtaining full text articles; L. Gunn for administrative assistance; and the numerous Déjà vu users who have reported inaccuracies or have alerted us to questionable publications. This work was funded by NIH grant 5R01LM009758-02, the Hudson Foundation, and the P. O'B. Montgomery Distinguished Chair.

Supporting Online Material
www.sciencemag.org/cgi/content/full/323/5919/1293/DC1

10.1126/science.1167408

NEUROSCIENCE

More AMPAR Garnish

Two protein families control excitatory synaptic transmission by receptors for the neurotransmitter glutamate.

Cezar Tigaret and Daniel Choquet

The AMPA subtype of glutamate receptors (AMPARs) are ligand-gated cation channels that mediate the majority of fast excitatory synaptic transmission in the vertebrate central nervous system. Changes in the functional properties and number of AMPARs are central to the processes of synaptic plasticity that underlie learning and memory (*1*). Traditionally, an AMPAR was seen as a heterotetramer of GluA1 to GluA4, pore-forming subunits that determine the biophysical properties and intracellular trafficking of the receptor (*2*). This model has been challenged by the discovery of a family of transmembrane AMPAR regulatory proteins (TARPs) (*3*), auxiliary AMPAR subunits that control channel gating and trafficking. On page 1313 of this issue, Schwenk *et al.* (*4*) reveal that members of another distinct family of transmembrane proteins—cornichon homologs CNIH-2 and CNIH-3—also associate with AMPARs and, akin to TARPs, increase the expression of AMPARs and slow the kinetics of their deactivation and desensitization.

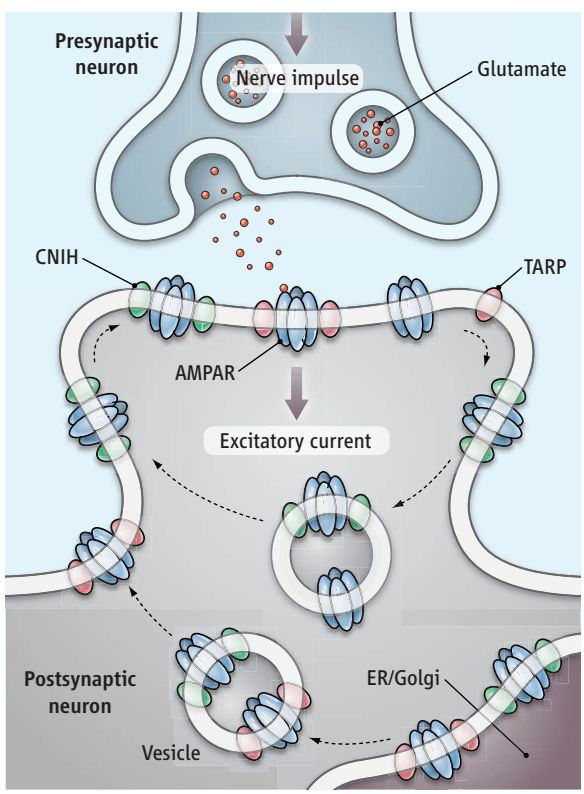
Schwenk *et al.* combined affinity purification of AMPAR complexes (from rat brain) with mass spectrometry to identify CNIH-2 and CNIH-3 as proteins tightly associated with GluA subunits. Surprisingly, they found that the majority (~70%) of AMPARs are in a complex with CNIH-2 and CNIH-3 rather than with TARPs, which suggests that there are distinct TARP- and CNIH-associated AMPAR populations with little overlap between them.

Cornichon proteins have been studied primarily in the fly *Drosophila melanogaster* and the yeast *Saccharomyces cerevisiae*. They have a cytoplasmic N terminus and three transmembrane domains. In *Drosophila*, cornichon is involved in embryo polarization by the protein Gurken (*5*). In yeast, the cornichon-related protein Erv14 is an integral component of the coat protein complex II (COPII)–coated vesicles that export most transmembrane proteins from the endoplasmic reticulum (*6*). In the vertebrate central nervous system, cornichon assists the transport and secretion of ligands for the receptor ErbB4 (*7*). Schwenk *et al.* found that CNIH-2 and CNIH-3 are present in both the synaptic and extrasynaptic regions of the plasma membrane of the postsynaptic neuron and are expressed throughout the rat central nervous system, except for cerebellar granule cells, where synaptic transmission by AMPARs depends on stargazin/TARP γ-2 (*8*).

In line with the role of cornichons in protein trafficking, CNIH-2 and CNIH-3 increase AMPAR expression at the neuronal cell surface. This is similar to TARPs, which participate in the early stages of AMPAR biosynthesis in the endoplasmic reticulum (*3*, *9*) (see the figure). However, cornichons and TARPs are structurally unrelated. All TARPs share a tetrameric membrane-spanning structure, with cytosolic N and C termini, and thus differ from cornichon. TARPs contribute to the density (representing the transmembrane region) of native AMPAR complexes seen by cryoelectron microscopy (*10*). It will be interesting to see whether cornichons fit these density maps.

More surprisingly, CNIH-2 and CNIH-3 slow down receptor deactivation and desensitization kinetics, and introduce a prominent nondesensitizing steady-state current component that could be interpreted as a stabilization of the open state of AMPARs. The functional consequence is to increase the total charge transfer through AMPARs during synaptic transmission. Upon glutamate binding, AMPARs open within less than a millisecond, and a large fraction of them rapidly proceed to a desensitized state where they can remain for tens of milliseconds. The others return to the basal closed state through deactivation. TARPs also slow AMPAR gating, albeit less strongly than does cornichon (*3*, *11*). Identifying the relative contributions of desensitization and deactivation in tuning fast synaptic transmission will thus require characterization of AMPARs' complex composition at specific synaptic sites.

Although both TARPs and cornichons affect AMPAR gating, TARPs particularly enhance the recovery from desensitization and increase the potency of the partial agonist kainate relative to glutamate (*3*). The latter remains to be tested for cornichons. TARPs' effects rely on their first extracellular loop (*10*, *11*). Previous work has suggested that not all AMPARs associate with TARPs and that agonist binding can induce rapid TARP-AMPAR dissociation (*12*, *13*). It will be of interest to determine how structurally distinct proteins mediate similar effects on AMPAR gating and whether the cornichon-AMPAR association is similarly regulated. Another key function of TARPs is the control of AMPAR stabilization at synapses. The cytoplasmic C-terminal tail contains a conserved motif that binds to PDZ domains of postsynaptic proteins (such as PSD-95) (*8*, *14*). The absence of these features in cor-



The AMPAR auxiliary. The pore-forming AMPA receptor (AMPAR) subunits associate with different auxiliary subunits, TARP and CNIH, in vertebrates. TARP and CNIH assist receptor biosynthesis in the endoplasmic reticulum (ER) and Golgi apparatus and promote export to the membrane. AMPAR complexes then diffuse to synaptic sites. Glutamate released from the presynaptic neuron binds to AMPAR and can induce dissociation of the AMPAR-TARP complexes, perhaps allowing diffusion of AMPARs out of the synapse to enter the endocytic-recycling pathway.

CREDIT: C. BICKEL/*SCIENCE*

UMR 5091 CNRS, Université de Bordeaux, 146 rue Léo Saignat, 33077 Bordeaux Cedex, France. E-mail: dchoquet@u-bordeaux2.fr

Published by AAAS

nichon intracellular domains suggests a major functional difference with TARPs.

Genetic screens in the nematode *Caenorhabditis elegans* has also led to the identification of SOL-1 as another AMPAR auxiliary subunit (*15*). *Sol-1* encodes a CUB-domain transmembrane protein that is required for the glutamate receptor GLR-1 currents. The identification of STG-1 and STG-2 as TARP homologs in *C. elegans* (*16*) demonstrated that the role of TARPs as obligate auxiliary subunits for AMPARs is evolutionarily conserved. Vertebrate SOL-1–like molecules that contribute to AMPAR function have not yet been identified. The future identification of other molecules that interact with the various ionotropic glutamate receptors will not be a surprise.

The differential properties of cornichons and TARPs in terms of binding to intracellular partners may enable exquisitely fine tuning of AMPAR trafficking and stabilization at the synapse. Functional studies will clarify how synaptic activity regulates TARP- and CNIH-mediated AMPAR trafficking.

References and Notes

1. J. D. Shepherd, R. L. Huganir, *Annu. Rev. Cell Dev. Biol.* **23**, 613 (2007).
2. R. Dingledine, K. Borges, D. Bowie, S. F. Traynelis, *Pharmacol. Rev.* **51**, 7 (1999).
3. A. D. Milstein, R. A. Nicoll, *Trends Pharmacol. Sci.* **29**, 333 (2008).
4. J. Schwenk *et al.*, *Science* **323**, 1313 (2009).
5. C. Bokel, S. Dass, M. Wilsch-Brauninger, S. Roth, *Development* **133**, 459 (2006).
6. G. A. Castillon, R. Watanabe, M. Taylor, T. M. Schwabe, H. Riezman, *Traffic* **10**, 186 (2009).
7. H. Hoshino *et al.*, *Mol. Biol. Cell* **18**, 1143 (2007).
8. L. Chen *et al.*, *Nature* **408**, 186 (2000).
9. I. H. Greger, E. B. Ziff, A. C. Penn, *Trends Neurosci.* **30**, 407 (2007).
10. T. Nakagawa, Y. Cheng, E. Ramm, M. Sheng, T. Walz, *Nature* **433**, 545 (2005).
11. A. D. Milstein, W. Zhou, S. Karimzadegan, D. S. Bredt, R. A. Nicoll, *Neuron* **55**, 905 (2007).
12. M. Morimoto-Tomita *et al.*, *Neuron* **61**, 101 (2009).
13. S. Tomita, M. Fukata, R. A. Nicoll, D. S. Bredt, *Science* **303**, 1508 (2004).
14. C. Bats, L. Groc, D. Choquet, *Neuron* **53**, 719 (2007).
15. Y. Zheng, J. E. Mellem, P. J. Brockie, D. M. Madsen, A. V. Maricq, *Nature* **427**, 451 (2004).
16. R. Wang *et al. Neuron* **59**, 997 (2008).

17. We thank the Choquet lab and F. Coussen for comments. Supported by an ATER fellowship from the University of Bordeaux 2 (C.T.).

10.1126/science.1171519

ECONOMICS

Eyes on the Prize?

Is there an optimal cost-benefit scheme for spurring innovation?

David K. Levine

There is widespread agreement that innovation is the foundation of our current and future economic well-being. Conventionally, patents have been seen as a key component of government policy to encourage innovation. This idea has been challenged by modern research showing that patents may be unnecessary and that there may be better policies for encouraging innovation and fostering a climate for robust economic development (*1*). On page 1335 in this issue, Meloso *et al*. report on experiments that examine alternative innovation policies (*2*). Though the study lends important insight, it does not yield a clearcut answer.

The economic experiments of Meloso *et al.* involved solving a difficult "knapsack" problem that required participants to combine objects for placement into bins. The goal was to compare the reward system—market versus patent—with respect to discovery of the best solution. The market system allowed innovators to trade shares in the objects, so that each item in the winning combination gets a prize. By contrast, in the patent system, the prize goes to the best solution. In these experiments, the market system did better, and so the study lays the groundwork for future experimental investigations of alternative compensation systems for inducing innovation. But do the results reinforce the view that patent systems are generally undesirable for inducing innovation? Not necessarily.

The first caveat to the study by Meloso *et al.* is that twice as much time was allowed in the market system to permit time for trading. The results are less strong when the allotted time is the same in both the market and patent systems. Because time spent trading is a substitute for time spent innovating, it seems as though the equal time treatments are the most relevant.

There are two deeper issues, however. For instance, the objects that are to be used are well defined in the experiment. It is less clear in practice how claims to part of a successful innovation might be defined and rewarded. As Meloso *et al*. point out, this means the market mechanism is more relevant in some contexts than others. Also, in these experiments, the rewards were simply payments from the experimenter. In practice, either there must be a government system of prizes or else revenue must be derived somehow from sale of the final product. The patent system uses the latter scheme; how it might be applied to payments for "pieces of the puzzle" in a market system, and whether such a system would induce the same monopoly distortions as a patent system, are uncertain.

Consequently, these results are not best interpreted as a market system versus a patent system but rather as alternative systems for the award of government prizes. The study by Meloso *et al.* is especially relevant to earlier research (*3*, *4*) proposing that prizes replace patents for pharmaceuticals. Thus, the results of Meloso *et al.* suggest that a system of traded partial prizes for pieces of the final product would be better than simply awarding a prize to the winner.

Where do the observations lead us? If we knew how costly innovations were and how valuable they were, it would not be difficult to design an efficient compensation scheme. The difficulty is that both the cost and the benefit are private information, and there can be substantial disagreement over either or both. Indeed, one proposed benefit of a patent system is that it encourages foolishly optimistic inventors to overinvest their time and effort, bringing great benefits to the rest of us (*5*).

The experiments by Meloso *et al.* focus on the lack of knowledge about the cost of innovating. That is, the way in which the objects should be combined is unknown to the innovators, although each has beliefs about the correct solution. This problem is addressed both by the patent system and by the market system. The second, also important, type of pri-

Department of Economics, Washington University in St. Louis, St. Louis, MO 63130–4899. E-mail: david@dklevine.com

CREDIT: PETER HOEY

vate information is about the benefit of the innovation. These experiments eliminate that uncertainty by promising a known fixed payment for the best solution or for some part of it. Left open is how in practice the overall size of the reward is to be determined, or how it is to be allocated to different pieces. (In the experiment, each piece was allocated an equal share of the prize; in practice, some pieces are more difficult than others, so would ideally command a larger prize.) The patent system deals with this second source of uncertainty by tying the reward to the profit that can be obtained by the monopoly granted by the patent. There are various methods of tying the reward to sales and price in the final goods market, as well as tapping other information about social value, but these need to be explicitly compared.

The standard economic approach to private information problems is mechanism design theory, which asks what is the best social outcome achievable for a given information structure. This theory has not yet been widely applied in the study of innovation. A good example of how to do so can be found in Hopenhayn and Mitchell (*6*). In addition to theoretical research, the experimental laboratory seems a good place to build on the research by Meloso *et al.* to investigate the second half of the innovation puzzle: How should an innovation be valued?

References

1. M. Boldrin, D. Levine, *Against Intellectual Monopoly*, (Cambridge University Press, 2008).
2. D. Meloso, J. Copic, P. Bossaerts, *Science* **323**, 1335 (2009).
3. M. Kremer *Q. J. Econ.* **113**, 1137 (1998).
4. M. Kremer "Creating Markets for New Vaccines," NBER Working Paper No. W7716 (National Bureau of Economic Research, Cambridge, MA, 2000).
5. R. E. Lerner, J. Mokyr, "Intellectual Property Rights, the Industrial Revolution, and the Beginnings of Modern Economic Growth," presented at *American Economic Association Annual Meetings,* San Francisco, CA, January 2009.
6. H. A. Hopenhayn, M. F. Mitchell, *RAND J. Econ.* **32**, 152 (2001).

10.1126/science.1171406

COMPUTER SCIENCE

Beyond the Data Deluge

The demands of data-intensive science represent a challenge for diverse scientific communities.

Gordon Bell,[1] Tony Hey,[1] Alex Szalay[2]

Since at least Newton's laws of motion in the 17th century, scientists have recognized experimental and theoretical science as the basic research paradigms for understanding nature. In recent decades, computer simulations have become an essential third paradigm: a standard tool for scientists to explore domains that are inaccessible to theory and experiment, such as the evolution of the universe, car passenger crash testing, and predicting climate change. As simulations and experiments yield ever more data, a fourth paradigm is emerging, consisting of the techniques and technologies needed to perform data-intensive science (*1*). For example, new types of computer clusters are emerging that are optimized for data movement and analysis rather than computing, while in astronomy and other sciences, integrated data systems allow data analysis and storage on site instead of requiring download of large amounts of data.

Today, some areas of science are facing hundred- to thousandfold increases in data volumes from satellites, telescopes, high-throughput instruments, sensor networks, accelerators, and supercomputers, compared to the volumes generated only a decade ago (*2*). In astronomy and particle physics, these new experiments generate petabytes (1 petabyte = 10^{15} bytes) of data per year. In bioinformatics, the increasing volume (*3*) and the extreme heterogeneity of the data are challenging scientists (*4*). In contrast to the traditional hypothesis-led approach to biology, Venter and others have argued that a data-intensive inductive approach to genomics (such as shotgun sequencing) is necessary to address large-scale ecosystem questions (*5*, *6*).

CREDIT: JONATHAN FAY/MICROSOFT

Moon and Pleiades from the VO. Astronomy has been one of the first disciplines to embrace data-intensive science with the Virtual Observatory (VO), enabling highly efficient access to data and analysis tools at a centralized site. The image shows the Pleiades star cluster form the Digitized Sky Survey combined with an image of the moon, synthesized within the World Wide Telescope service.

Other research fields also face major data management challenges. In almost every laboratory, "born digital" data proliferate in files, spreadsheets, or databases stored on hard drives, digital notebooks, Web sites, blogs, and wikis. The management, curation, and archiving of these digital data are becoming increasingly burdensome for research scientists.

Over the past 40 years or more, Moore's Law has enabled transistors on silicon chips to get smaller and processors to get faster. At the same time, technology improvements for disks for storage cannot keep up with the ever increasing flood of scientific data generated by the faster computers. In university research labs, Beowulf clusters—groups of usually identical, inexpensive PC computers that can be used for parallel computations—have

[1]Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA. [2]Department of Physics and Astronomy, Johns Hopkins University, 3701 San Martin Drive, Baltimore, MD 21218, USA. E-mail: szalay@jhu.edu

become ubiquitous. However, these cluster computing systems have limited connection to disks and lack database software. Scientists and computer scientists must now develop similarly cost-effective solutions for data-intensive research. Jim Gray was one of the first to anticipate this need. In 1995, he advocated building clusters of "storage bricks," consisting of inexpensive, balanced systems of central processing units, memory, and storage for data-intensive research (*7*).

We have realized such an architecture in the GrayWulf system (*8*), which is built out of commodity components like Beowulf clusters. However, unlike Beowulf clusters, which are optimized for computation, the GrayWulf design emphasizes high-speed access to data residing on each node of the cluster, which supports a large database system; its performance scales well as the number of nodes is increased. GrayWulf won the Storage Challenge at the SC08 conference (*9*) by executing a query on the Sloan Digital Sky Survey (SDSS) database in 12 minutes; the same task took 13 days on a traditional (nonparallel) database system.

The bandwidth of inexpensive, commodity computer networks is also falling behind the data explosion. Copying large amounts of experimental data from a data center to personal workstations or distributing data to numerous independent centers is no longer tenable without recourse to extreme—and thus expensive—networking solutions. For research to be affordable, data analysis must increasingly be done where data sets reside, leaving academic research networks to handle low-bandwidth queries and analytic requests, including visualization.

The urgency for new tools and technologies to enable data-intensive research has been building for a decade or more (*2*, *7*). In 2007, Jim Gray laid out his vision for a fourth research paradigm—data-intensive science—which he described as collaborative, networked, and data-driven (*1*, *10*). He defined eScience as the synthesis of information technology and science that enables challenges on previously unimaginable scales to be tackled.

Despite the enormous potential of this approach, data-intensive science has been slow to develop due to the subtleties of databases, schemas, and ontologies, and a general lack of understanding of these topics by the scientific community. For example, virtually all large-scale models use databases to organize the vast array of files that hold data from computational modeling, but these databases rarely hold any data: They only hold pointers to the files that hold data, making direct analysis impractical. Indeed, many areas of science lag commercial use and understanding of data analytics by at least a decade.

Astronomy has been among the first disciplines to undergo the paradigm shift to data-intensive science. The first step in this direction was made in 2001, when data from the SDSS were put into a publicly available database (*11*), with simple Web services offering the primary access to the multiterabyte (1 terabyte = 10^{12} bytes) data sets. Astronomers have embraced not only these services but have also frequently used the powerful Structured Query Language (SQL)—previously used almost exclusively by the financial and commercial sector—to gain direct access to data stored in a relational database. The site also offers an analysis workbench, where users can analyze data and store derived data sets next to the main database. About 15 to 20% of the world's professional astronomers now have their own server-side database, and the SDSS servers are running close to saturation.

Now, astronomers have gone even further in embracing data-intensive science. An international grassroots effort has gone a long way toward integrating all astronomical data (hundreds of terabytes today) into the Virtual Observatory (VO) (see the figure), of which the SDSS is an integral part. In the VO, data are accessible through services, and ontological and semantic information is stored with each data set (*12*); this information provides crucial support for data searching, analysis, and reuse by using a standard vocabulary agreed on by the community and by recording semantic information about the structure and type of data, as well as the instrument that generated the data. Most major astronomical data providers have adopted a standardized interface for services, and there is a registry to find particular data sets.

A similar transformation is happening in many sciences: For high-energy physics, the CERN Large Hadron Collider (LHC) is set to create an integrated data system resembling the VO; in genomics, NCBI (National Center for Biotechnology Information) and GenBank play this part. Many day-to-day issues are the same, whether we deal with astronomy or oceanography data. The emerging solution to these challenges lies in more diverse computing system architectures—like the GrayWulf system—that are specialized for highly data-intensive computations. Such systems will offer specialized data-analysis facilities located next to the largest data sets, coexisting with and complementary to today's supercomputers.

Data-intensive science will be integral to many future scientific endeavors, but demands specialized skills and analysis tools. In addition, the research community now has the option of accessing storage and computing resources on demand. The IT industry is building huge data centers, far beyond the financial scope of universities and national laboratories (*13*). These "cloud services" provide high-bandwidth access to cost-effective storage and computing services. However, there are no clear examples of successful scientific applications of clouds yet; making optimum use of such services will require some radical rethinking in the research community. In the future, the rapidity with which any given discipline advances is likely to depend on how well the community acquires the necessary expertise in database, workflow management, visualization, and cloud computing technologies.

References and Notes

1. J. Gray, A. Szalay, *eScience—A Transformed Scientific Method*, Presentation to the Computer Science and Technology Board of the National Research Council, Mountain View, CA, 11 January 2007; see http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt.
2. A. Hey, A. Trefethen, in *Grid Computing: Making the Global Infrastructure a Reality*, F. Berman, G. Fox, T. Hey, Eds. (Wiley, Chichester, UK, 2003), pp. 809–824.
3. See www.ncbi.nlm.nih.gov.
4. J. C. Wooley, Herbert S. Lin, Eds., *Catalyzing Inquiry at the Interface of Computing and Biology* (Report of the National Research Council of the National Academies, 2005), chap. 3.
5. J. C. Venter *et al.*, *Science* **304**, 66 (2004).
6. D. B. Kell, S. G. Oliver, *BioEssays* **26**, 99 (2004).
7. Jim Gray talked about the need for such a "brick architecture" in his Mackay Lectures at Berkeley as long ago as 4 October 1994, under the title "SNAP—Scalable Networks and Platforms," followed on 4 November 1995 by a lecture titled "Parallel Database Systems: A SNAP application." These lectures are available in the Internet Archive www.archive.org under www.research.microsoft.com/~gray.
8. A. S. Szalay *et al.*, *GrayWulf: Scalable Clustered Architecture for Data-Intensive Computing*, Proceedings of the 42nd Hawaii International Conference on System Sciences, Hawaii, 5 to 8 January 2009, paper no. 720; available as Microsoft Tech Report MSR-TR-2008-187 at http://research.microsoft.com/apps/pubs/default.aspx?id=79429.
9. See http://sc08.supercomputing.org/.
10. Jim Gray was one of the founders of modern database technology and transaction processing and received the Turing Award in 1998. Recognizing that the greatest future data challenges would come not from commercial database applications, but from the next generation of scientific experiments, he spent 10 years systematically exploring his "Fourth Paradigm," working with astronomers, biologists, oceanographers, and geographers. Jim was tragically lost at sea in January 2007 in the midst of his latest challenge.
11. J. Gray *et al.*, in *Distributed Data & Structures 4: Records of the 4th International Meeting*, W. Litwin, G. Levy, Eds. (Carleton Scientific, Waterloo, Ontario, 2003), pp. 189–210.
12. M. J. Graham, M. J. Fitzpatrick, T. A. McGlynn, Eds., *The National Virtual Observatory: Tools and Techniques for Astronomical Research*, ASP Conference Series **382** (Astronomical Society of the Pacific, Provo, UT, 2007).
13. R. Ozzie, Interview on Cloud Computing, *Wired* Magazine, December 2008.

14. We thank J. Fay (World Wide Telescope) for providing the figure.

10.1126/science.1170411

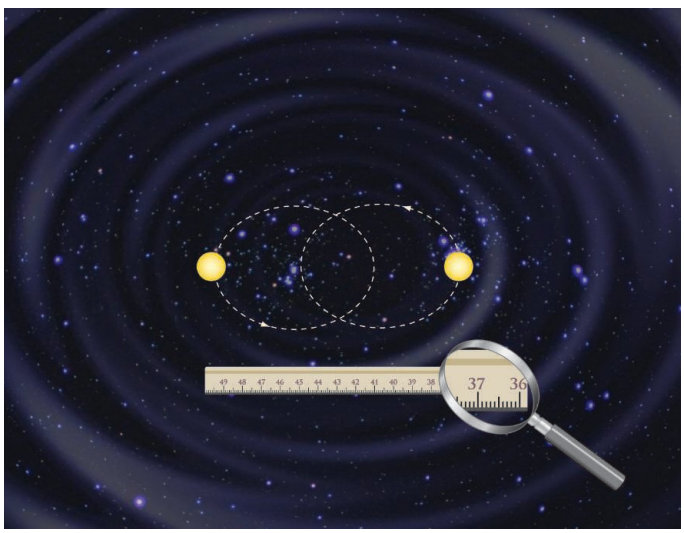
ASTRONOMY

Pinpointing Gravity

Accurate measurement of the distance to a double pulsar system will place stringent limits on theories of gravity.

Nicolò D'Amico

One of the consequences of Einstein's theory of gravity is that a pair of orbiting stars generates ripples in space—gravitational waves. While gravitational waves have not yet been detected directly, experiments are being developed with which the indirect effects of such phenomena can be measured with unprecedented accuracy. On page 1327 of this issue, Deller *et al.* (*1*) use very long baseline interferometry (VLBI) to determine an accurate distance to the double pulsar system J0737-3039 A/B, thereby pushing further the limits on the precision with which tests of general relativity can be made.



A dance of pulsars. As the two pulsars orbit around each other, gravitational waves are emitted. This emission results in energy loss and a secular shrinking of the binary system, whose decay law is precisely predicted by Einstein's theory of relativistic gravity. While the change in motion can be determined by timing techniques, other kinematic effects may mimic an orbital variation. VLBI, as shown by Deller *et al.*, quantifies these undesired effects, allowing accurate estimates of the true orbital decay, and thereby placing stringent limits on gravity theories.

Present large gravitational wave observatories, such as LIGO (*2*) and VIRGO (*3*), are sufficiently sensitive to be able to detect the burst of gravitational waves resulting from the merger of a pair of orbiting stars into a black hole at distances up to many tens of millions of light-years away. However, such events are rare, and these instruments might not detect the weak steady gravitational waves emitted by a binary system. At present, we can only observe the steady emission of gravitational waves from binary systems indirectly (see the figure). The intensity of the gravitational waves, and so the amplitude of the corresponding orbital-period decay, depend on the masses of the two bodies, the orbital period, and the eccentricity (*4*), thus implying that some binary systems are more relativistic than others.

Pulsars are rapidly rotating, highly magnetized neutron stars radiating collimated beams of radio waves, which we observe as pulses, like those produced by a lighthouse, once per rotation. Owing to the astonishing clock stability of the pulses, comparable to the best time standards available on Earth, a pulsar orbiting another star provides an accurate time reference for measuring the relativistic gravity effects such as the orbital decay due to the emission of gravitational waves. Indeed, the historical evidence of such orbital decay was provided by the binary pulsar PSR B1913+16 (*5*), the discovery of which earned Hulse and Taylor the Nobel Prize in 1993. Precision timing over decades has shown that the orbit of this system is systematically shrinking with a measured decay that agrees with the relativistic gravity prediction to an accuracy of 0.2%.

The demand to improve further on the accuracy of such measurements is rather high, because alternative gravity theories exist, whose measurable effects are marginally different from those predicted by relativistic gravity. Much more relativistic binary systems would therefore provide much more accurate tests. But if relativistic gravity is sufficient for most practical purposes, such as describing satellite tracking in the solar system (*6*), why should we care about alternative theories of gravity?

The point is that some alternative theories are related to fundamental models of the nature of matter, which in turn have major implications for cosmology and so for the origin, evolution, and fate of the universe. For instance, the alternative theories may better explain the large-scale structure of the universe, without necessarily invoking dark matter (*7*, *8*). Also, the attempt to unify all forces of nature might require modifications to the relativistic gravity equations, which seem ideal for testing by binary pulsars (*9*).

The discovery of the binary pulsar system PSR J0737-3039, the most relativistic found to date (*10*), provides a system in which the observed evolution of the orbital parameters can be measured with unprecedented accuracy. The system shrinks by 7 mm/day. The pulsar clock allows the orbital parameters to be measured with such accuracy that the galactic and kinematic contributions to the observed variation of the orbital period affect the estimate of the orbital decay. We need to exactly locate the system within the Galaxy, determining an accurate measure of its distance and proper motion. Using VLBI and pulsar timing observations, Deller *et al.* estimate that over a decade, the orbital decay of the system due to gravitational wave emission can be measured to an accuracy of ~0.01%.

In addition, the neutron star companion to the pulsar in this system has been found also to be a pulsar (*11*). So, we actually have two clocks orbiting each other, providing further constraints on the system parameters (*12*). As a result, relativistic gravity in this system is expected to be probed with much more accuracy than in the solar system environment. Indeed, the accuracy allowed by this system, and its coherence with the relativistic gravity equations, should enable the presence of other perturbing stars in the vicinity of the binary system to be probed, including a potential companion, as in the speculative triple-system scenario mentioned by Deller *et al.*

Besides the uncertainty in locating a binary system in the Galaxy, other quantities affect the estimate of the galactic and kinematic contributions to the observed variation of the orbital period. Uncertainties in the galactic radius and speed of the solar system are ~15%, while the uncertainty in the vertical gravitational potential of the Galaxy is ~10%. Further searches for pulsars are in

Observatorio Astronomico di Cagliari, National Institute for Astrophysics, Loc. Poggio dei Pini, Strada 54, Capoterra, I-09012 Italy. E-mail: damico@oa-cagliari.inaf.it

progress at the Parkes, Green Bank, and Arecibo observatories, and it is expected that even more relativistic binary pulsars will be found. With the advent of the Square Kilometer Array, a huge sample of relativistic binaries will be available (*13*). At the same time, VLBI techniques are rapidly evolving, which will provide an accurate location of these clocks within the Galaxy. A rich array of perfectly clocked shrinking binary systems, exactly located in different zones of the Galaxy, will be available, constituting a powerful gravity probe. As the present uncertainties in the gravitational potential of the Galaxy would be averaged by such an array, orbital decays and other non-Newtonian effects would be estimated with better accuracy, thus providing an unprecedented test of relativistic gravity. It would then pose challenging questions to those alternative theories of gravity.

References

1. A. T. Deller, M. Bailes, S. J. Tingay, *Science* **323**, 1327 (2009); published online 5 February 2009 (10.1126/science.1167969).
2. LIGO Scientific Collaboration: B. Abbott *et al.*, http://arxiv.org/abs/0711.3041v1 (2007).
3. A. Giazotto, S. Braccini, in *Proceedings of the 14th SIGRAV Conference on General Relativity and Gravitational Physics*, Genova, Italy, September 2000, (Springer, Berlin, 2002), pp. 111–119.
4. D. R. Lorimer, *Living Rev. Relativity* **11**, 8 (2008).
5. R. A. Hulse, J. H. Taylor, *Astrophys. J.* **195**, L51 (1975).
6. I. Ciufolini, J. A. Wheeler, *Gravitation and Inertia* (Princeton Univ. Press, Princeton, NJ, 1996).
7. S. Dodelson, M. Liguori, *Phys. Rev. Lett.* **97**, 231301 (2006).
8. T. Damour, G. Esposito-Farese, *Phys Rev. D* **54**, 1474 (1996).
9. N. Yunes, D. N. Spergel, http://arxiv.org/abs/0810.5541v1 (2008).
10. M. Burgay *et al.*, *Nature* **426**, 531 (2003).
11. A. G. Lyne *et al.*, *Science* **303**, 1153 (2004).
12. R. P. Breton *et al.*, *Science* **321**, 104 (2008).
13. R. Smits *et al.*, *Astron. Astrophys.* **493**, 1161 (2009).

10.1126/science.1170936

ECOLOGY

Warming Up Food Webs

How do predator-prey interactions influence ecosystem responses to climate change?

Jason M. Tylianakis

Human changes to the global environment have long been known to affect organisms, for example by altering their physiology, range, or longevity (*1*, *2*). However, responses vary widely across species, making it difficult to predict how entire ecosystems will respond in the future (*3*). A key problem is that species do not respond to extrinsic drivers (such as climate) in isolation. Rather, species responses may be determined to a greater or lesser extent by other species with which they interact. On page 1347 of this issue, Harmon *et al.* elucidate one such interaction in a study of pea aphids and two of their ladybird predator species (*4*).

Early population models showed that interactions among species could weaken or strengthen within-species responses to environmental change (*5*). More recently, empirical evidence has demonstrated that species interactions can reverse the response of individual grassland species to climate change and subsequently alter their community trajectory (*6*). At the same time, numerous studies have identified rapid evolutionary responses to climate change (*2*). For example, evolutionary studies indicate that under strong climate-induced selection pressure, life history traits (such as phenology, longevity, and reproductive rates) may evolve within just a few generations (*7*).

The question thus arises whether ecological interactions among species can alter their respective evolutionary responses to external drivers such as climate change. Harmon *et al*. have now met the enormous challenge of addressing this question with an elegant study in which they tested the ecological and evolutionary responses of an insect herbivore to the combined effects of climate change and food web interactions.

Their study system involved the pea aphid, an important pest species, which has populations that differ in their susceptibility to short periods of high temperature (heat shocks), in part because of differences in internal microorganisms that are passed on from parents to offspring and confer heat tolerance. Harmon *et al.* introduced these symbiotic microorganisms to different aphid populations to simulate a mutation event conferring heat tolerance to the aphid line. Aphids are attacked by many natural enemies, including two species of ladybirds that differ in their foraging behavior (see the figure). Harmon *et al.* used field experiments to test the ecological and evolutionary responses of aphids to increased frequency of heat shocks, and to contrast the effects of the two predators on population growth rates after these shocks.

The authors found that behavioral differences between the predator species affected the prey population response: One predator reduced its attack rates at low prey densities after heat shocks, such that it did not compound the negative effect of heat on aphid population growth. In contrast, the other predator kept attacking aphids at the same rate, increasing aphid mortality beyond the rates caused by the heat shocks. Aphids did evolve tolerance to heat shocks (that is, tolerant strains increased in frequency), but in a model based on field data, predator-prey interactions, despite their effects on aphid population growth, did not affect the evolution of heat shock tolerance.

Future research is needed across species to determine whether rapid evolutionary responses to food web interactions and external drivers are generally additive in this way, such that the selection effects of predation operate independently of climate change and vice versa. If in other cases these forces have interactive effects—for example, if traits conferring resistance to one threat (such as climate change) are negatively correlated

Interactive effects. Harmon *et al.* show that predation rates of aphids by ladybirds depend on the response of the predator species to altered aphid density following heat shocks.

CREDIT: ANTHONY R. IVES/UNIVERSITY OF WISCONSIN

School of Biological Sciences, University of Canterbury, Christchurch 8020, New Zealand. E-mail: jason.tylianakis@canterbury.ac.nz

with traits providing protection from another (such as predation)—then an evolutionary stalemate could prevent adaptation at the rate necessary for species to cope with predicted climatic changes.

In the study by Harmon *et al*., predator density dependence plays a key role in determining the effects of climate change on the growth of the prey population. This raises a complex issue for ecologists, because environmental disturbances strongly affect prey density. Quantitative food webs depict links between species as having a certain strength (the frequency of the interaction is often used as a surrogate), yet the interaction strength may be a nonlinear function of prey density. Altered interaction frequencies in food webs after environmental perturbations (*8*) could be driven by relative shifts in the abundance of different species and/or by nonlinear responses of different predators to altered prey densities. Environmental changes could even alter the response of predators to prey density—for example, if habitat simplification alters search efficiency (*9*). Metrics of food web structure that are advocated for their insensitivity to differing relative abundances of prey species (*10*) may not detect these ecologically important changes.

When Earth undergoes climatic change, species either adapt or go extinct (*11*). However, during each period of change, the spatial rearrangement of extant species and the emergence of new lineages will provide a novel context within which evolution takes place. Thus, even if the genetic variation necessary for selection exists in a population, food web interactions may impose constraints on the rate or direction of evolutionary change. The work of Harmon *et al.* suggests that the difficulties inherent in studying the combined complexities of ecological and evolutionary forces can be overcome. This will be necessary if we are to predict the survival of species in an uncertain future.

References and Notes

1. G. R. Walther *et al*., *Nature* **416**, 389 (2002).
2. C. Parmesan, *Annu. Rev. Ecol. Evol. Syst.* **37**, 637 (2006).
3. J. M. Tylianakis, R. K. Didham, J. Bascompte, D. A. Wardle, *Ecol. Lett.* **11**, 1351 (2008).
4. J. P. Harmon, N. A. Moran, A. R. Ives, *Science* **323**, 1347 (2009).
5. A. R. Ives, *Ecology* **76**, 926 (1995).
6. K. B. Suttle, M. A. Thomsen, M. E. Power, *Science* **315**, 640 (2007).
7. S. J. Franks, S. Sim, A. E. Weis, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1278 (2007).
8. J. M. Tylianakis, T. Tscharntke, O. T. Lewis, *Nature* **445**, 202 (2007).
9. R. K. Didham, J. M. Tylianakis, N. J. Gemmell, T. A. Rand, R. M. Ewers, *Trends Ecol. Evol.* **22**, 489 (2007).
10. N. Blüthgen, J. Fründ, D. P. Vazquez, F. Menzel, *Ecology* **89**, 3387 (2008).
11. R. Jansson, M. Dynesius, *Annu. Rev. Ecol. Syst.* **33**, 741 (2002).
12. Supported by Marsden Fund of New Zealand grant UOC-0705. This manuscript benefited from comments by R. K. Didham, T. A. Rand, and T. E. Steeves.

10.1126/science.1170909

PLANT SCIENCE

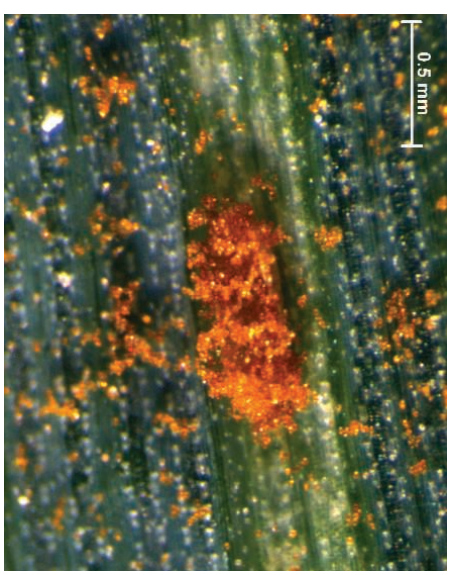
Anti-Rust Antitrust

The identification of genes that confer durable, multipathogen resistance may help breeders overcome devastating wheat fungal diseases.

Daniel J. Kliebenstein and Heather C. Rowe

Human civilization rests upon the products of agriculture, but numerous fungi, bacteria, and viruses compete with us for the crops that we cultivate. Given the growing human population and demand for increased food production from less land, minimizing crop loss from pathogen infection is critical. Two reports in this issue, by Fu *et al*. on page 1357 (*1*) and by Krattinger *et al*. on page 1360 (*2*), identify molecular mechanisms that control durable, broad-spectrum, disease resistance in wheat, the second-most-produced cereal crop in the world.

Understanding the molecular basis of specific plant resistances has enabled the development of crop varieties that resist common diseases. These resistance mechanisms follow a "gene-for-gene" model, whereby specific plant recognition of a pathogen gene product yields near-absolute resistance to pathogens containing the recognized gene. However, this form of resistance is vulnerable to rapid counter-evolution by the pathogen (*3*). This generates a continuous and resource-intensive cycle of plant resistance gene discovery, crop breeding, and eventual resistance gene "defeat" by the pathogen, which has driven interest in identifying durable forms of resistance to pathogens. In most plant-pathogen systems, durable resistance has been identified through quantitative trait locus (QTL) mapping. Although identified QTL provide quantitative resistance to a range of pathogen species and genotypes therein, the molecular bases of these broad quantitative resistances have remained largely unknown (*4*).



Wheat rust. Close-up of a wheat leaf rust pustule releasing orange urediospores (oranges spots) on a leaf of the susceptible spring wheat cultivar Thatcher.

CREDIT: SIMON KRATTINGER/UNIVERSITY OF ZURICH

Krattinger, Fu, and their colleagues begin to show that mechanisms underlying quantitative resistance are not simply weak alleles of genes involved in specific gene-for-gene resistance, as has been proposed. Rather, each plant species may contain multiple, possibly independent, mechanisms of quantitative resistance. The latter may reveal similar processes in other plants or may be taxonomically limited.

Krattinger *et al*. describe the cloning of the wheat *Lr34* QTL that has been used to confer resistance to multiple rusts (see the figure) and a mildew in the field for nearly 50 years (*5*). The locus harbors the gene *Lr34*, which encodes an ATP-binding cassette (ABC) transporter (*4*). ABC transporters, or pleiotropic drug resistance transporters, move diverse chemical compounds, including plant natural products, across membranes. The LR34 transporter is weakly homologous to PEN3, an ABC transporter from the plant *Arabidopsis thaliana*, which facilitates resistance to the same mildew via the targeted export of plant metabolites (*6, 7*). The mechanistic similarity between LR34 and PEN3 reinvigorates hypotheses that specific and

Department of Plant Sciences and Genetics Graduate Group, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA. E-mail: kliebenstein@ucdavis.edu

diverse plant natural products may play essential roles in plant resistance to pathogens (*8*).

Fu *et al*. find that another QTL, *Yr36*, which provides quantitative resistance to stripe rust, harbors the gene *WHEAT KINASE START 1* (*WKS1*). The WKS1 protein contains a steroidogenic acute regulatory protein–related lipid transfer domain (START) and a functional enzymatic (kinase) domain, suggesting a role for plant lipids in cell signaling mechanisms that confer disease resistance. This specific combination of domains, unique among plants examined to date, appears to have evolved just before the divergence of wheat and its closest relatives. However, although *WKS1* provides resistance to diverse stripe rusts, the resistant allele is present in only a minority of the species tested, and even within the species in which *WKS1* occurs, it is not ubiquitous across all genotypes. This suggests that the gene evolved early in the wheat lineage and was then repeatedly lost. Natural variation of gene function is also a hallmark of gene-for-gene resistance loci, where continual shifts in pathogen populations are believed to drive rapid evolution of these genes (*9, 10*). Given the dearth of cloned quantitative resistance loci, it remains to be seen if these loci show enhanced levels of polymorphism and evolutionary dynamics similar to those for gene-for-gene resistance.

Why don't pathogens evolve counter-resistance to durable plant resistance? Perhaps these quantitative resistance loci target components of pathogens that are so critical for pathogen success that their encoding genes are evolutionarily constrained relative to similar pathogen genes involved in gene-for-gene interactions. Alternatively, these quantitative genes may provide resistance to the plant at less detriment to pathogen fitness than do gene-for-gene resistances. In both scenarios, decreasing the selective pressure for pathogens to evolve counter-resistance promises greater long-term success for improving crop yields using durable quantitative plant defenses. Additionally, for both cloned QTL described by Krattinger *et al.* and Fu *et al.*, alleles associated with loss of plant resistance to the tested pathogens would be predicted to produce modified, rather than nonfunctional proteins. Future inquiry into possible alternative functions of these genes or fitness costs of possessing resistant alleles at these loci may provide further insight into the evolution and durability of these mechanisms.

The studies by Fu *et al*. and by Krattinger *et al*. provide a first glimpse into the molecular mechanisms controlling quantitative resistance. Together, they demonstrate that multiple mechanisms can contribute to quantitative resistance within a given plant and that these mechanisms are not obviously associated with the well-studied and taxonomically widespread specific gene-for-gene resistance mechanisms. Mechanisms of quantitative resistance may be unique to particular plant lineages. Thus, while providing valuable tools to develop durable resistance against important wheat pathogens, these studies also argue that the molecular mechanisms of quantitative resistance need to be studied in a wide range of crop and model plants to fully explain the phenomenon.

References

1. D. Fu *et al.*, *Science* **323**, 1357 (2009); published online 19 February 2009 (10.1126/science.1166289).
2. S. G. Krattinger *et al.*, *Science* **323**, 1360 (2009); published online 19 February 2009 (10.1126/science.1166453).
3. B. A. McDonald, C. Linde, *Annu. Rev. Phytopathol.* **40**, 349 (2002).
4. J. A. Poland, P. J. Balint-Kurti, R. J. Wisser, R. C. Pratt, R. J. Nelson, *Trends Plant Sci.* **14**, 21 (2008).
5. P. L. Dyck, D. J. Samborski, R. G. Anderson, *Can. J. Genet. Cytol.* **8**, 665 (1966).
6. N. K. Clay, A. M. Adio, C. Denoux, G. Jander, F. M. Ausubel, *Science* **323**, 95 (2009).
7. P. Bednarek *et al.*, *Science* **323**, 101 (2009).
8. H. D. Van Etten, D. E. Matthews, P. S. Matthews, *Annu. Rev. Phytopathol.* **27**, 143 (1989).
9. H. H. Flor, *Adv. Genet.* **8**, 29 (1956).
10. E. G. Bakker, C. Toomajian, M. Kreitman, J. Bergelson, *Plant Cell* **18**, 1803 (2006).

10.1126/science.1171410

ENGINEERING

Infrastructure Design Issues in Disaster-Prone Regions

Advances in infrastructure assessment and design should help designers and builders support sustainability and resilience goals.

Seth D. Guikema

If designed and managed well, infrastructure—the networks that transport people and goods, distribute energy, and maintain communications and the buildings in which people live, work, and play—contributes to societal sustainability and resilience in areas at substantial risk from catastrophic events such as hurricanes. The extent to which infrastructure functions after such events depends on design choices that trade off, at least implicitly, current construction costs for future repair and replacement costs. These choices are based on assumptions that may not reflect all of the relevant factors.

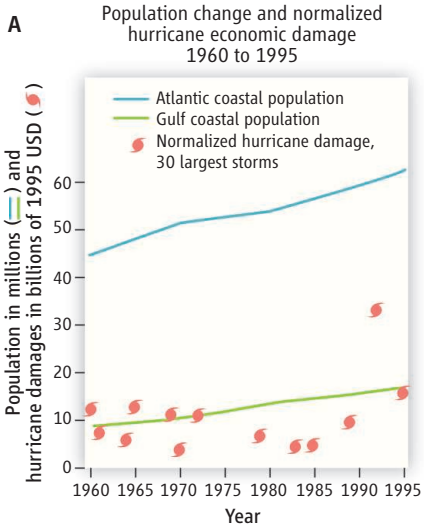
Recent advances in assessment and design allow the economic and environmental trade-offs in both design and postdisaster restoration to be explicitly considered and managed proactively during the design process. Similarly, infrastructure design can take advantage of the interactions between the natural and built environments in disaster-prone regions.

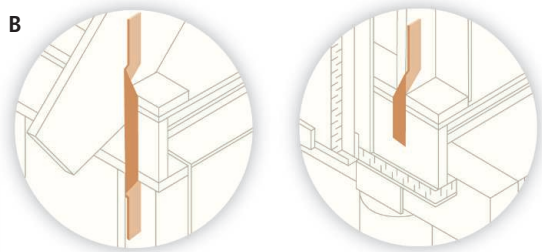
For example, we have an extensive record of hurricane impacts and can design structures to withstand those impacts in many cases. Despite advances in our knowledge of structural design for hurricane-prone areas, economic damage to buildings from major hurricanes in the United States has remained largely steady over the past four decades when adjusted for population growth and inflation (*1*, *2*) (see the figure, panel A).

Consider as an example a new commercial development to be built in a hurricane-prone area. Traditional design practice would be to design the individual buildings, utility poles, water distribution pipes, and flood protection systems to meet the design wind speeds and flooding levels specified in the relevant standards (*3*) or occurring with a preselected yearly probability. Current standards (*3*) specify that most structures (except for critical structures such as emergency response facilities) should be designed to withstand the stresses imposed by a hurricane that would occur, on average, every 50 years. This approach makes three fundamental assumptions that should be questioned.

First, it implicitly assumes that the design standards strike the proper balance between the benefits and costs of different design alternatives. Design standards are intended to apply across many different buildings, each with location-specific costs of reinforcement and failure. For example, two neighboring low-rise industrial buildings may have been

Department of Geography and Environmental Engineering, Johns Hopkins University, Baltimore, MD 21218, USA. E-mail: sguikema@jhu.edu





Hurricane strap ties transfer the load from the roof rafters down through the studs and into the foundation systems

Exposed to the elements. (**A**) Normalized insured economic losses from the 30 most damaging hurricanes in the United States and population growth (*1*). Economic losses are normalized for population growth, inflation, and property values using 1995 dollars, with each point representing the total economic loss in that year from the 30 major hurricanes. Years with no losses are not shown. Coastal population estimates are from the U.S. Census Bureau (*2*). (**B**) Use of hurricane straps in residential construction (*16*). (**C**) A hurricane-resistant dome home (*17*).

built at the same time and may be structurally similar. Yet, if one contains low-value manufacturing while the other houses high-value computer chip manufacturing involving hazardous chemicals, the costs of failure during a disaster will be dramatically different. Standardized design codes do not directly account for these differences.

Second, it assumes that the life-cycle environmental impacts of the development should not directly enter design decision-making, particularly the selection of the level of robustness of the system. However, contaminant releases, degradation of ecosystems, and impacts on communities surrounding the facility must also be considered. These burdens are real, even if hard to measure.

Third, it assumes that the design decision-making should not consider depending on or enhancing protection offered by natural ecosystems. For example, in designing a levee system to protect New Orleans, the flood-damping benefits of marshes and barrier islands were not explicitly considered as part of the design process. These natural buffer systems have been degrading over time, potentially decreasing the level of protection afforded by the engineered system (*4*). Both marsh restoration and designed facilities can be important parts of the protection system.

Recent advances have allowed engineers to consider a wider range of impacts in the design process. These include the life-cycle financial impacts of failures during disasters (*5*, *6*), enabling increased up-front reinforcement cost to be balanced against the uncertain costs of failures during disasters. Indirect economic costs of infrastructure failures, including loss of business revenue and economic growth in impacted areas, can also be included based on modified regional economic input-output models incorporating adaptation (*7*). These types of models suggest that indirect economic costs are at least twice as high as direct losses from major disasters. Environmental life-cycle inventory analysis allows the outcomes to be further broadened to include contaminant emissions and energy use over the life cycle of a built facility (*8*–*10*).

The interactions between natural ecosystems and built infrastructure are now better understood in many regions. In particular, mangrove swamps and other coastal vegetation can reduce the impacts of hurricane winds, hurricane surges, and tsunamis on infrastructure (*11*–*14*), with submerged marsh vegetation alone offering the potential of up to 30% reductions in wave heights (*14*). The damping effects of natural ecosystems can be taken advantage of by integrating ecosystem restoration in the infrastructure design process and then adapting the resistance of the components of the infrastructure systems to account for the modified loading, building from recent advances in understanding stresses imposed by hurricanes (*15*).

Initial steps have been taken toward the integrated design approach suggested here. For example, residential homes can be built with a modified, elevated foundation, materials that are stronger and more resistant to mold, hurricane straps (see the figure, panel B), and improved building envelope sealing, moisture management, and insulation. This approach can decrease energy costs by $600 to $1000 per year and yield a home substantially more resistant to costly failure during hurricanes while maintaining a traditional architectural appearance (*16*). With a less traditional design, "dome homes" (*17*) (see the figure, panel C) provide another highly hurricane-resistant approach that has seen successful, although limited, use. However, further work is needed to fully account for and take advantage of the interactions between the natural and built environments. The optimal design is one that considers the full impacts of infrastructure on the surrounding environment and community and the influence of the surrounding environment on the built facility.

PHOTO CREDIT: VALERIE SIGLER

References and Notes

1. R. Pielke Jr., C. Landsea, *Weather Forecasting* **13**, 621 (1998).
2. Census Bureau, *The 2008 Statistical Abstract* (U.S. Census Bureau, 2008).
3. ASCE, *ASCE 7-05: Minimum Design Loads for Buildings and Other Structures* (American Society of Civil Engineers, Reston, VA, 2006).
4. J. Bourne, *Science* **289**, 1860 (2000).
5. S. Chang, *Nat. Hazards Rev.* **4**, 186 (2003).
6. S. Chang, M. Shinozuka, *J. Infrastruct. Sys.* **2**, 118 (1996).
7. S. Hallegatte, *Risk Anal.* **28**, 779 (2008).
8. A. Horvath, thesis, Carnegie Mellon University, Pittsburgh, PA (1997).
9. A. Horvath, C. Hendrickson, *J. Infrastruct. Sys.* **4**, 111 (1998).
10. S. Junnila, A. Horvath, *J. Infrastruct. Sys.* **9**, 157 (2003).
11. F. Danielsen *et al.*, *Science* **310**, 643 (2005).
12. R. Costanza *et al.*, *Ambio* **37**, 241 (2008).
13. R. Costanza, W. Mitsch, J. Day Jr., *Front. Ecol. Environ.* **4**, 465 (2006).
14. L. Augustin, J. Irish, P. Lynett, *Coastal Engineering*, in press; published online 31 October 2008; (10.1016/j.coastaleng.2008.09.004).
15. I. Robertson, H. Riggs, S. Yim, Y. Young, *J. Waterway, Port, Coastal, Ocean Eng.* **133**, 463 (2007).
16. P. Baker. *Building a Durable and Energy-Efficient Home in Post-Katrina New Orleans*, Research Report 0704 (Building Science Press, Somerville, MA, 2007).
17. J. Fehrenbacher, *Inhabitat,* October 2005. Available from www.inhabitat.com/2005/10/10/monolithic-domes.

18. Supported by the Office of Science (Biological and Environmental Research), U.S. Department of Energy, Grant DE-FG02-08ER64644 and NSF Grant ECCS-0725823.

10.11206/science.1169057

The Challenge of Finding a Cure for HIV Infection

Douglas D. Richman,[1]* David M. Margolis,[2] Martin Delaney,[3]† Warner C. Greene,[4] Daria Hazuda,[5] Roger J. Pomerantz[6]

Although combination therapy for HIV infection represents a triumph for modern medicine, chronic suppressive therapy is required to contain persistent infection in reservoirs such as latently infected CD4$^+$ lymphocytes and cells of the macrophage-monocyte lineage. Despite its success, chronic suppressive therapy is limited by its cost, the requirement of lifelong adherence, and the unknown effects of long-term treatment. This review discusses our current understanding of suppressive antiretroviral therapy, the latent viral reservoir, and the needs for and challenges of attacking this reservoir to achieve a cure.

Highly active antiretroviral therapy (HAART) for the chronic suppression of HIV replication has been the major accomplishment in HIV/AIDS medicine (*1*, *2*). Many patients are now in their second decade of treatment, with levels of plasma HIV RNA below the limits of detection of clinical assays. The impact on morbidity and mortality in the developed world has led to efforts that have brought this therapy to nearly three million people in resource-limited settings (*3*). Many patients are now enjoying a life-style little encumbered by symptoms or the side effects of medications, many of which require only once-daily administration. With the remarkable success of chronic suppression, why propose curing HIV infection—a challenging objective that requires potentially risky interventions and that may be unachievable?

Can We Do Better Than HAART?

HAART is no panacea. Current treatments must be maintained for life, with treatment interruption resulting in the rapid rebound of replicating virus. Although drug resistance can emerge because of the challenges of maintaining adherence and access to chronic antiviral therapy or owing to transmitted drug-resistant viruses, the success of HAART has been improved by the development of more potent and more tolerable therapies. Successful new drug development may not continue indefinitely, however, and HAART may never reach the majority of infected individuals in less-developed countries. Despite the prolonged suppression of HIV replication below the standard limits of detection for patients on HAART, ongoing viremia can be detected at levels of 1 to 50 copies per milliliter in the majority of patients (*4*, *5*). The origin of this viremia has not been fully characterized, but it does not appear to jeopardize the prolonged success of therapy in the adherent patient (*6*). Nevertheless, the virions may engage CD4 and chemokine receptors and may activate pathways that could lead to chronic consequences, including cardiovascular and malignant disease. The suboptimal penetration of many antiretrovirals into the central nervous system may also permit low levels of viral replication and/or release from stable viral reservoirs, resulting in neuropathology (*7*, *8*).

Despite the very low rates of toxicity of many of the newer HAART regimens, many of these drugs modulate lipid and glucose metabolism (*9*). Even modest toxicities may have cumulative effects over decades of treatment. Moreover, prolonged treatment may reveal toxicities not appreciable with animal toxicology or several years of clinical surveillance. There is already growing concern about increased rates of heart disease, diabetes, liver disease, and many forms of cancer in aging HIV-infected patients who are receiving treatment (*10*–*13*). Whether these are because of long-term HIV infection, therapeutic drug treatment, or both, is uncertain. Finally, the cost of HAART may be too much to sustain treatments on a global scale, as millions are affected.

Given the shortcomings of HAART, time-limited interventions that do not result in the resumption of viremia are a desirable but a currently unattainable objective, unlike what can be achieved with the treatment of hepatitis C virus infection. Such therapy might or might not eliminate every functional virion or infected cell, but would permit the discontinuation of HAART without the reappearance of viremia and disease. We propose that a drug-free remission should be the new goal of HIV therapeutics.

What Is the State of HIV in Successfully Treated Patients?

The source of the low-level viremia seen in most patients on HAART (*4*, *14*, *15*) may be incompletely characterized, but we do have some hints (Table 1). The failure, thus far, of treatment intensification to clear this viremia (*16*) and the lack of evidence for nucleotide sequence evolution over long periods of treatment (*17*–*19*) indicate that this phenomenon may not be driven by ongoing rounds of replication.

Patient data reveal that 1 in 10^6 CD4$^+$ T cells are latently infected with HIV, despite the durable suppression of detectable plasma viremia, although the frequency can be much lower in some patients (*20*–*22*). In vivo, it is thought that these cells are intermittently activated by antigen recognition or as bystanders in a local inflammatory process, which leads to the release of progeny virions.

Another source of virion production, which does not require ongoing replication, is the episodic production of HIV by long-lived cells. In situ hybridization of lymphoid tissue in simian immunodeficiency virus (SIV)–infected macaques and HIV-infected humans revealed that, in addition to the activated and infected CD4$^+$ T cells that produce large numbers of virions with a short cellular half-life, many lymphocytes can be visualized that produce small amounts of viral RNA, yet do not display markers of activation (*23*). Such cells are not seen in vitro, and whether such cells occur in vivo during prolonged antiretro-

Table 1. HIV latency.

- Latently infected resting memory CD4$^+$ T cells are the best-characterized latent reservoir for HIV-1.
- Less than 1 cell per 1,000,000 resting CD4$^+$ T cells from patients on HAART harbor latent HIV-1 provirus.
- Sequence of latent proviruses does not evolve, which suggests no ongoing viral replication.
- Discontinuation of HAART allows viral relapse from latent reservoir.
- Patients successfully treated with HAART for longer than 10 years exhibit no appreciable decrease in the size of the latent reservoir.
- The persistence of latently infected memory CD4$^+$ T lymphocytes precludes their elimination by HAART alone for the lifetime of the patient.
- Other drug-insensitive reservoirs, including brain, macrophages, and hematopoietic stem cells, may also exist.
- Latency is likely established and maintained by numerous blocks at multiple steps in the HIV-1 replicative pathway, which potentially complicates eradication strategies.

[1]San Diego VA Healthcare System and University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0679, USA. [2]Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [3]Project Inform, 1375 Mission Street, San Francisco, CA 94103, USA. [4]Gladstone Institute of Virology and Immunology, San Francisco, CA 94158, and University of California at San Francisco, San Francisco, CA 94143, USA. [5]Merck and Co., West Point, PA 19486, USA. [6]Tibotec Pharmaceuticals Inc. and Johnson and Johnson Corporation, 1020 Stony Hill Road, Suite 300, Yardley, PA 19067, USA.

†Deceased 23 January 2009.
*To whom correspondence should be addressed. E-mail: drichman@ucsd.edu

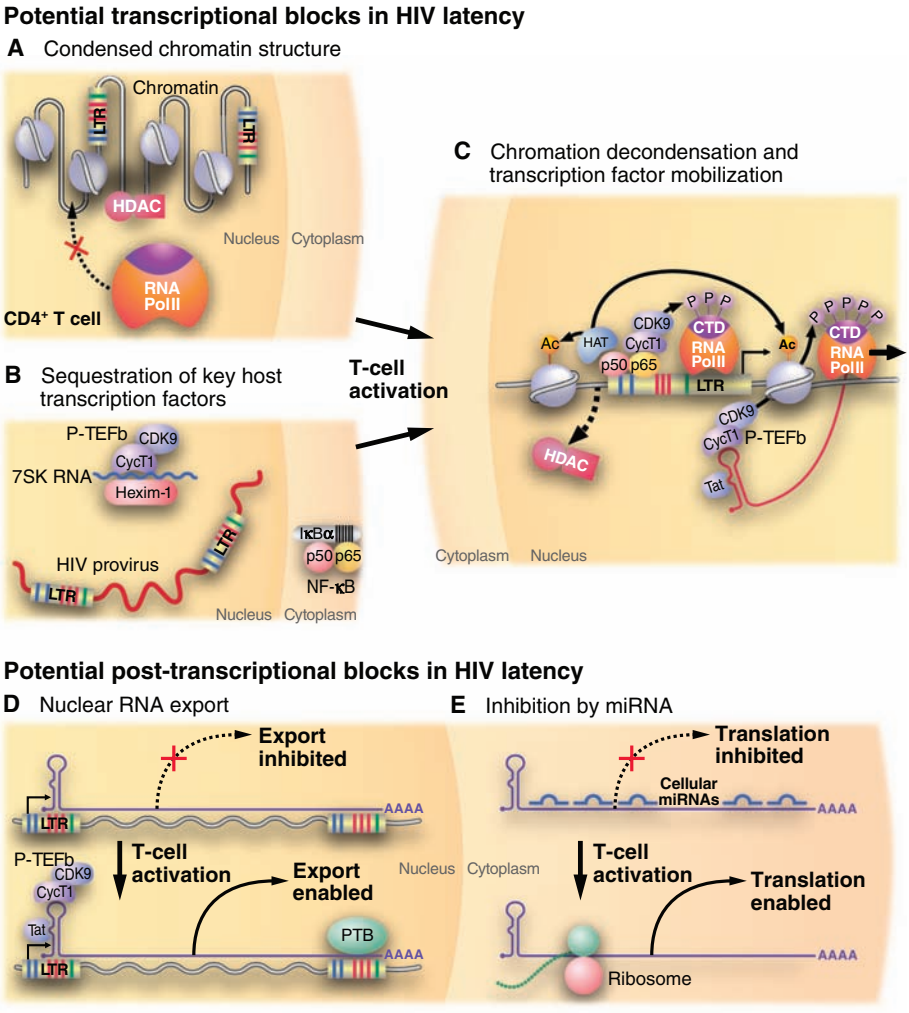


Fig. 1. Proviral latency is the result of multiple restrictions on HIV expression. (**A**) Proviral latency is maintained, in part, by the action of several transcription factors that recruit HDACs and other complexes to the HIV-1 long-terminal repeat (LTR) promoter, which results in histone modifications within chromatin at the HIV promoter that limit the ability of RNA polymerase to initiate transcription. (**B**) Key cellular factors that are required for robust HIV transcription, such as NF-κB or the P-TEFb–cyclin complex, are sequestered in resting $CD4^+$ T cells by cellular regulatory complexes [inhibitor of nuclear factor κB (IκB) and HEXIM–7SK RNA, respectively). Release and mobilization of these factors is required for proviral expression. (**C**) When histone acetyltransferases (HATs) supercede the effect of HDACs, coactivators such as NF-κB can recruit RNA polymerase (RNA Pol) complexes. Production of Tat allows the recruitment of P-TEFb, mediating an explosive increase in transcription and the escape of provirus from latency. (**D**) The initial wave of Tat production may be further restricted by inefficient export of multiple spliced HIV mRNAs, relieved upon cellular activation by enhanced expression of PTB. (**E**) Cellular miRNAs that bind HIV mRNAs may also restrict translation of early expressed HIV mRNAs and so reduce Tat production. CDK, cyclin-dependent kinase; CTD, C-terminal repeat domain; and CycT1, cyclin T1.

viral therapy is unknown. Further, the life span of and the kinetics of viral expression in such cells remain undefined.

Low-level plasma viremia cannot always be linked to activation of latently infected $CD4^+$ T cells. In a longitudinal analysis of cloned RNA from plasma-derived virions of a subset of HAART-suppressed patients, the Siliciano group identified distinctive homogeneous viral subpopulations (*24*). These observations raise the possibility of a chronically infected clonal reservoir, analogous to a persistently infected stem cell. How a persistently infected cell population could produce virions at a steady state for years, in the presence of some level of cell-mediated immunity, remains unexplained. Other cellular or tissue sources of virus, such as cells of the monocyte and macrophage lineages, may also contribute to low levels of viremia.

Can Mechanisms That Drive Latency Be Therapeutically Exploited?

Activation from latency to completion of the replication cycle should result in lytic cell death of $CD4^+$ T cells. Multiple mechanisms may contribute to the maintenance of proviral latency [reviewed in Williams and Greene (*25*)], and so, combination approaches could be required to eradicate infection (Fig. 1 and 2). Such strategies would depend on current or future antiretroviral therapy to completely inhibit all new infection events. Antilatency agents would be given, intermittently and for a limited period of time, to purge the last sanctuaries of HIV infection (Fig. 3).

Chromatin remodeling enzymes like histone deacetylases (HDACs) play a critical role in HIV latency (Fig. 1A) (*26*–*29*). HDACs are recruited to the highly conserved initiator region of the HIV promoter by several distinct complexes, by means of factors that are both ubiquitous in cell types infected by HIV and also participate in basal and activated viral gene expression. The existence of multiple mechanisms that recruit repressive HDAC complexes to the proviral promoter raises the possibility that HDAC inhibitors might lead to the activation of HIV in latently infected cells (Fig. 2).

In addition to HDACs, HIV expression is limited by other cellular barriers to effective mRNA transcription, which the virus overcomes through the action of its own activator, Tat. Tat recruits the positive transcription elongation factor b (P-TEFb) kinase to the integrated viral promoter, inducing viral gene expression (Fig. 1B and C) (*30*). Several kinase agonists, including hexamethylbisacetamide (HMBA)—a compound previously tested in human cancer trials (*31*), activate intracellular signaling cascades that mobilize P-TEFb in the absence of Tat (*32*, *33*) and can induce the expression of HIV in latently infected cells (Fig. 2) (*34*).

The HIV promoter responds to coactivators that are abundant in activated cells, but, in the context of the resting T cell, inadequate nuclear levels of nuclear factor κB (NF-κB) and nuclear factor of activated T cells (NFAT) may contribute to the establishment of latency (Fig. 1B) (*35*). Diminished binding could be the result of changes in chromatin structure, in part mediated by the action of HDACs. Prostratin, a nontumorigenic phorbol ester isolated from the Samoan medicinal plant, *Homalanthus nutans*, induces HIV expression in latently infected cell lines and cells isolated from HIV-infected, HAART-treated patients in the absence of cellular proliferation (*36*). In cell-line models, prostratin stimulates HIV expression through protein kinase C–mediated activation of NF-κB and so provides an approach to activation and clearance of latently infected cells (Fig. 2) (*37*).

HIV mRNA export may also be impaired in resting T cells because of the low levels of polypyrimidine tract–binding protein (PTB) available in resting cells (Fig. 1D) (*38*). MicroRNAs (miRNAs) endogenously expressed in human cells may further impede HIV mRNA expression or translation (Fig. 1E) (*39*, *40*). If such mechanisms contribute to proviral persistence, entirely new classes of therapeutic agents able to safely alter host RNA expression or transport will be required.

Given the intimacy of the interaction between the retrovirus and the host cell, therapeutic approaches that disrupt latent infection are also likely to affect host cell function. Although mild host toxicities for limited periods of time might be acceptable, global immune activation must be avoided. Once quiescent virus is successfully induced to complete a round of replication, virus-induced cytolysis and cytotoxic T cells need to be able to clear HIV antigen–expressing cells. The viral progeny generated by such activated cells have to be prevented from successfully infecting other cells by the presence of HAART (Fig. 2).

How Are Interventions to Be Investigated?

Undoubtedly, there are other factors that regulate latency occurring in primary cells in vivo. Although we need to be aware of the potential for additional reservoirs of infectious virus, addressing the latently infected T cell reservoir may be the most direct way of exposing an even smaller additional reservoir, like infected macrophages, or anatomic compartments, such as the central nervous system, that may be suboptimally exposed to HAART. Careful in vivo testing of therapeutic agents capable of antagonizing the different mechanisms underlying HIV latency identified in $CD4^+$ T cells is important for establishing the proof of concept.

An animal model is not required for antiretroviral drug development because, thus far, activity in vitro has correlated with activity in vivo. In contrast, an animal model could be invaluable in the development and testing of antilatency therapies and would guide clinical trial design. Given the excellent outcomes of HAART, initial studies of new antilatency therapies in humans might be difficult to design and execute, because volunteers in such early studies may have little to gain, and the candidate interventions will have unproven efficacies and uncertain toxicities. SIV infection in the rhesus macaque gives rise to latent infections in $CD4^+$ T cells that mirror HIV latency (*41*), although it remains unknown whether the pathways and molecular targets promoting postintegration latency in macaques are the same as in humans.

BLT (bone marrow-liver-thymus) mice provide a second animal model. These immunodeficient mice (which lack endogenous T and B cells) are transplanted with human thymus and liver tissue and injected with hematopoietic stem cells, giving rise to systemic repopulation with human T and B cells, monocytes-macrophages, and dendritic cells capable of antibody production, activation by human antigen-presenting cells, and potent human major histocompatibility complex–restricted T cell immune responses (*42*). BLT mice have already been used to study HIV transmission and to test preexposure antiretroviral prophylaxis (*43*). Determining whether this model can be used to study HIV latency is a high experimental priority. Despite the availability of animal models for preliminary testing, clinical studies in HIV-infected patients are ultimately required.

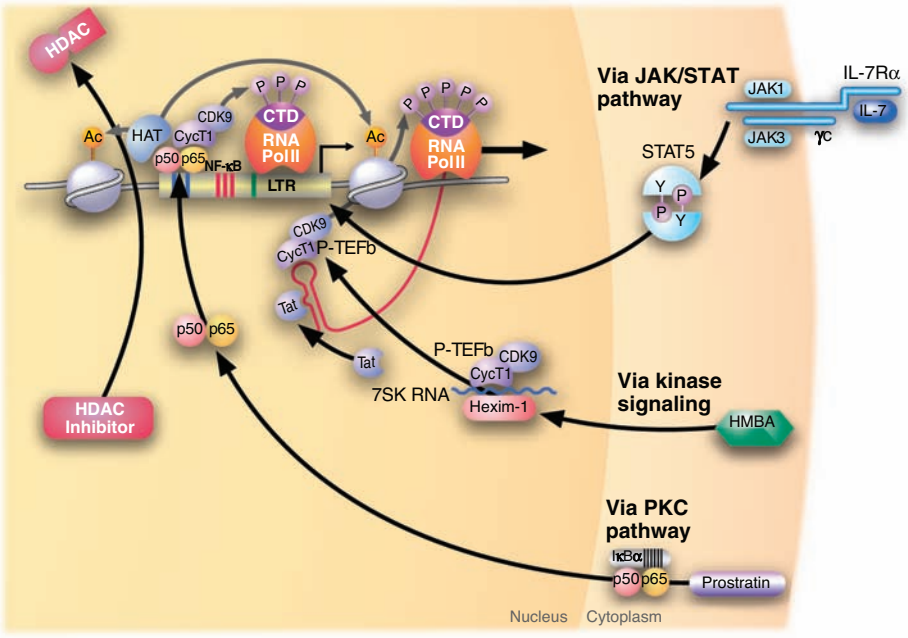


Fig. 2. Potential therapies to disrupt latent proviral HIV infection. HDAC inhibitors may relieve repression by HDACs and may allow histone acetylation by HAT, which results in HIV expression. Via kinase signaling, HMBA stimulates the release of P-TEFb from sequestration within a ribonucleoprotein complex containing HEXIM and 7SK snRNA (small nuclear RNA). Tat then recruits P-TEFb to an HIV RNA structure (TAR), present at the 5′ end of all nascent HIV RNAs, which allows for phosphorylation and activation of RNA Pol II and other factors, leading to processive transcription. Prostratin stimulates HIV through protein kinase C (PKC)–mediated release of active NF-κB. Interleukin 7 (IL-7), a cytokine essential for maintenance of T cell homeostasis, can induce HIV expression from quiescent resting cells without global T cell activation, via the JAK/STAT5 signaling pathway.

Phase I trials to deplete persistent HIV infection have demonstrated that these approaches can be tested safely (*44*–*46*), and studies using novel inducers of HIV expression such as interleukin 7 (*47*) may soon be feasible (Figs. 2 and 3).

Quantifying the latent HIV reservoir in humans is challenging when less than 1 in a million $CD4^+$ T cells are latently infected, and there are approximately 100 copies of integrated provirus for each latently infected $CD4^+$ T cell (*48*). After amplification by the polymerase chain reaction, measurements of integrated proviral DNA might serve as a surrogate marker for changes in the latent reservoir (*18*). However, the small size of the reservoir and the imprecision of current assays require improved techniques to assess the effectiveness of interventions. Moreover, once the reservoir is reduced by 10- to 100-fold, the remaining latently infected cells may be concealed below the limit of detection of any assay yet described.

Access to lymphoid tissue or most anatomic compartments in otherwise healthy subjects is difficult. Although such studies may fail to detect an infected reservoir, they cannot prove its eradication. When an intervention or combination of interventions is considered sufficiently compelling, the ultimate test of efficacy will be the withdrawal of HAART. Antiretroviral therapy is effective and relatively safe. As a result, the administration of any experimental intervention in either a proof-of-concept feasibility trial or in a trial incorporating treatment interruption raises significant ethical, regulatory and study design issues, because antiretroviral therapy is so effective and relatively safe. Therefore, involvement of various stakeholders in thoughtful deliberations is necessary. Such studies are required if we wish to cure HIV; but, although the potential benefit to humanity is great, the benefit to the early trial volunteers is nearly nonexistent. The appropriate volunteers in a trial involving treatment interruption might be those who initiated HAART before significant immune depletion. This criterion would minimize risk of treatment interruption, especially with close monitoring to resume treatment should virus replication be detected. A second rationale for selecting such subjects is that their infected-cell reservoir may be smaller and thus more amenable to intervention (*18*, *49*).

Do We Need a New Approach to Develop a Cure?

The recent disappointing results from the trials of HIV vaccine and microbicide candidates have prompted a renewed commitment to basic research to identify effective approaches to these critically needed prevention strategies. We advocate a similar impetus for new approaches to purge the latent reservoir in order to cure HIV infection.

Years of effort have led to public health strategies to reduce the risk of cancer, a vaccine

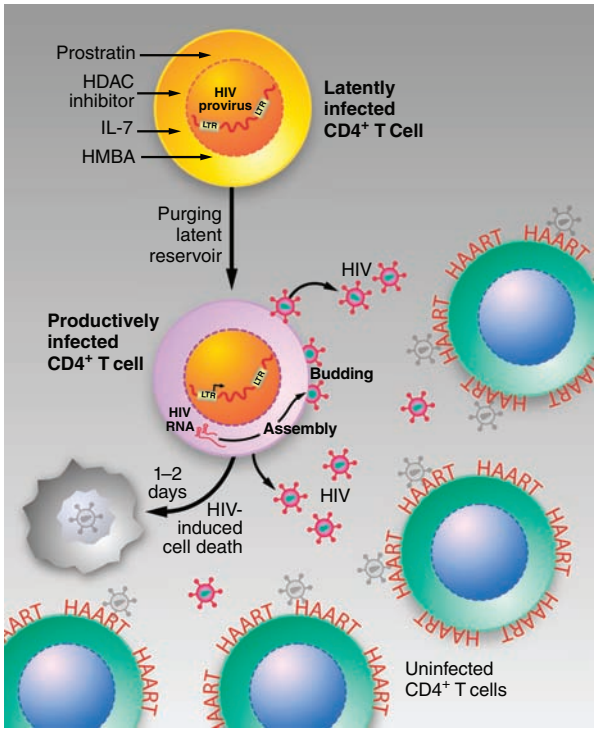


Fig. 3. Purging persistent proviral infection. If targeted approaches, alone or in combination, succeed in activating latent HIV proviruses present in differentiated CD4+ T cells, the life span of these cells should be short. These inductive agents must be used in combination with HAART to prevent further HIV spread to uninfected CD4+ T cells.

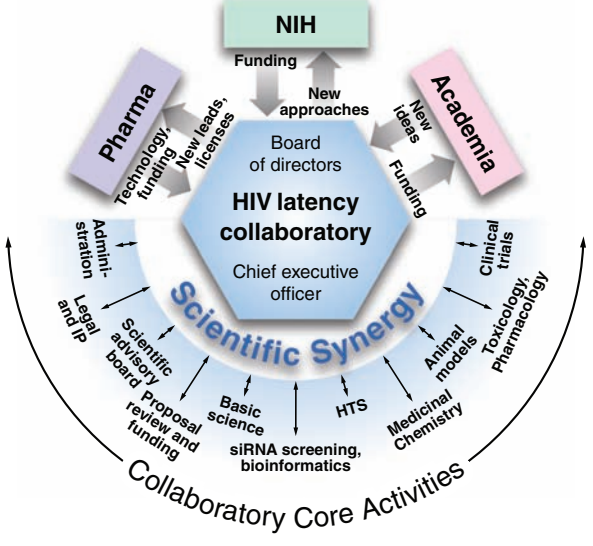


Fig. 4. Overview of an HIV Latency Collaboratory representing a joint research venture between the U.S. National Institutes of Health (NIH), the pharmaceutical industry (pharma), and academia. The goal of this collaboratory is to accelerate basic discovery and the clinical translation of these discoveries that allows for eradication of HIV in infected patients or at least a stable drug-free remission. This collaboratory is designed to include all interested investigators with meritorious ideas. These investigators will be funded via the collaboratory and will have access to a range of core technologies provided by the collaboratory that should promote scientific synergy and enhanced experimental efficiency. The collaboratory would be governed by a Board of Directors comprised of representatives from each of the partners in the joint venture. HTS, high-throughput screening; IP, intellectual property; and siRNA, small interfering RNA.

that prevents cervical cancer, better therapies to treat malignancies, and curative therapies for some cancers. Such a multifaceted approach should also be applied to the effort to cure HIV infection. This will require behavioral and biological tools to prevent HIV infection; safe, affordable, and nontoxic therapies for initial control of HIV infection; and new interventions that can achieve a drug-free remission of viremia in some patients.

The challenge of developing an HIV vaccine spans the need for new basic research insights to product development to clinical trials. The complexity of fostering and coordinating these efforts has led to the creation of major NIH intramural (Vaccine Research Center) and extramural (Center for HIV/AIDS Immunology) programs and of an international, multi-institutional effort (The Global HIV Vaccine Enterprise). Our understanding of HIV latency has chiefly resulted from independent, investigator-initiated efforts. In order to translate these academic accomplishments into clinical treatments similar initiatives are required. Antilatency therapies will require the drug discovery capabilities of industry, like high-throughput drug candidate screening; medicinal chemistry; product synthesis, production, and formulation; toxicology; and pharmacology. A coordinated initiative involving academia, industry, government, and patient advocates could greatly accelerate the identification of potential interventions and their clinical assessment (Fig. 4). We conceive an initiative, termed here a collaboratory, in which the government contributes funding, regulatory oversight, and coordination; industry contributes funding, drug discovery, technology, and expertise; and academia contributes ideas and investigative capacity. Long-term support for a flexible, collaborative public-private joint venture might improve efficiency and conserve resources, while at the same time catalyzing progress that no single group could achieve. Clearly much work and many challenges lie ahead, but if novel scientific insights can be brought to bear in clinically effective ways, the era marked by the benefits of HAART may be followed by one in which HAART is no longer a lifelong necessity.

References and Notes

1. F. J. Palella Jr. *et al.*, *N. Engl. J. Med.* **338**, 853 (1998).
2. R. P. Walensky *et al.*, *J. Infect. Dis.* **194**, 11 (2006).
3. World Health Organization, *Towards Universal Access: Scaling Up Priority HIV/AIDS Interventions in the Health Sector: Progress Report 2008* (World Health Organization, Geneva, June 2008); www.who.int/hiv/mediacentre/2008progressreport/en/index.html.
4. G. Dornadula *et al.*, *JAMA* **282**, 1627 (1999).
5. M. Fischer *et al.*, *AIDS Res. Hum. Retroviruses* **16**, 1135 (2000).
6. D. V. Havlir *et al.*, *JAMA* **286**, 171 (2001).
7. O. Lambotte *et al.*, *AIDS* **19**, 217 (2005).
8. S. Letendre *et al.*, *Arch. Neurol.* **65**, 65 (2008).
9. P. W. Mallon, *AIDS Rev.* **9**, 3 (2007).
10. R. Bedimo, *Curr. HIV/AIDS Rep.* **5**, 140 (2008).
11. D. Florescu, D. P. Kotler, *Antivir. Ther.* **12**, 149 (2007).
12. The Data Collection on Adverse Events of Anti-HIV Drugs Study Group, *Arch. Intern. Med.* **166**, 1632 (2006).
13. K. Mondy, P. Tebas, *Annu. Rev. Med.* **58**, 141 (2007).
14. F. Maldarelli *et al.*, *PLoS Pathog.* **3**, e46 (2007).
15. S. Palmer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3879 (2008).
16. F. Maldarelli *et al.*, *Antivir. Ther.* **13** (suppl. 3), A79 (2008).
17. H. F. Gunthard *et al.*, *J. Virol.* **73**, 9404 (1999).
18. M. C. Strain *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4819 (2003).
19. L. Zhang *et al.*, *N. Engl. J. Med.* **340**, 1605 (1999).
20. T. W. Chun *et al.*, *Nature* **387**, 183 (1997).
21. D. Finzi *et al.*, *Science* **278**, 1295 (1997).
22. J. K. Wong *et al.*, *Science* **278**, 1291 (1997).
23. Z. Zhang *et al.*, *Science* **286**, 1353 (1999).
24. J. R. Bailey *et al.*, *J. Virol.* **80**, 6441 (2006).
25. S. A. Williams, W. C. Greene, *Cytokine* **39**, 63 (2007).
26. J. J. Coull *et al.*, *J. Virol.* **74**, 6790 (2000).
27. S. A. Williams *et al.*, *EMBO J.* **25**, 139 (2006).
28. G. Jiang, A. Espeseth, D. J. Hazuda, D. M. Margolis, *J. Virol.* **81**, 10914 (2007).
29. M. Tyagi, J. Karn, *EMBO J.* **26**, 4985 (2007).
30. B. M. Peterlin, D. H. Price, *Mol. Cell* **23**, 297 (2006).
31. C. W. Young *et al.*, *Cancer Res.* **48**, 7304 (1988).
32. X. Contreras, M. Barboric, T. Lenasi, B. M. Peterlin, *PLoS Pathog.* **3**, 1459 (2007).
33. V. Klichko, N. Archin, R. Kaur, G. Lehrman, D. Margolis, *J. Virol.* **80**, 4570 (2006).
34. S. K. Choudhary, N. M. Archin, D. M. Margolis, *J. Infect. Dis.* **197**, 1162 (2008).
35. D. Bisgrove, M. Lewinski, F. Bushman, E. Verdin, *Expert Rev. Anti Infect. Ther.* **3**, 805 (2005).
36. J. Kulkosky *et al.*, *Blood* **98**, 3006 (2001).
37. S. A. Williams *et al.*, *J. Biol. Chem.* **279**, 42008 (2004).
38. K. G. Lassen, K. X. Ramyar, J. R. Bailey, Y. Zhou, R. F. Siliciano, *PLoS Pathog.* **2**, e68 (2006).
39. J. Huang *et al.*, *Nat. Med.* **13**, 1241 (2007).
40. Z. Klase *et al.*, *BMC Mol. Biol.* **8**, 63 (2007).
41. A. Shen *et al.*, *J. Virol.* **77**, 4938 (2003).
42. M. W. Melkus *et al.*, *Nat. Med.* **12**, 1316 (2006).
43. P. W. Denton *et al.*, *PLoS Med.* **5**, e16 (2008).
44. N. M. Archin *et al.*, *AIDS* **22**, 1131 (2008).
45. T. W. Chun *et al.*, *Nat. Med.* **5**, 651 (1999).
46. J. Kulkosky *et al.*, *J. Infect. Dis.* **186**, 1403 (2002).
47. F. X. Wang *et al.*, *J. Clin. Invest.* **115**, 128 (2005).
48. Y. Han, M. Wind-Rotolo, H. C. Yang, J. D. Siliciano, R. F. Siliciano, *Nat. Rev. Microbiol.* **5**, 95 (2007).
49. M. C. Strain *et al.*, *J. Infect. Dis.* **191**, 1410 (2005).

50. We acknowledge the encouragement and support of C. Dieffenbach of the Division of AIDS, National Institute of Allergy and Infectious Diseases, NIH, and V. Miller and the Forum for Collaborative HIV Research. We also thank J. C. W. Carroll from the J. David Gladstone Institutes for graphic artwork. This article is dedicated to the memory of our friend and colleague, Martin Delaney.

10.1126/science.1165706

Extremely High Mutation Rate of a Hammerhead Viroid

Selma Gago,[1] Santiago F. Elena,[1] Ricardo Flores,[1] Rafael Sanjuán[1,2]*

Mutation rates vary by orders of magnitude across species (*1*, *2*), with the highest rates measured so far corresponding to RNA viruses (*3*), but little is known about other RNA replicons. Viroids are plant pathogens with minimal non-protein-coding RNA genomes replicated by host RNA polymerases (*4*). We estimated the mutation rate of *Chrysanthemum chlorotic mottle viroid* (CChMVd), a 399-nucleotide chloroplastic viroid with hammerhead ribozymes. Hammerheads are RNA motifs formed by three double-helix regions flanking a core of 15 highly conserved nucleotides critical for catalytic activity (*5*), which mediate self-cleavage of replicative intermediates and, hence, are essential for viroid replication. Hammerhead viroids show elevated genetic variability (*6*), but this variability results from the combined action of mutation and selection and therefore cannot be used to directly estimate mutation rates.

To achieve this goal, we inoculated plants with an in vitro transcript of CChMVd (*7*), and at the onset of symptoms we screened for mutations at the 15 core nucleotides plus the nucleotide preceding the self-cleavage site in each of the two hammerheads (32 sites). Considering that these mutations are lethal for the viroid, their population frequency must equal the mutation rate because, despite multiple replication rounds downstream from inoculation, they have necessarily been generated during the last one. In three independent experiments, we found three, seven, and five mutations in 63, 64, and 61 reverse transcription polymerase chain reaction (RT-PCR) clones, respectively (188 × 32 = 6016 total target sites), yielding a mutation rate of 0.0025 ± 0.0006 (SEM) per site and replication cycle, that is, one mutation per 400 nucleotides (fig. S1).

In a control experiment, we sequenced RT-PCR clones from the in vitro transcript used for inoculations and found a single substitution in 6525 sites. This result gives an error rate 17-fold lower than the estimated CChMVd mutation rate and discards any significant effect of RT-PCR artifacts. To confirm the lethality of the hammerhead mutations sampled in vivo, we recreated the mutations by site-directed mutagenesis and assayed for infectivity. Northern-blot hybridizations indicated that plants inoculated with these mutants had no detectable viroid RNA (fig. S2A). Further, self-cleavage analyses confirmed that all except one of the mutant hammerheads showed severely reduced or null catalytic activity (fig. S2B).

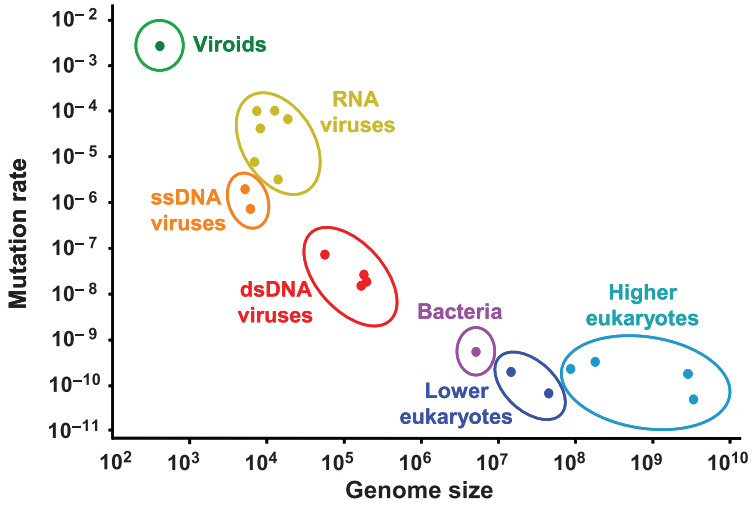


Fig. 1. Per-site mutation rate versus genome size for CChMVd and other biological entities [reviewed in (*2*) and updated with more recent data from (*3*)]. RNA viruses (left to right) are tobacco mosaic virus, human rhinovirus, poliovirus, vesicular stomatitis virus, bacteriophage Φ6, and measles virus. Single-stranded DNA viruses are bacteriophage ΦX174 and bacteriophage m13. Double-stranded DNA viruses are bacteriophage λ, herpes simplex virus, bacteriophage T2, and bacteriophage T4. Bacteria is *Escherichia coli*. Lower eukaryotes are *Saccharomyces cerevisiae* and *Neurospora crassa*. Higher eukaryotes are *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus*, and *Homo sapiens*. When several estimations were available, the mean value is shown.

To determine the strength of selection against mutations elsewhere in the viroid genome, we competed 24 random-point mutants against the wild type. Sequencing of 138 RT-PCR clones revealed that 20/24 mutations had been purged by selection at the onset of symptoms. In contrast, 51 new polymorphisms appeared in this time interval, showing that genetic variability is rapidly regenerated because of highly error-prone replication (fig. S3). We also inferred that hammerheads are unlikely to constitute mutational hotspots because polymorphisms did not map more frequently in hammerheads than in the rest of the genome (Fisher exact test, $P = 0.963$) whereas the fraction of point mutations that were selected against was also similar for these two regions (7/8 and 13/16, respectively).

The CChMVd mutation rate is the highest reported for any biological entity (Fig. 1). Hammerhead viroids are replicated by a proofreading-deficient chloroplastic DNA-dependent RNA polymerase that is redirected to use RNA instead of its native DNA template (*4*). This, together with the presence of mutagenic free radicals or unbalanced nucleotide pools, would lead to extremely error-prone replication. Viroids can tolerate such elevated per-site mutation rates owing to their minimal genomes, whereas more complex genomes would accumulate an excessive mutational load (*8*). Given their genomic simplicity and autocatalytic activity, hammerhead viroids are reminiscent of the postulated RNA world replicons (*9*). These primitive replicons would also resemble hammerhead viroids in their extremely error-prone replication. Thus, our results support the notion that the emergence of replication fidelity mechanisms was central to the evolution of complexity in the early history of life.

References and Notes

1. J. W. Drake, B. Charlesworth, D. Charlesworth, J. F. Crow, *Genetics* **148**, 1667 (1998).
2. P. D. Sniegowski, P. J. Gerrish, T. Johnson, A. Shaver, *Bioessays* **22**, 1057 (2000).
3. S. Duffy, L. A. Shackelton, E. C. Holmes, *Nat. Rev. Genet.* **9**, 267 (2008).
4. R. Flores, C. Hernández, A. E. Martínez de Alba, J. A. Daròs, F. Di Serio, *Annu. Rev. Phytopathol.* **43**, 117 (2005).
5. M. Martick, W. G. Scott, *Cell* **126**, 309 (2006).
6. N. Duran-Vila, S. F. Elena, J. A. Daròs, R. Flores, in *Origin and Evolution of Viruses*, E. Domingo, C. R. Parrish, J. J. Holland, Eds. (Elsevier, London, 2008), pp. 43–64.
7. Materials and methods are available as supporting material on *Science* Online.
8. M. Eigen, *Naturwissenschaften* **58**, 465 (1971).
9. T. O. Diener, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9370 (1989).
10. This work was supported by grants from the Spanish Ministerio de Ciencia e Innovación: BFU2006-14819-C02-01/BMC to S.F.E., BFU2008-03154/BMC to R.F., and BFU2008-03978/BMC to R.S. GenBank sequence accession numbers are FJ647228 to FJ647553.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5919/1308/DC1
Materials and Methods
Figs. S1 to S3
References

1 December 2008; accepted 2 February 2009
10.1126/science.1169202

[1]Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas–Universidad Politécnica de València, 46022 València, Spain. [2]Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, 46980 València, Spain.

*To whom correspondence should be addressed. E-mail: rafael.sanjuan@uv.es

Dynamic Order-Disorder in Atomistic Models of Structural Glass Formers

Lester O. Hedges,[1]* Robert L. Jack,[1,2]* Juan P. Garrahan,[3] David Chandler[1]†

The glass transition is the freezing of a liquid into a solid state without evident structural order. Although glassy materials are well characterized experimentally, the existence of a phase transition into the glass state remains controversial. Here, we present numerical evidence for the existence of a novel first-order dynamical phase transition in atomistic models of structural glass formers. In contrast to equilibrium phase transitions, which occur in configuration space, this transition occurs in trajectory space, and it is controlled by variables that drive the system out of equilibrium. Coexistence is established between an ergodic phase with finite relaxation time and a nonergodic phase of immobile molecular configurations. Thus, we connect the glass transition to a true phase transition, offering the possibility of a unified picture of glassy phenomena.

When supercooled far below their melting temperatures, many liquids become extremely viscous, so much so that at low enough temperatures these materials become amorphous solids (*1*, *2*). This phenomenon is termed the "glass transition." The dynamical behavior of molecules in a glass is heterogeneous in that there are domains of mobile and immobile molecules segregated in space (*3*–*6*). At equilibrium, the spatial extent of these domains is large compared with molecular dimensions (*5*) but not so large to imply an actual phase transition. Indeed, and despite the name given to it, there is no observation that demonstrates a link between the glass transition and a phase transition controlled by traditional thermodynamic variables like temperature and pressure.

Nevertheless, for idealized lattice models, recent work has established the existence of a nontraditional phase transition, one controlled by variables that drive a system out of equilibrium (*7*–*9*). Here, we present numerical evidence for the same behavior in atomistic models of structural glass formers. We do so with a suitable form of transition path sampling (*10*) that allows us to study ensembles of long trajectories for supercooled fluids with several hundred particles driven out of equilibrium by a field that couples to their mobility. By adjusting field strength, trajectories of these supercooled fluids can be moved reversibly between ergodic and nonergodic behaviors. The former are mobile states with finite relaxation times: the system forgets its initial state. The latter are immobile states that remember initial conditions for all time. At intermediate field strengths, trajectory space is filled by two coexisting domains, one that is ergodic and one that is nonergodic.

In this way, it appears that dynamic heterogeneity observed in the equilibrium dynamics of supercooled fluids is a precursor to a first-order phase transition in space-time. First-order transitions are associated with a discontinuity in an order parameter and a corresponding singularity in a partition function, such as the discontinuity in density for a liquid-vapor transition. These mathematical features emerge from the principles of statistical mechanics in the limit of a very large system, what is usually called the "thermodynamic" limit (*11*). For finite systems studied numerically, there are no such singularities. Evidence of a phase transition in these cases is found in the behaviors of crossovers from one phase to another (*12*). Figure 1 illustrates the system-size behavior of a crossover. For the transition we consider, the partition function is a sum over dynamical histories (i.e., trajectories) of the system, and the order parameter measures the amount of activity or mobility that occurs among N particles in a volume V with trajectories that run for an observation time, t_{obs}. As such, the pertinent measure of system size is a volume in space-time, the product $N \times t_{obs}$ or equivalently $V \times t_{obs}$. In the work reported here, we consider spatial volumes that are 10 to 30 times larger than the correlation volume of the equilibrium system and observation times that are 10 to 100 times longer than a structural relaxation time of the undriven system. These sizes are sufficient to exhibit behaviors suggestive of a nonequilibrium phase transition.

Equilibrium and nonequilibrium phase transitions. To discuss how these behaviors are revealed, let us first recall how Gibbs' statistical mechanics is used to study traditional equilibrium phase transitions (*11*). Taking a system of N particles at a pressure p, we use the volume V as an order parameter and take microstates to be points in configuration space, $x = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$, where the vector $\mathbf{r}_i$ denotes the position of the ith particle. Different phases, such as liquid and vapor, are distinguished from the other by the typical size of V. Changes in V are coupled to the thermodynamic field p or βp, where $1/\beta$ stands for Boltzmann's constant times temperature, $k_B T$. In particular, the probability of a configuration, x, is proportional to $P_0(x)\exp[-\beta \Delta p V(x)]$, where $P_0(x)$ is the probability of x at the reference field or pressure $p_0 = p - \Delta p$. The mean volume of the system with this distribution is $\langle V \rangle_p \equiv V_p$, which is depicted schematically in Fig. 1. A first-order phase transition is manifested by a discontinuity at the pressure $p = p^*$. At this value of the pressure, two phases coexist with respective volumes per particle v_1 and v_2.

At coexistence, the distribution function for the order parameter is bimodal. The two peaks in the distribution coincide with the two equilibrium phases. There is a low probability to observe an intermediate value of V, between Nv_1 and Nv_2 in Fig. 1. This low probability decreases exponen-

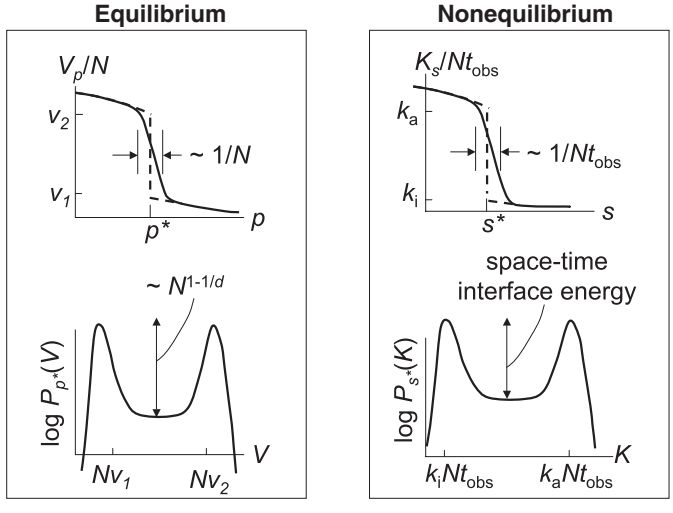


Fig. 1. Finite size effects of equilibrium and nonequilibrium phase transitions. The mean volume V_p manifests an equilibrium first-order phase transition at pressure $p = p^*$, whereas the mean dynamical activity K_s manifests a dynamical first-order phase transition at the dynamical field $s = s^*$. At conditions of phase coexistence, the volume distribution function, $P_p(V)$, and the dynamical activity distribution, $P_s(K)$, are bimodal. Configurations or trajectories with intermediate behaviors lie at much higher free energies (or lower probabilities) than those of the basins. For finite systems, discontinuous phase transitions become crossovers with widths that vanish as system size, N, and observation time, t_{obs}, grow to infinity.

[1]Department of Chemistry, University of California, Berkeley, CA 94720–1460, USA. [2]Department of Physics, University of Bath, Bath BA2 7AY, UK. [3]School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: chandler@cchem.berkeley.edu

tially with the free energy cost to form an interface between the phases. The interfacial free energy grows as $N^{1-1/d}$, where d is the dimensionality. In the limit of a large system, therefore, volumes between Nv_1 and Nv_2 can be achieved only through a direct constraint on the volume. Further, at coexistence, the presence of two macroscopic states means that the mean square fluctuation in the volume grows as N^2. Because the response of the volume to a change in pressure is $-\partial V_p/\partial p = k_B T\langle (V - V_p)^2\rangle_p$, these large fluctuations mean that the width of the crossover illustrated in Fig. 1 vanishes as $1/N$.

Analogous statements for trajectory space begin with a choice of order parameter, which we have taken to be

$$K[x(t)] = \Delta t \sum_{t=0}^{t_{obs}} \sum_{j=1}^{N} |\mathbf{r}_j(t + \Delta t) - \mathbf{r}_j(t)|^2$$

where $\mathbf{r}_j(t)$ and $x(t)$ refer to particle position and point in configuration space, respectively, now as functions of time t. This chosen order parameter depends on the system's path or history over the observation period, $0 \le t \le t_{obs}$. Square brackets are used to indicate that the order parameter is a function of configurations $x(t)$ over the entire period. The incremental time, Δt, is assigned a value for which a particle in a normal liquid would typically move a distance of the order of a molecular diameter. The sum over time is done incrementally, every Δt, thus giving a total of $t_{obs}/\Delta t$ points in time that contribute. When particles are mobile, as in a normal liquid, $K[x(t)]$ is typically large; when particles are immobile, as in a glass, $K[x(t)]$ is typically small. An order-disorder transition reflecting extensive changes in particle mobility is reflected in a discontinuous mean value of $K[x(t)]$.

The next step is to consider the probability distribution for trajectories when this order parameter is coupled to a field s (*13*, *14*). This distribution is proportional to $P_0[x(t)]\exp\{-sK[x(t)]\}$, where $P_0[x(t)]$ is the equilibrium probability distribution, that is, the distribution at $s = 0$. The equilibrium distribution is for trajectories that are causal and time-reversal symmetric and that preserve an equilibrium distribution of microstates. Its partition function is trivial because the distribution is normalized, that is, $1 = \sum_{x(t)} P_0[x(t)]$, where the sum over $x(t)$ is a sum over all trajectories. In contrast, the perturbed distribution

$$P_s[x(t)] \propto P_0[x(t)]\exp\{-sK[x(t)]\}$$

has a nontrivial partition function, which for positive s decreases with increasing t_{obs}. For the space of trajectories governed by that distribution at positive s, this space is compressed with increasing t_{obs}, and, for large enough s, configurations favored by that distribution are ones that are visited by immobile or nonergodic trajectories. $P_s[x(t)]$ is therefore a distribution for trajectories of a system driven out of equilibrium.

Laboratory procedures for forming glass are nonequilibrium processes that stabilize configurations from which equilibration is impossible. One example is the preparation of ultrastable glasses via vapor deposition (*15*). We use the field s as a mathematical device to access these same configurations, configurations that would have negligible statistical weight in an undriven equilibrium dynamics. We do not address how a particular experimental protocol stabilizes these nonergodic configurations. We do, however, address whether the domain of these configurations is sufficiently large to produce a nonequilibrium phase transition.

Transition path sampling of the *s* ensemble. In ordinary molecular dynamics or Monte Carlo trajectories of model systems, trajectories obey detailed balance and are presumed to be ergodic. Their distribution, $P_0[x(t)]$, can be sampled by either running a single trajectory for a time $n \times t_{obs}$ or equivalently carrying out a random walk through trajectory space, sampling n independent trajectories each of duration t_{obs}. The latter procedure is a method of transition path sampling (*10*). To sample $P_s[x(t)]$, we use transition path sampling but now accepting or rejecting random walk steps so as to preserve the weight $P_0[x(t)]\exp\{-sK[x(t)]\}$. We call the collection of trajectories harvested in this way the "s ensemble." In this ensemble, for models with sufficiently correlated dynamics, the distribution function for the order parameter can be bimodal, as indicative of an order-disorder transition (*7*–*9*). This behavior is illustrated schematically in Fig. 1 in a fashion that stresses its analogy with the corresponding behavior of an equilibrium phase transition.

In particular, the average value of the order parameter we have chosen is extensive in space-time, that is, $K_s = \langle K[x(t)]\rangle_s$ is proportional to $N \times t_{obs}$, where the proportionality constant is the

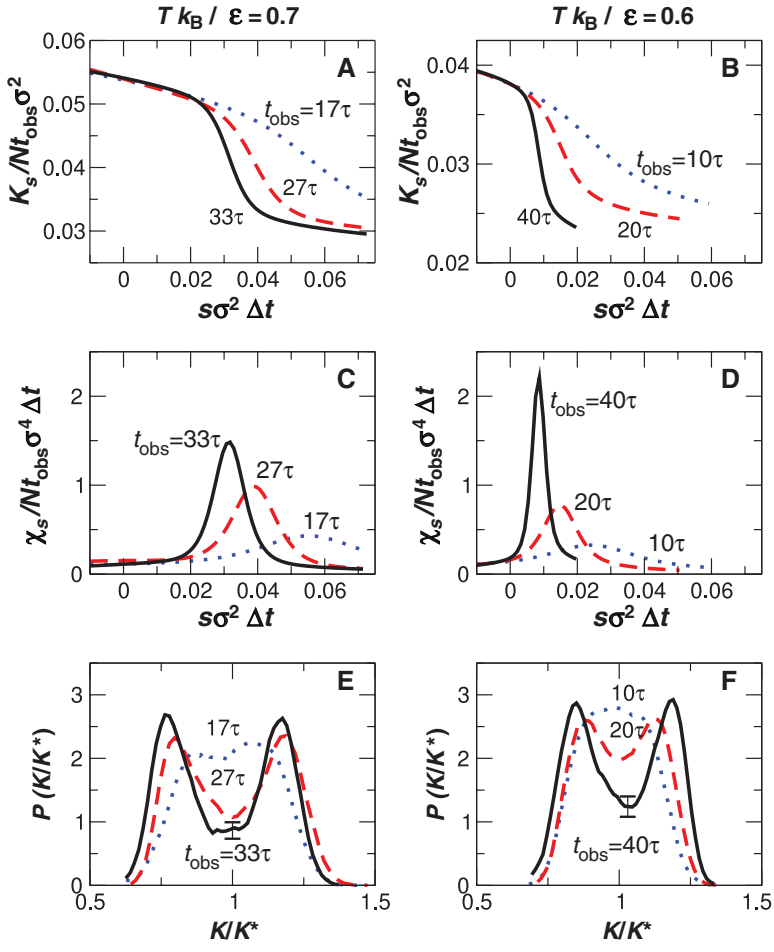


Fig. 2. Evidence for first-order phase transition in space-time. (**A** and **B**) Average space-time order parameter $K_s = \langle K[x(t)]\rangle_s$ as a function of field s, from molecular dynamics simulations of the KA Lennard-Jones mixture for $N = 150$ total particles, at reduced temperatures $k_BT/\varepsilon = 0.6$ and $k_BT/\varepsilon = 0.7$, and principal component density $N_A\sigma^3/V = 0.96$. ε and σ are the Lennard-Jones parameters for the larger (principal component) particles in the KA mixture. As the length of trajectories increases, the crossover in K_s becomes sharper and happens at smaller values of s. (**C** and **D**) The peak in the susceptibility $\chi_s = \frac{\partial K_s}{\partial s}$ becomes larger, and its position moves to smaller s with increasing t_{obs}. The crossover of K_s reflects a first-order transition in the infinite size limit. (**E** and **F**) Distribution of K at coexistence (where $K^* = K_s$ for $s = s^*$). For large t_{obs}, the order parameter distribution at the coexistence field s^* becomes bimodal, as expected for a first-order transition.

mean-square displacement of a particle in an incremental time Δt. If two dynamical phases coexist, one with proportionality constant k_a and the other with k_i, then K_s will be a discontinuous function of s at the condition of coexistence, $s = s^*$. As illustrated with Fig. 1 for a finite system, the corresponding crossover will have width of the order of $1/Nt_{obs}$ because the mean square fluctuations in the order parameter grow as $(Nt_{obs})^2$. Further, the excess in free energy to maintain a coexisting ensemble grows as an interfacial area in space-time, so that trajectories manifesting this coexistence are suppressed by a factor that depends exponentially on this area. In some cases (*8*), this area scales as $N^{1-1/d}t_{obs}$. The value s^* is proportional to the rate at which configurations from the nonergodic "inactive" phase relax back to the ergodic equilibrium fluid if the driving field s is removed. In kinetically constrained lattice models (*9*), which are idealized models of structural glass formers (*16*), this rate is zero, that is, $s^* = 0$. In other words, for those models, undriven equilibrium dynamics coexists with a nonergodic phase. Here, we show that for more realistic atomistic models, and therefore for real glass forming materials, s^* is small although perhaps nonzero.

The particular system we have considered is Kob and Andersen's (KA) two-component mixture of Lennard-Jones particles (*17*). It has $N_A = 0.8N$ principal particles, each with Lennard-Jones diameter σ and energy parameter ε. In addition, it has $N_B = 0.2N$ smaller secondary particles, where their size and attractive energy parameters are chosen so as to frustrate crystallization (*17*). The structural and dynamical properties we report for this model, including the order parameter $K[x(t)]$, refer to the principal particles. We have carried out two independent studies, one where trajectories are governed by Newtonian molecular dynamics and the other where trajectories are governed by a Monte Carlo dynamics. With appropriate scaling of time, both studies yield similar results. The results shown in the figures of this paper are from the molecular dynamics studies. For the incremental time, we use $\Delta t = 13.33(m\sigma^2/48\varepsilon)^{1/2}$, where m is the mass of the particles. Results from the Monte Carlo dynamics plus additional information about our computations are presented in (*18*). In terms of the reduced temperature k_BT/ε, the KA model behaves as an ordinary simple fluid at temperatures $k_BT/\varepsilon > 1$, but its relaxation slows and large glassy fluctuations appear at lower temperatures. Around $k_BT/\varepsilon = 0.4$, relaxation becomes so slow that equilibration of the model on current-day computers becomes intractable. For what is shown below, we work at less severe but nonetheless nontrivial supercooled conditions, $0.6 \leq k_BT/\varepsilon \leq 0.7$.

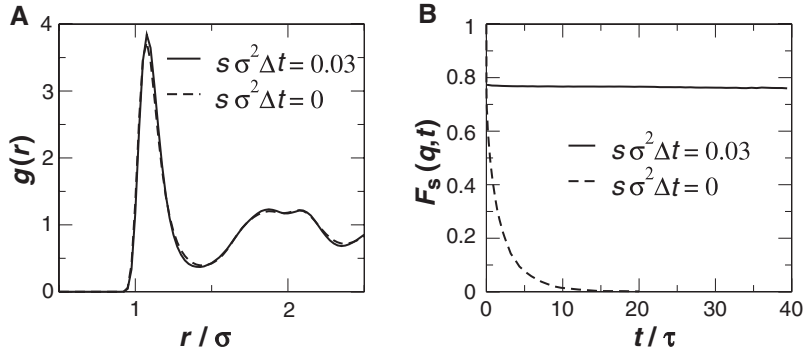


Fig. 3. Comparison of structure and dynamics of active and inactive phases. (**A**) Radial distribution function for the atoms of the principal component in the KA mixture. The equilibrium KA mixture is at $s = 0$. The nonequilibrium mixture is at $s = 0.03/\sigma^2\Delta t > s^*$. Here, σ is the Lennard-Jones diameter for the principal component atoms, and $\Delta t = 13.33(m\sigma^2/48\varepsilon)^{1/2} \simeq \tau/15$, where τ is the structural relaxation time. There is no appreciable difference in the static structures of the active, $s < s^*$, and inactive, $s > s^*$, dynamical phases. (**B**) Self-intermediate scattering function for the same values of s. In the active phase, correlation functions relax to zero. In the inactive phase, correlation functions remain at a nonzero value even for the longest times; the inactive phase is nonergodic. These results were obtained by using simulations of $N = 150$ total particles, at reduced temperature and principal component density $k_BT/\varepsilon = 0.6$ and $N_A\sigma^3/V = 0.96$, respectively.

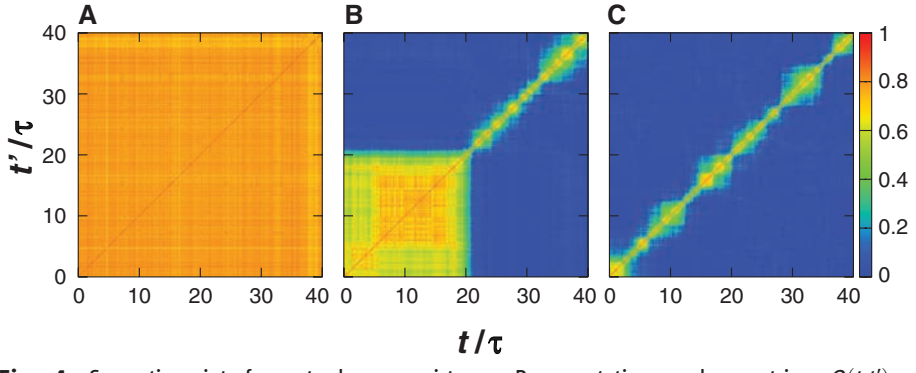


Fig. 4. Space-time interface at phase coexistence. Representative overlap matrices $Q(t,t') = N^{-1}\sum_{i=1}^{N}\cos\{\mathbf{q} \bullet [\mathbf{r}_i(t) - \mathbf{r}_i(t')]\}$ taken from the ensemble of trajectories at the coexistence field s^*, for $k_BT/\varepsilon = 0.6$, $t_{obs} = 40\tau$ and $N = 150$ (Fig. 2B). (**A**) Typical trajectory from the inactive phase (from the low K_s peak in Fig. 2F), where for all observation times the correlation function remains close to 1, indicating nonergodic dynamics. (**B**) Typical trajectory at coexistence (for values of K_s in the trough of Fig. 2F), where the overlap matrix shows a sharp interface-like structure at $t/\tau \approx 20$ separating an inactive region of space-time at earlier times and an active region at later times. (**C**) Typical trajectory from the active phase (from the high K_s peak in Fig. 2F), where the system's dynamics is ergodic and the correlation function decays rapidly to zero. The inhomogeneous features evident in (C) are finite size effects that would vanish in the limit of infinite size, $N \to \infty$.

Although the system does not crystallize under equilibrium conditions, biasing the supercooled KA model toward an inactive phase, as we do with transition path sampling of the s ensemble, can induce crystallization. The effect is pronounced for small periodically replicated systems. It occurs because $K[x(t)]$ by itself does not discriminate between glass and crystal. Although one phase is an equilibrium phase and the other is not, both have low mobility. Thus, in addition to accounting for the value of K, our transition path sampling must also account for a measure of crystallinity. In particular, we use a common neighbor analysis (*19*) and bias against trajectories with this measure of crystallinity (*18*).

Bistability and phase transition in trajectory space. Our findings for the mean order parameter and its distribution in the s ensemble (Fig. 2) reflect the qualitative features associated with a first-order phase transition. The quantitative analysis of finite-size scaling is beyond the capabilities of the algorithms we have used in this work, but the results we have obtained are similar to those established in idealized kinetically constrained models of glass formers (*7*–*9*). The susceptibility

$$\chi_s = -\frac{\partial K_s}{\partial s} = \langle\{K[x(t)] - K_s\}^2\rangle_s$$

has a peak that grows with increasing N and t_{obs}. The peak position is the finite system estimate of s^* (*20*). Its value decreases with increasing N and t_{obs} [system size scaling shown in (*18*)]. The order parameter distribution is bimodal at this value of s, and the minimum between its two peaks decreases with increasing N and t_{obs}. For

the range of values of N and t_{obs}, the effects of increasing t_{obs} are greater than those of increasing N. This asymmetry in the dependence on N and t_{obs} is also reflected in the structure of space-time at coexistence. We will describe this structure but first discuss a measure of ergodicity.

The particular measure we consider is the behavior of the function

$$F_s(q,t) = \frac{1}{Nt_{obs}} \sum_{t'=0}^{t_{obs}} \sum_{i=1}^{N} \langle \exp\{i\mathbf{q}\bullet[\mathbf{r}_i(t+t') - \mathbf{r}_i(t')]\}\rangle_s$$

At equilibrium, that is, $s = 0$, $F_s(q, t)$ is the Van Hove self-correlation or intermediate-scattering function (*21*). In general, it is a mean overlap between configurations displaced by a time t. The extent to which it is nonzero in the limit of large t is a measure of nonergodicity (*22*). We choose the wave vector q to coincide with the first maximum in the structure factor of the liquid (or glass). With this choice, the time τ for which $F_0(q, \tau) = 1/e$ is a common definition of the structural relaxation time. $F_s(q, t)$ is shown in Fig. 3 for the active (i.e., ergodic) and inactive (i.e., nonergodic) phases. The relaxation time of the active trajectories is much less than t_{obs}, yet inactive trajectories remain trapped in a single state throughout the observation time. Figure 3 also shows the radial distribution function for the principal component of the mixture in the active and inactive states. This average measure of structure in the phase that can equilibrate ($s = 0$) is virtually identical to that for the phase that is driven to a nonergodic state ($s > 0$). Thus, fluctuations from the mean structure are crucial to the difference between glassy and fluid materials.

Figure 4 illustrates the structure of trajectory space at conditions of active-inactive coexistence, that is, at $s = s^*$. Each panel illustrates an overlap matrix,

$$Q(t,t') = N^{-1} \sum_{i=1}^{N} \cos\{\mathbf{q}\bullet[\mathbf{r}_i(t) - \mathbf{r}_i(t')]\},$$

and records the similarity or overlap on a given trajectory between configurations at different times t and t'. The s ensemble average of this quantity gives $F_s(q, t - t')$. The pictures in Fig. 4 illustrate the fluctuations from the mean, with the distance from the diagonal reflecting the time interval $t - t'$. At $s = s^*$, there are some trajectories with relatively high values of K indicative of the active phase, others with relatively low values indicative of the inactive phase, and lastly those that are intermediate and that demonstrate phase coexistence. Figure 4C shows a representative trajectory with a value of K corresponding to the active phase; see Fig. 2F. Here, motion is plentiful, and the system quickly decorrelates as $t - t'$ grows.

In contrast, the overlap matrix for inactive trajectories, as illustrated in Fig. 4A, is homogeneous. The system remains correlated for the entire observation time, so that $Q(t, t')$ is large even when $|t - t'|$ is much greater than a structural relaxation time τ. Figure 4B shows a trajectory where the active and inactive phases are separated by a sharp temporal interface. These figures might seem to resemble those found for trajectories of small equilibrium subsystems over time periods that are small compared with the structural relaxation time (*23*, *24*). However, overlap matrices for trajectories of an equilibrated system reveal a structure in space-time in which the system undergoes multiple transitions between collections of states with low energy and activity, commonly termed "meta-basins." Typical lifetimes for these inactive basins are of the order of a fraction of τ. These lifetimes are negligible compared with the very long lifetimes of inactive states within the s ensemble.

Figure 5 illustrates the correlations between the order parameter K and the behaviors of potential energy and icosahedral ordering at the coexistence field s^*. The potential energy of the inactive configurations is smaller, consistent with their stability. For active trajectories, particles sample many configurations, which leads to the self-averaging of structural measures and is reflected in the width of the distribution, which is significantly narrower for greater values of K. However, a clear correlation between potential energy and dynamics does not imply a causal link. Indeed, this thermodynamic variable cannot control the first-order transition that we describe here. Rather, we must look to variables that measure dynamics over a period of time. Although there is a gradual increase in icosahedral ordering (*19*) as K decreases, the observed correlation is far weaker than that of the potential energy. The broad distribution at small K is once again indicative of fluctuation dominance within the inactive phase.

Fluctuations, wetting, and critical points. The existence of a first-order transition has many consequences. For example, fluctuations in an equilibrium system near to phase coexistence grow rapidly as the surface tension between these phases is reduced. Figure 2 indicates that the KA mixture is near a coexistence line between active and inactive phases at both of the temperatures that we considered. Further, as the temperature is reduced, the values of K for the two phases approach each other, indicating that the surface tension between the phases will decrease as temperature decreases. We may therefore associate this decrease with the growth of fluctuations within the active phase. These fluctuations are the dynamical heterogeneities (*5*, *6*) in the equilibrium dynamics of the glass former. A further consequence of phase coexistence is the occurrence of wetting phenomena (*25*). In the case of the dynamical transition discussed here, "wetting" is remembering the initial conditions, and an example of this behavior can be seen for $F_s(q, t)$ in Fig. 3B. At $s > s^*$, the initial time surface is fully wetted by the inactive phase. At $s = 0 < s^*$, there is only a precursor to the wetting transition, a film of finite thickness in time.

For some idealized kinetically constrained models (*7*–*9*), coexistence between active and inactive phases along the $s = 0$ line ends at a $T = 0$ critical point. We expect similarly that, for the KA mixture, the order parameters of both the active and inactive phases will approach the same value, $K_s \to 0$ as $T \to 0$ and $s = 0$. However, where active-inactive coexistence is present in the kinetically constrained models for all temperatures when $s = 0$, the same cannot be true for models with finite intermolecular forces. At high enough temperatures, such forces are insufficiently

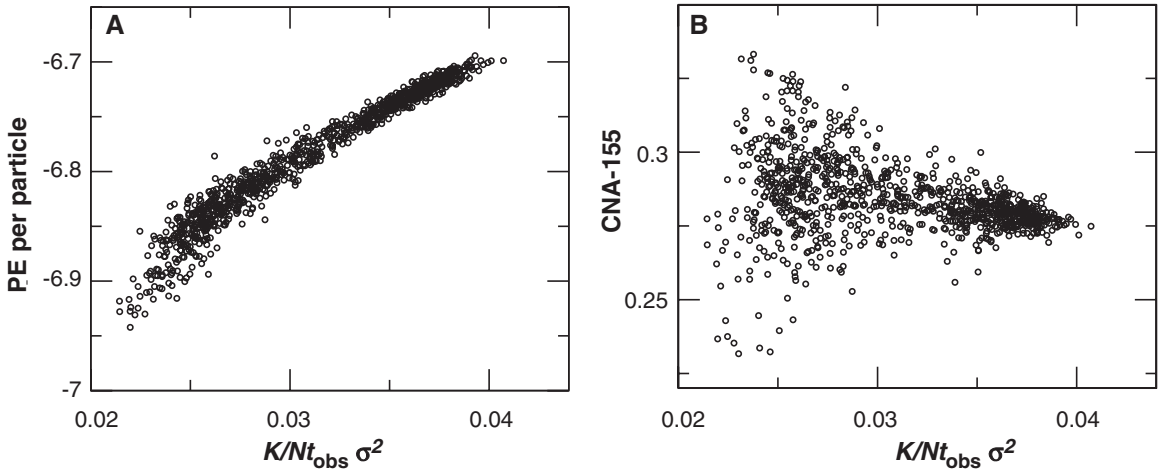


Fig. 5. Test of correlation between space-time order parameter and potential energy and icosahedral order. (**A**) Potential energy per particle in units of ε versus dynamical order parameter K of a trajectory. (**B**) Icosahedral order, as quantified by the CNA-155 parameter of the common neighbor analysis, as a function of K. Results are for trajectories at $k_BT/\varepsilon = 0.6$, $t_{obs} = 40\tau$, and $N = 150$.

constraining to produce collective behavior. Indeed, we have found that at small values of s, the order parameter distribution for the KA mixture ceases to be bimodal when k_BT/ε is significantly larger than 1 (*18*). One possibility is that the first-order coexistence line ends at an upper critical point at finite s and T. This possibility remains to be investigated.

The first-order transition we have described is to be contrasted with the scenario that emerges from other approaches, such as mode-coupling theory (*26*, *27*) and the random first-order transition theory (*28*, *29*). These theories predict the existence of dynamic or thermodynamic transitions controlled by thermodynamic fields such as temperature or pressure. In contrast, our results show that the order-disorder transition is in the trajectories of the dynamics and is thus controlled by dynamic fields. Perhaps a thermodynamic manifestation can be related to the picture of an avoided phase transition (*30*). In any case, our numerical results here suggest that in real glass formers this dynamical order-disorder phenomenon is close to that predicted from idealized kinetically constrained models (*7*–*9*). Thus, we pass the baton to the experimenters to find protocols for controlling the dynamic observable K or driving field s that allow experimental probes of the transition described in this work.

References and Notes

1. M. D. Ediger, C. A. Angell, S. Nagel, *J. Phys. Chem.* **100**, 13200 (1996).
2. P. G. Debenedetti, F. H. Stillinger, *Nature* **410**, 259 (2001).
3. K. Schmidt-Rohr, H. Spiess, *Phys. Rev. Lett.* **66**, 3020 (1991).
4. E. R. Weeks, J. C. Crocker, A. C. Levitt, A. Schofield, D. A. Weitz, *Science* **287**, 627 (2000).
5. M. D. Ediger, *Annu. Rev. Phys. Chem.* **51**, 99 (2000).
6. S. C. Glotzer, *J. Non-Cryst. Solids* **274**, 342 (2000).
7. M. Merolle, J. P. Garrahan, D. Chandler, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10837 (2005).
8. R. L. Jack, J. P. Garrahan, D. Chandler, *J. Chem. Phys.* **125**, 184509 (2006).
9. J. P. Garrahan *et al*., *Phys. Rev. Lett.* **98**, 195702 (2007).
10. P. G. Bolhuis, D. Chandler, C. Dellago, P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
11. D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford Univ. Press, Oxford, 1987).
12. D. P. Landau, K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge Univ. Press, Cambridge, 2000); see section 4.2.3.
13. J.-P. Eckmann, D. Ruelle, *Rev. Mod. Phys.* **57**, 617 (1985).
14. V. Lecomte, C. Appert-Rolland, F. van Wijland, *J. Stat. Phys.* **127**, 51 (2007).
15. S. F. Swallen *et al*., *Science* **315**, 353 (2007); published online 6 December 2006 (10.1126/science.1135795).
16. F. Ritort, P. Sollich, *Adv. Phys.* **52**, 219 (2003).
17. W. Kob, H. C. Andersen, *Phys. Rev. Lett.* **73**, 1376 (1994).
18. Supporting materials are available for this article on *Science* Online.
19. J. D. Honeycutt, H. C. Andersen, *J. Phys. Chem.* **91**, 4950 (1987).
20. C. Borgs, R. Kotecky, *Phys. Rev. Lett.* **68**, 1738 (1992).
21. J. P. Hansen, I. R. McDonald, *The Theory of Simple Liquids* (Academic Press, London, 2006).
22. S. F. Edwards, P. W. Anderson, *J. Phys. F* **5**, 965 (1975).
23. B. Doliwa, A. Heuer, *Phys. Rev. E* **67**, 031506 (2003).
24. G. A. Appignanesi, J. A. Rodríguez Fris, R. A. Montani, W. Kob, *Phys. Rev. Lett.* **96**, 057801 (2006).
25. H. Nakanishi, M. E. Fisher, *Phys. Rev. Lett.* **49**, 1565 (1982).
26. W. Gotze, L. Sjogren, *Rep. Prog. Phys.* **55**, 241 (1992).
27. D. R. Reichman, P. Charbonneau, *J. Stat. Mech.* **2005**, P05013 (2005).
28. T. R. Kirkpatrick, D. Thirumalai, P. Wolynes, *Phys. Rev. A* **40**, 1045 (1989).
29. J. P. Bouchaud, G. Biroli, *J. Chem. Phys.* **121**, 7347 (2004).
30. D. Kivelson, S. A. Kivelson, X. L. Zhao, Z. Nussinov, G. Tarjus, *Physica A* **219**, 27 (1995).

31. During the course of this work we have benefited from discussions with T. F. Miller III and F. van Wijland. The work was made possible through grants from the NSF (CHE-0543158 in the early stages for R.J. and D.C. and CHE-0626305 in the later stages for L.H. and D.C.), from the U.S. Office of Naval Research (N00014-07-1-0689 in the late stages for R.J.), and from the Engineering and Physical Sciences Research Council (GR/S54074/01 for J.P.G.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1166665/DC1
SOM Text
Figs. S1 to S4
References

1 October 2008; accepted 22 January 2009
Published online 5 February 2009;
10.1126/science.1166665
Include this information when citing this paper.

Functional Proteomics Identify Cornichon Proteins as Auxiliary Subunits of AMPA Receptors

Jochen Schwenk,[1]* Nadine Harmel,[1]* Gerd Zolles,[1]* Wolfgang Bildl,[1] Akos Kulik,[4] Bernd Heimrich,[4] Osamu Chisaka,[6] Peter Jonas,[3] Uwe Schulte,[1,2] Bernd Fakler,[1,5]† Nikolaj Klöcker[1]†

Glutamate receptors of the AMPA-subtype (AMPARs), together with the transmembrane AMPAR regulatory proteins (TARPs), mediate fast excitatory synaptic transmission in the mammalian brain. Here, we show by proteomic analysis that the majority of AMPARs in the rat brain are coassembled with two members of the cornichon family of transmembrane proteins, rather than with the TARPs. Coassembly with cornichon homologs 2 and 3 affects AMPARs in two ways: Cornichons increase surface expression of AMPARs, and they alter channel gating by markedly slowing deactivation and desensitization kinetics. These results demonstrate that cornichons are intrinsic auxiliary subunits of native AMPARs and provide previously unknown molecular determinants for glutamatergic neurotransmission in the central nervous system.

Fast excitatory synaptic transmission in the mammalian CNS is mostly mediated by AMPA receptors (AMPARs), ligand-gated ion channels that are activated by glutamate released from the presynaptic terminals (*1*–*4*). On activation, AMPARs provide the transient excitatory postsynaptic current (EPSC) that depolarizes the membrane and initiates downstream processes, such as the generation of action potentials or synaptic plasticity (*5*, *6*). The time course and amplitude of AMPAR-mediated EPSCs exhibit considerable variability among neurons and synapses and strongly depend on the properties of the postsynaptic AMPARs (*7*, *8*).

AMPARs are tetrameric assemblies of α subunits with distinct properties that are encoded by the glutamate receptor (GluR) genes GluR-A to GluR-D (*9*–*11*) [or GluA1-4 according to the International Union of Basic and Clinical Pharmacology nomenclature (*12*)] and their variations resulting from alternative splicing and RNA editing (*13*–*15*). In most central neurons, multiple variants of these GluR proteins are expressed and assembled into heteromultimeric channels that display a wide range of gating kinetics and Ca^{2+} permeabilities (*16*–*19*). In addition to the α subunits, the properties of the AMPARs are modulated by a family of transmembrane AMPAR regulatory proteins (TARPs) (*20*, *21*). The TARPs coassemble with the GluR proteins and through direct protein-protein interactions affect the gating, permeability and pharmacology of the AMPARs (*21*–*25*). Furthermore, the TARPs influence the number and subcellular localization of AMPARs by promoting their trafficking to the plasma membrane and their targeting to the synapse (*26*, *27*).

The profound impact of the TARPs led to the assumption that almost all AMPARs in the mammalian brain may be assembled with these auxiliary subunits (*28*, *29*). However, only a minor portion of the AMPAR complexes in the rat brain (~30%) are associated with γ-2 and γ-3, the TARPs with the most widespread expression pattern (*30*, *31*) (Fig. 1A, arrowhead). It is, therefore, possible that native AMPARs contain further yet-unknown protein constituents that may be iden-

[1]Institute of Physiology II, University of Freiburg, Engesserstrasse 4, 79108 Freiburg, Germany. [2]Logopharm GmbH, Engesserstrasse 4, 79108 Freiburg, Germany. [3]Institute of Physiology I, University of Freiburg, Engesserstrasse 4, 79108 Freiburg, Germany. [4]Institute of Anatomy and Cell Biology, University of Freiburg, Albertstrasse 23, 79104 Freiburg, Germany. [5]Center for Biological Signaling Studies (bioss), Albertstrasse 10, 79104 Freiburg, Germany. [6]Department of Cell and Developmental Biology, Kyoto University, Kyoto 606-8502, Japan.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: bernd.fakler@physiologie.uni-freiburg.de (B.F.) or nikolaj.kloecker@physiologie.uni-freiburg.de (N.K.)

tified in a proteomic approach combining affinity purification of protein complexes and mass spectrometric analysis of their subunit composition. Such comprehensive analysis has successfully been applied to multiprotein complexes associated with different types of ion channels (*32*–*35*) but has not yet been performed with AMPARs.

Proteomic analysis of AMPAR complexes from rat brain. For proteomic analysis of native AMPARs, we used affinity purifications (APs) with three different antibodies specific either for GluR-A and GluR-B (anti–GluR-A, anti–GluR-B) or the TARPs γ-2 and γ-3 (anti–γ-2/3) (fig. S2, B and C) on solubilized membrane fractions prepared from total rat brain (*36*) (fig. S3). These protein preparations contained high-molecular-weight complexes of AMPARs (~0.7 megadalton) assembled from GluR and TARP subunits as visualized by two-dimensional gel separations using blue native polyacrylamide gel electrophoresis (BN-PAGE) and denaturing SDS-PAGE (*36*) (Fig. 1A and fig. S1). Total eluates of APs with the AMPAR subunit–specific antibodies or several pools of preimmunization immunoglobulins G (IgGs) serving as negative controls were analyzed by high-resolution nanoflow liquid chromatography tandem mass spectrometry (nano-LC MS/MS). The results of these MS-analyses showed that all four GluR proteins (GluR-A to GluR-D) were specifically and abundantly [rPQ score and PQ_{norm} score, respectively] (*36*) retained by each of the three AMPAR subunit–specific antibodies (table S1); the peptides retrieved by mass spectrometry provided extensive coverage for the primary sequence of the individual GluR isoforms (61, 75, 56, and 52% for GluR-A to GluR-D, respectively). Moreover, MS/MS spectra obtained from both anti-GluR and anti–γ-2/3 eluates identified five members of the TARP family, with significant peptide yields for γ-2 and γ-3, and smaller yields for γ-4, γ-7, and γ-8 (table S1). A sixth TARP isoform, γ-5, was only observed in eluates of anti–GluR-A APs, albeit in small amounts (table S1).

In addition to the known AMPAR subunits, our MS analyses consistently identified cornichon homolog 2 (CNIH-2) (Fig. 1B) and cornichon homolog 3 (CNIH-3), closely related members of a conserved family of small transmembrane proteins that was first described in *Drosophila* (Fig. 1, C and D) (*37*–*39*). Both cornichon proteins were specifically copurified at high yield with all three AMPAR subunit–specific antibodies (table S1), which suggests that CNIH-2 and CNIH-3 are robustly integrated into AMPAR complexes in rat brain.

Coassembly of native and heterologously expressed AMPARs and cornichon proteins. Coassembly of the two cornichon proteins with native AMPARs was confirmed by subsequent reverse-purification using an antibody specific for CNIH-2 and CNIH-3 (anti–CNIH-2/3) (fig. S2A) (*36*) on membrane fractions from rat brain. The respective eluates, besides the cornichon homologs, contained the AMPAR α subunits GluR-A to GluR-D, as well as the TARP isoforms γ-2, γ-3, and γ-4 (table S1). In addition, the cornichon-AMPAR assembly was corroborated by Western probing of the two-dimensional gel separation of the antibody-shift assay from Fig. 1A with the anti–CNIH-2/3 antibody; this assay separated the γ-2/3–associated AMPARs from those devoid of these TARPs by the additional mass introduced via target-specific binding of the anti–γ-2/3 IgG before the BN-PAGE. The major portion of the two cornichon proteins was not shifted by the γ-2/3 antibody, which indicated that CNIH-2 and CNIH-3 are predominantly assembled into γ-2/3–free AMPAR complexes (Fig. 2A). These γ-2/3–free AMPAR complexes were effectively and completely shifted when anti–CNIH-2/3 IgG was used

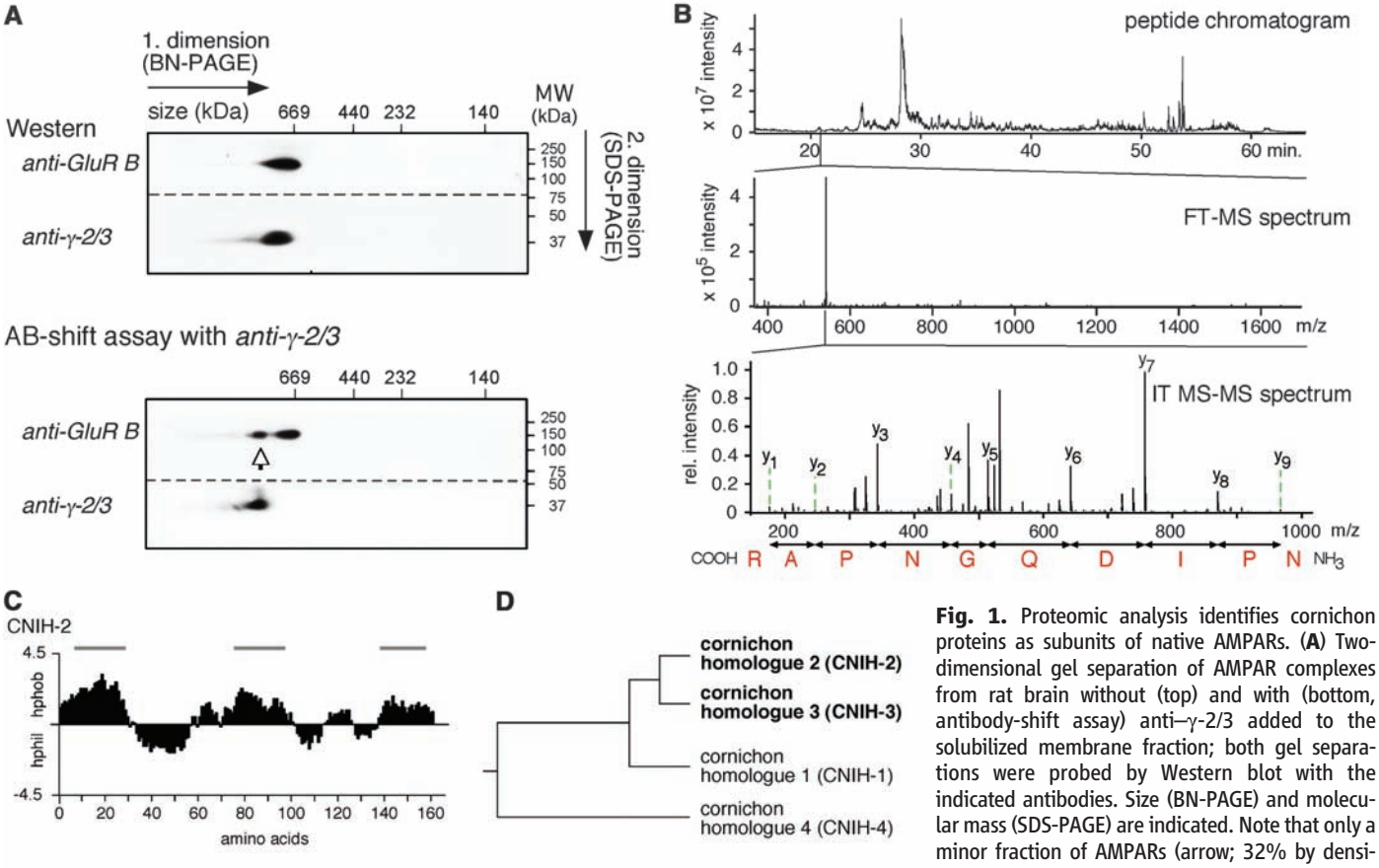


Fig. 1. Proteomic analysis identifies cornichon proteins as subunits of native AMPARs. (**A**) Two-dimensional gel separation of AMPAR complexes from rat brain without (top) and with (bottom, antibody-shift assay) anti–γ-2/3 added to the solubilized membrane fraction; both gel separations were probed by Western blot with the indicated antibodies. Size (BN-PAGE) and molecular mass (SDS-PAGE) are indicated. Note that only a minor fraction of AMPARs (arrow; 32% by densitometric analysis) is shifted by the anti–γ-2/3 antibody. (**B**) (Top) High-performance liquid chromatography chromatogram of peptide fragments of an anti–GluR-B eluate. (Middle and bottom) MS- and MS/MS-spectra of a peptide unique for CNIH-2 (*m/z* value of 541.27246). The complete MS/MS fragment y^+-ion series is indicated together with the amino acid sequence derived from the mass differences (in carboxy-to-amino-terminal direction). (**C**) Hydropathy plot (Kyte-Doolittle, window of 12 amino acids) of the rat CNIH-2 protein; horizontal bars denote sequence stretches long enough to span the membrane. (**D**) Dendrogram (Clustal method) of the cornichon family of proteins.

in the antibody-shift assay (fig. S4), which strongly suggested that all AMPARs not associated with TARPs are coassembled with CNIH proteins.

The participation of cornichons and TARPs in native AMPARs was further investigated by relative quantification of CNIH-2 and γ-2 protein in the aforementioned APs. The mass traces (*m/z* peak volumes) of peptides unique for CNIH-2 and γ-2 were quantified on the basis of calibration curves determined for these peptides in a dilution series with heterologously expressed tagged versions of both proteins (*36*). In APs with anti–GluR-A and anti–GluR-B, the ratio of copurified CNIH-2/γ-2 was between 6/1 and 7/1 (Fig. 2B) or, equivalently, ~85% of the purified AMPARs were associated with CNIH-2, whereas ~15% partnered with the TARP γ-2. The excess of copurified CNIH-2 over γ-2 was independent of the GluR composition of the AMPARs (Fig. 2B) and suggested that cornichons, similar to TARP proteins, may be directly assembled with the GluR subunits (*25*). This was tested in APs with anti–GluR-A and anti–CNIH-2/3 on membrane fractions prepared from *Xenopus* oocytes and cultured cells (*36*) that heterologously expressed CNIH-2 and heteromeric GluR-A/GluR-B AMPARs either alone or in combination. Robust and specific copurification of GluR-A and CNIH-2 was observed with both antibodies when used on oocytes coexpressing AMPARs and CNIH-2 (Fig. 2C).

Together, proteomic and biochemical analyses indicated that the cornichon proteins CNIH-2 and CNIH-3 are integral constituents of the majority of native AMPARs. They are intimately associated with the pore-forming GluR subunits.

Expression profile of cornichons in the CNS. Next, the expression profile of CNIH-2 and CNIH-3 in the rat brain was investigated by immunohistochemistry using the anti–CNIH-2/3 antibody. Anti–CNIH-2/3 immunoreactivity was observed throughout most regions of the brain. Examples of expression of CNIH-2/3 in the neocortex, hippocampal formation, and cerebellum are depicted in Fig. 3. In all these areas, CNIH-2/3 immunoreactivity was found in various types of neurons, including neocortical and hippocampal pyramidal cells and cerebellar Purkinje cells, as well as in glial cells, such as Bergmann glia in the cerebellum or astrocytes in the hippocampus (Fig. 3, A and B). Neither of the two cornichon proteins was detected in cerebellar granule cells [(Fig. 3A), right], where AMPAR-mediated synaptic transmission crucially depends on the presence of the TARP γ-2 (*40*, *41*). In the hippocampal CA1 region, the anti–CNIH-2/3 immunoreactivity was localized to the plasma membrane of both postsynapses and extrasynaptic sites (dendritic shafts, spines of pyramidal cells), as seen with post-embedding immunogold electron microscopy (EM) (Fig. 3C).

Enhanced surface expression of AMPARs by cornichons. Auxiliary subunits affect both processing and biophysical characteristics of the pore-forming α subunits of various ion channels (*42*), which prompts respective analyses for the cornichon proteins. To examine CNIH-mediated effects on AMPAR trafficking, we used heterologous expression of the flop splice variant of GluR-A (GluR-A_o) in cultured cells and *Xenopus* oocytes either alone or together with CNIH-2 or CNIH-3. Surface expression of the resulting AMPARs was monitored either by staining a hemagglutinin (HA) epitope in the extracellular N terminus of GluR-A_o (cultured cells) or by recording glutamate-activated currents in whole oocytes (*36*). Coexpression of either cornichon isoform markedly enhanced the HA-based surface immunoreactivity (Fig. 4A); quantification yielded an ~10-fold increase of GluR-A_o receptors in the plasma membrane of the CNIH-expressing cells (Fig. 4B). Similarly, in *Xenopus* oocytes coexpression of CNIH-2 or CNIH-3 substantially increased the GluR-A_o–mediated currents evoked by application of a saturating concentration of glutamate (1 mM) together with trichloromethiazide (TCM), an inhibitor of channel desensitization (*43*) (Fig. 4, C and D). In either system, the CNIH-mediated enhancement was highest at low expression levels of the homomeric AMPAR [(Fig. 4A), left] but decreased as the GluR-A_o expression was increased. Together, these results indicated that coassembly with CNIH-2 and CNIH-3 effectively promoted expression of the AMPAR complexes at the plasma membrane.

Modulation of AMPAR gating by cornichons. In addition to protein processing, whole-oocyte recordings showed that both cornichons markedly reduced the increase in current amplitude induced by TCM in the presence of glutamate (Fig. 4C). This reduction suggested that CNIH-2/3 may also influence the gating properties of the AMPARs.

The impact of the cornichon proteins on channel gating was investigated in giant outside-out (oo) patches from *Xenopus* oocytes with rapid glutamate application via a piezo-controlled fast application system (*36*, *44*). Figure 5A shows typical current transients recorded with 1-ms glutamate pulses on AMPARs assembled from GluR-A_i and GluR-B_i (GluR-A_i/B_i), a particularly abundant subunit combination in the CNS (*16*), either alone or in combination with the accessory subunits γ-2 or CNIH-2. All three types of AMPARs activated rapidly with similar values (means ± SD) for the 20 to 80% rise time (0.25 ± 0.03 ms, n = 11 for GluR-A_i/B_i; 0.32 ± 0.03 ms,

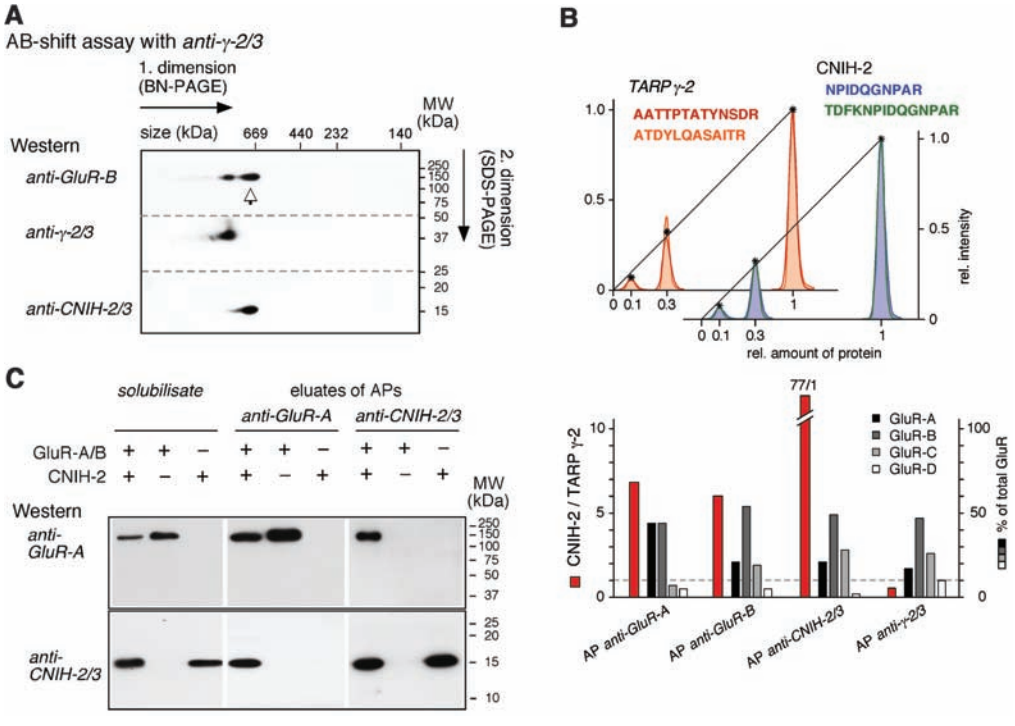


Fig. 2. Cornichons are major partners of AMPARs in the rat brain and directly coassemble with the AMPAR α subunits. (**A**) Antibody-shift assay from Fig. 1A probed by Western blot with the anti–CNIH-2/3 antibody. Note that the major portion of AMPARs (arrow) is associated with the cornichon proteins. (**B**) Relative quantification of AMPAR subunits in APs with the indicated antibodies. (Top) Calibration curves of *m/z* peak volumes used for relative quantification of CNIH-2 and γ-2 (*16*). Asterisks are intensity-weighted means of *m/z* peak volumes, lines represent linear regressions to these mean values. (Bottom) Molar ratio of CNIH-2 versus γ-2 (red bars) and GluR subunit composition in the indicated APs. (**C**) Copurification of AMPARs and CNIH-2 from *Xenopus* oocytes expressing GluR-A_i, GluR-B_i and CNIH-2. Solubilisate and eluates from APs with anti–GluR-A and anti–CNIH-2/3 were separated by SDS-PAGE and probed by Western blot with the indicated antibodies.

$n = 13$; 0.28 ± 0.04 ms, $n = 6$, for AMPARs with γ-2 and CNIH-2, respectively); however, they differed considerably in their deactivation kinetics, as reflected by the current decay when the agonist was removed (Fig. 5A, inset). Thus, either accessory subunit slowed the deactivation of GluR-A_i/B_i receptors, but, whereas slowing by γ-2 was moderate as reported earlier (*21*, *22*, *24*), deceleration by CNIH-2 was substantial and

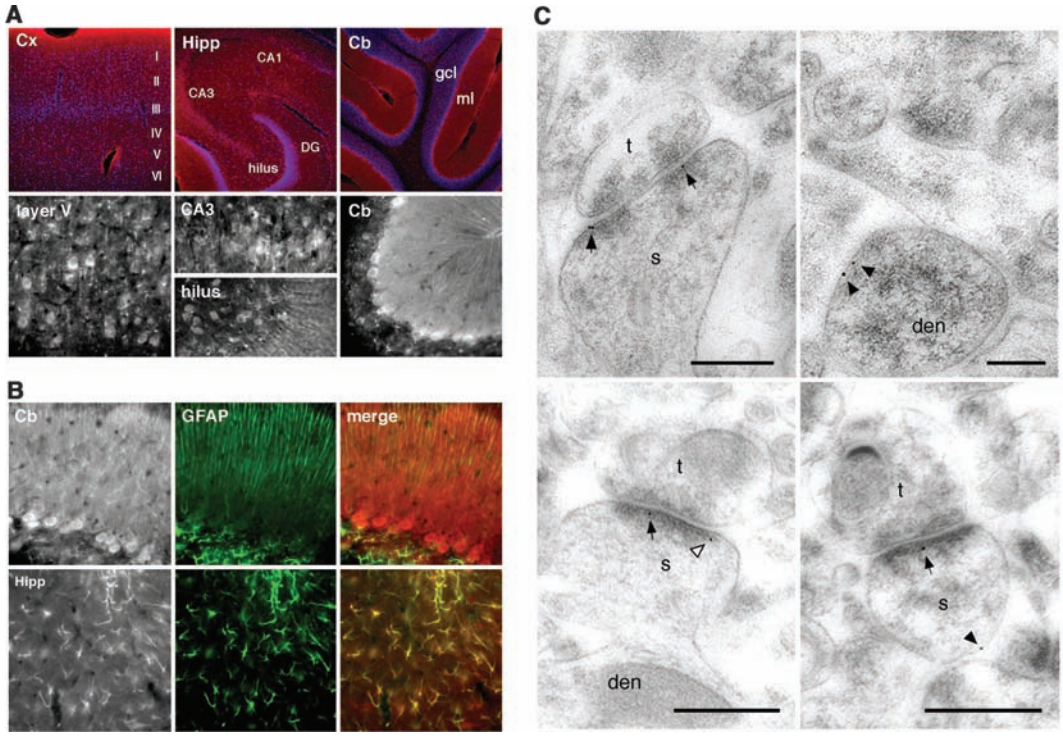


Fig. 3. Expression profile of CNIH-2 and CNIH-3 in CNS neurons and glial cells. (**A**) Immunostaining (red, top; white, bottom) of neocortex (Cx), hippocampal formation (Hipp) and cerebellum (Cb) with anti–CNIH-2/3. Immunoreactivity was observed in various cell types throughout the cortical layers (including both pyramidal and stellate cells in layer V), in CA1/CA3 pyramidal cells and subpopulations of hilar neurons in the hippocampus (including mossy cells), and in Purkinje neurons, as well as in the molecular layer (ml) of the cerebellum. Granule cells in the cerebellum (granule cell layer, gcl) were not stained. DG, dentate gyrus. DAPI staining of nuclei is in blue. (**B**) CNIH-2/3 immunoreactivity in radial Bergmann glia in the cerebellum (upper left) and hippocampal astrocytes (lower left) identified by double-labeling with an antibody against the glial fibrillary acidic protein (GFAP, middle, and both signals merged, right). (**C**) Electron micrographs of CNIH-2/3 immunoreactivity in the CA1 region of the adult rat hippocampus detected by post-embedding immunogold-EM. Immunoparticles were found over (arrows) and at the edge (open arrowhead) of asymmetrical synapses between axon terminals (t) and dendritic spines (s) of pyramidal cells. Gold particles were also found at the extrasynaptic plasma membrane (filled arrowheads) of dendritic shafts (den) and spines of pyramidal cells. Scale bars, 0.2 μm.

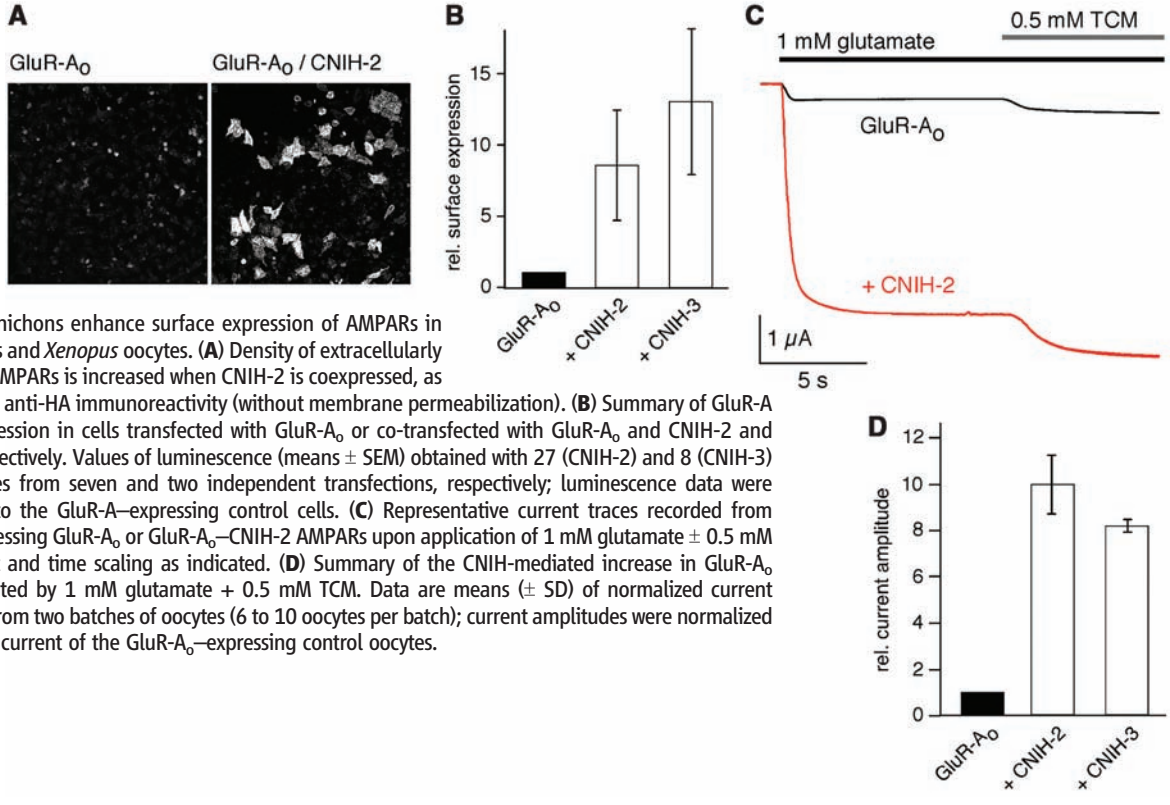
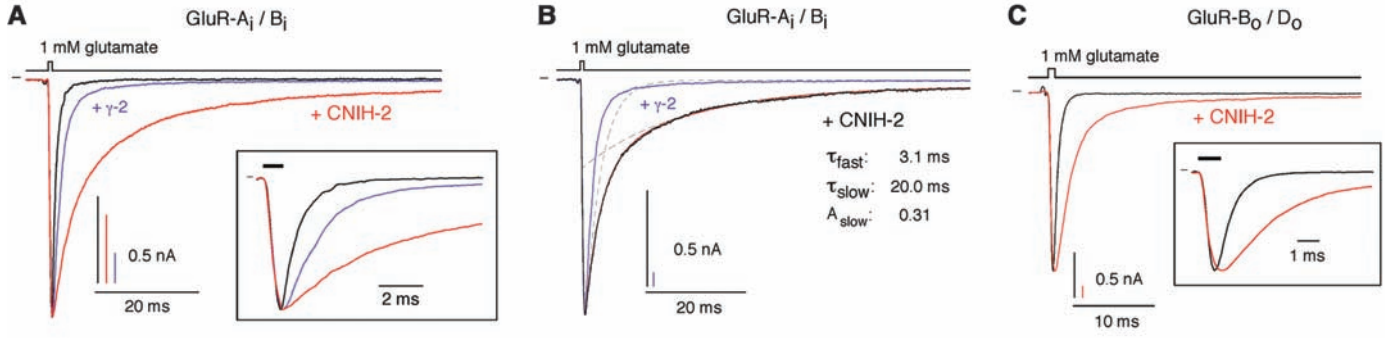


Fig. 4. Cornichons enhance surface expression of AMPARs in cultured cells and *Xenopus* oocytes. (**A**) Density of extracellularly HA-tagged AMPARs is increased when CNIH-2 is coexpressed, as visualized by anti-HA immunoreactivity (without membrane permeabilization). (**B**) Summary of GluR-A surface expression in cells transfected with GluR-A_o or co-transfected with GluR-A_o and CNIH-2 and CNIH-3, respectively. Values of luminescence (means ± SEM) obtained with 27 (CNIH-2) and 8 (CNIH-3) culture dishes from seven and two independent transfections, respectively; luminescence data were normalized to the GluR-A–expressing control cells. (**C**) Representative current traces recorded from oocytes expressing GluR-A_o or GluR-A_o–CNIH-2 AMPARs upon application of 1 mM glutamate ± 0.5 mM TCM; current and time scaling as indicated. (**D**) Summary of the CNIH-mediated increase in GluR-A_o currents elicited by 1 mM glutamate + 0.5 mM TCM. Data are means (± SD) of normalized current amplitudes from two batches of oocytes (6 to 10 oocytes per batch); current amplitudes were normalized to the mean current of the GluR-A_o–expressing control oocytes.

varied according to its expression level (Fig. 5A and fig. S5). Closer analysis (of the current decay) revealed that the slowing of deactivation by γ-2 and CNIH-2 resulted from a more complex process of channel closure. Although the deactivation time course of GluR-A_i/B_i receptors was adequately described by a single exponential function (time constant of 0.96 ± 0.24 ms, n = 15) (Fig. 5D), two exponential components were required for adequate fitting of the current decay in accessory subunit–containing GluR-A_i/B_i receptor complexes (Fig. 5, B and D) [see also (*36*)]. The distinct effects of γ-2 and CNIH-2 on deactivation kinetics predominantly resulted from the distinct amplitudes of the respective slow component, which was 8 ± 4% (n = 5) in γ-2–containing complexes, but 30 ± 7% (n = 13) in GluR-A_i/B_i–CNIH-2 receptors ($P < 0.01$, Mann-Whitney U test) (Fig. 5D). At lower expression levels of CNIH-2, this slow component was largely reduced, and deactivation could be approximated with a monoexponential function albeit with a time constant for deactivation ($\tau_{deactivation}$) slower than that of GluR-A_i/B_i (fig. S5). As CNIH-2 did, CNIH-3 also prolonged deactivation kinetics of heteromeric GluR-A_i/B_i receptors (Fig. 5D).

The impact of the cornichon proteins on channel gating was also examined in AMPARs assembled from the flop splice variants of GluR-B and GluR-D (GluR-B_o/D_o), heteromeric channels that display the fastest deactivation and desensitization kinetics of AMPARs (*19*). Both CNIH-2 and CNIH-3 slowed the deactivation kinetics ~3.6-fold (Fig.



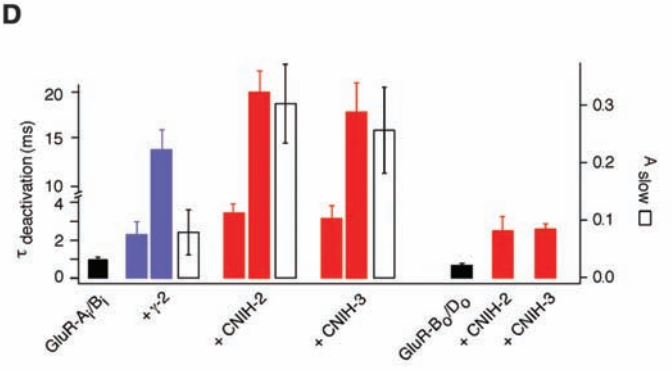


Fig. 5. Cornichons slow deactivation kinetics of AMPARs. (**A**) Representative current responses of AMPARs recorded upon 1-ms applications of 1 mM glutamate (indicated above the current trace) in giant oo-patches excised from *Xenopus* oocytes expressing GluR-A_i/B_i (black trace) or coexpressing GluR-A_i/B_i and either γ-2 (blue trace) or CNIH-2 (red trace). All complementary RNAs were injected at equal amounts; current and time scaling as indicated. (Inset) Current responses at expanded time scale; agonist application indicated by the horizontal bar. (**B**) Channel deactivation of GluR-A_i/B_i–CNIH-2 AMPAR complexes is a bi-exponential process. Continuous red line is fit of the sum of two exponentials (dashed lines, single components) with the time constants (τ_{fast}, τ_{slow}) and relative amplitude of the slow component (A_{slow}) as indicated. The current response of GluR-A_i/B_i–γ-2 complexes from (A) is shown for comparison. (**C**) Representative responses of GluR-B_o/D_o and GluR-B_o/D_o–CNIH-2 AMPARs upon 1-ms applications of 1 mM glutamate in experiments as in (A). (**D**) Summary of the fit parameters of channel deactivation obtained with AMPARs of the indicated molecular composition. Data are means (± SD) of 6 to 15 patches. Time constants are shown as solid bars, open bars denote A_{slow} in GluR-A_i/B_i–containing AMPARs.

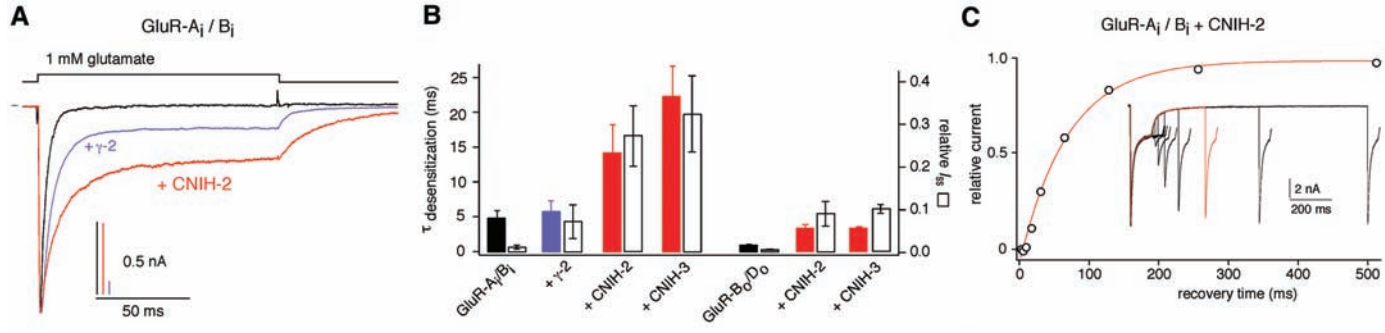


Fig. 6. Cornichons reduce desensitization of AMPARs without affecting recovery from desensitization. (**A**) Superimposed current responses of the indicated AMPARs to a 100-ms application of 1 mM glutamate (indicated above the current trace) in experiments as in Fig. 5. (**B**) Summary of the values for $\tau_{desensitization}$ (solid bars) and the relative amplitude of the nondesensitizing current component (relative I_{ss}, white bars) obtained from monoexponential fits to the current decay recorded from the indicated AMPARs in experiments as in (A). Data are means (± SD) of 6 to 15 patches. (**C**) Recovery of GluR-A_i/B_i–CNIH-2 AMPARs from steady-state desensitization recorded with a double-pulse protocol (pair of a 100-ms and a 50-ms glutamate pulse separated by increasing time intervals) in a giant oo-patch. Data points are peak currents recorded during the second pulse and normalized to the maximal current (recorded during the first glutamate application). Red line is the result of a monoexponential fit to the data points ($\tau_{recovery}$ = 69.2 ms). (Inset) Original current recordings; red trace is response with a recovery interval of 128 ms. (**D**) Values for $\tau_{recovery}$ obtained from fits as in (C) with the indicated AMPARs. Data are means (± SD) of three to eight patches.

5, C and D) ($\tau_{deactivation}$ of 0.71 ± 0.11 ms, n = 8 for pure GluR-B_o/D_o; 2.56 ± 0.74 ms, n = 11 and 2.61 ± 0.29 ms, n = 7 for GluR-B_o/D_o receptors coassembled with CNIH-2 and CNIH-3, respectively; $P < 0.001$ for CNIH-mediated effects, Mann-Whitney U test). As before, channel activation appeared largely unaffected by the CNIH proteins, as indicated by the similar values obtained for the 20 to 80% rise time in pure and CNIH-2–associated GluR-B_o/D_o receptors (values of 0.23 ± 0.02 ms, n = 8 for GluR-B_o/D_o; 0.28 ± 0.04 ms, n = 11 and 0.28 ± 0.05 ms, n = 7 for the receptors associated with CNIH-2 and CNIH-3, respectively). The differences in the time-to-peak interval observed between GluR-B_o/D_o receptors expressed alone and their CNIH-associated counterparts [(Fig. 5C), inset] likely reflected effects of the cornichons on both deactivation and desensitization, rather than differences in activation kinetics between both types of receptors (*45*).

In a second set of experiments, we therefore investigated the effects of CNIH-2 and CNIH-3 on the desensitization of GluR-A_i/B_i and GluR-B_o/D_o AMPARs, using 100-ms pulses of 1 mM glutamate. Representative current responses of such experiments (Fig. 6A) illustrate the changes caused by the coassembly of GluR-A_i/B_i receptors with CNIH-2. Thus, CNIH-2 markedly slowed the desensitization-mediated decay of the current and introduced a prominent nondesensitizing steady-state current component (Fig. 6A). When analyzed by monoexponential fits, the changes mediated by CNIH-2 and CNIH-3 in GluR-A_i/B_i receptors quantified to a more than threefold increase of the desensitization time constant ($\tau_{desensitization}$) and to a steady-state current of roughly 30% of the peak current (Fig. 6B); both the change in $\tau_{desensitization}$ and the amplitude of the nondesensitizing current depended on the expression level of CNIH-2 (fig. S5). A similar relative increase in $\tau_{desensitization}$ was obtained with the GluR-B_o/D_o receptors when assembled with the two cornichon isoforms (values for $\tau_{desensitization}$ of 0.92 ± 0.14 ms, n = 8 for GluR-B_o/D_o; 3.36 ± 0.51 ms, n = 11 and 3.48 ± 0.18 ms, n = 6 for GluR-B_o/D_o–CNIH-2 and GluR-B_o/D_o–CNIH-3, respectively; $P < 0.001$ for CNIH-mediated effects, Mann-Whitney U test); the respective steady-state current, however, was markedly smaller than in GluR-A_i/B_i receptors (~10% for both cornichon isoforms, Fig. 6B).

Because the CNIH proteins slowed desensitization, the reverse process, recovery from desensitization, was investigated with a double-pulse protocol in GluR-A_i/B_i and GluR-B_o/D_o receptors (*46*). As exemplified for GluR-A_i/B_i coassembled with CNIH-2, recovery from desensitization was complete, and the time course of recovery was adequately described by a single exponential (Fig. 6C). Unlike the gating transitions described above, recovery from desensitization was largely independent of associated CNIH proteins (Fig. 6D).

Taken together, the results on channel kinetics indicated that the CNIH proteins extensively modify the gating properties of AMPARs, probably by stabilizing the open state of the receptor channels; this stabilizing effect promotes slowing of deactivation and desensitization, without major effects on channel activation or recovery from desensitization. Moreover, the channel gating observed with heterologously expressed GluR-A_i/B_i–CNIH receptors closely resembles that of their native counterparts from hippocampal mossy cells and CA3 pyramidal cells (fig. S5) (*16*).

Discussion. We identified cornichon homologs 2 and 3 as intrinsic auxiliary subunits of the majority of AMPAR complexes in the mammalian brain. Physical association of CNIH-2 and CNIH-3 with the pore-forming GluR proteins promotes surface expression of AMPARs and extensively modulates their gating properties by slowing deactivation and desensitization kinetics.

CNIH proteins—novel auxiliary subunits of AMPARs. For comprehensive analysis of the molecular composition of native AMPARs, we used a proteomic approach that combines APs of appropriately solubilized protein complexes (controlled by BN-PAGE) (Figs. 1A and 2A) with nano-LC MS/MS analysis of total eluates. When applied to membrane preparations from total rat brain, this procedure isolated the expected set of known AMPAR subunits including the four GluR proteins, as well as six members of the TARP family (table S1). This unbiased approach also revealed two unexpected results. First, it identified the two cornichon proteins CNIH-2 and CNIH-3 as copurified partners of native AMPARs (Fig. 1 and table S1); neither of these proteins has previously been implicated in AMPAR physiology. Second, by relative quantification of MS-data and electrophoretic analyses (Fig. 2), the proteomic approach showed that the two CNIH proteins coassemble with the majority of AMPARs in the rat brain. We estimate that ~70% of AMPARs contain cornichons as accessory subunits, whereas ~30% coassemble with TARPs (Fig. 2). Antibody-shift assays (Fig. 2 and fig. S4) and APs with the anti–γ-2/3 and anti–CNIH-2/3 antibodies suggest that a minor portion of AMPARs may be coassembled with both TARPs and cornichons; in the majority of AMPARs, however, both accessory subunits seemed to be mutually exclusive.

Subsequent functional analysis showed that both cornichon proteins affect surface expression and gating properties of AMPARs in the plasma membrane (Figs. 4 to 6). These AMPAR-related functions are different from what has been reported for this family of proteins in *Drosophila*, chicken, and yeast, where cornichons were shown to operate as cargo receptors for the export of certain growth factors from the endoplasmic reticulum (*37*, *38*, *47*). Thus, it appears that the members of the cornichon family of transmembrane proteins may have multiple functions in cell physiology.

Implication for AMPAR-mediated signaling in the CNS. The influence of CNIH-2 and CNIH-3 on the gating of AMPARs may be viewed as stabilization of the open state impairing channel closure either upon agonist removal or upon conformational processes that trigger receptor desensitization (*48*). As a consequence, the time course of both deactivation and desensitization was slowed by up to several-fold in AMPARs of various subunit composition (Figs. 5 and 6).

Immuno-EM on the hippocampal CA1 region (Fig. 3C) suggests that CNIH proteins may be incorporated into both postsynaptic and extrasynaptic AMPARs. If coassembled into postsynaptic AMPARs, the cornichons will slow the decay time course of EPSCs, often determined by AMPAR deactivation kinetics. If present in extrasynaptic receptors, CNIH-2 and 3 may enhance the effects of glutamate spillover from the release sites to more distant locations, leading to the activation of receptors that would otherwise desensitize. Such spillover effects may be relevant in mossy fiber synapses on hippocampal CA3 pyramidal neurons and hilar mossy cells and in parallel and climbing fiber synapses on cerebellar Purkinje cells, in which the presynaptic elements form closely spaced release sites (*49*, *50*) and the postsynaptic neurons abundantly express the two cornichon proteins (Fig. 3). Thus, effects on postsynaptic and extrasynaptic receptors will slow EPSCs, prolonging the time course of excitatory postsynaptic potentials and thus enhancing temporal summation of synaptic events (fig. S6). Whether neurons specialized on synaptic integration (such as hippocampal pyramidal cells) and others specialized on coincidence detection (such as neurons in the auditory pathway or GABAergic interneurons in the cortex) differentially express the cornichon proteins remains to be determined. In conclusion, our results establish cornichons CNIH-2 and CNIH-3 as a new class of accessory AMPAR subunits and thus provide novel molecular determinants for the modulation of neurotransmission in the CNS.

References and Notes

1. P. Jonas, N. Spruston, *Curr. Opin. Neurobiol.* **4**, 366 (1994).
2. I. M. Raman, L. O. Trussell, *Neuron* **9**, 173 (1992).
3. P. Sah, S. Hestrin, R. A. Nicoll, *J. Physiol.* **430**, 605 (1990).
4. R. A. Silver, S. F. Traynelis, S. G. Cull-Candy, *Nature* **355**, 163 (1992).
5. M. Sheng, M. J. Kim, *Science* **298**, 776 (2002).
6. I. Song, R. L. Huganir, *Trends Neurosci.* **25**, 578 (2002).
7. F. Conti, R. J. Weinberg, *Trends Neurosci.* **22**, 451 (1999).
8. P. Jonas, *News Physiol. Sci.* **15**, 83 (2000).
9. J. Boulter *et al.*, *Science* **249**, 1033 (1990).
10. K. Keinänen *et al.*, *Science* **249**, 556 (1990).
11. S. Nakanishi, *Science* **258**, 597 (1992).
12. G. L. Collingridge, R. W. Olsen, J. Peters, M. Spedding, *Neuropharmacology* **56**, 2 (2009).
13. R. Dingledine, K. Borges, D. Bowie, S. F. Traynelis, *Pharmacol. Rev.* **51**, 7 (1999).
14. P. H. Seeburg, M. Higuchi, R. Sprengel, *Brain Res. Brain Res. Rev.* **26**, 217 (1998).
15. B. Sommer *et al.*, *Science* **249**, 1580 (1990).
16. J. R. P. Geiger *et al.*, *Neuron* **15**, 193 (1995).
17. M. Hollmann, M. Hartley, S. Heinemann, *Science* **252**, 851 (1991).
18. H. Lomeli *et al.*, *Science* **266**, 1709 (1994).
19. J. Mosbacher *et al.*, *Science* **266**, 1059 (1994).
20. A. S. Kato, E. R. Siuda, E. S. Nisenbaum, D. S. Bredt, *Neuron* **59**, 986 (2008).
21. A. D. Milstein, W. Zhou, S. Karmizadegan, D. S. Bredt, R. A. Nicoll, *Neuron* **55**, 905 (2007).

22. C. H. Cho, F. St-Gelais, W. Zhang, S. Tomita, J. R. Howe, *Neuron* **55**, 890 (2007).
23. D. Soto, I. D. Coombs, L. Kelly, M. Farrant, S. G. Cull-Candy, *Nat. Neurosci.* **10**, 1260 (2007).
24. S. Tomita *et al*., *Nature* **435**, 1052 (2005).
25. W. Vandenberghe, R. A. Nicoll, D. S. Bredt, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 485 (2005).
26. C. Bats, L. Groc, D. Choquet, *Neuron* **53**, 719 (2007).
27. L. Chen *et al*., *Nature* **408**, 936 (2000).
28. K. Menuz, G. A. Kerchner, J. L. O'Brien, R. A. Nicoll, *Neuropharmacology* **56**, 22 (2009).
29. K. Menuz, J. L. O'Brien, S. Karmizadegan, D. S. Bredt, R. A. Nicoll, *J. Neurosci.* **28**, 8740 (2008).
30. M. Fukaya, M. Yamazaki, K. Sakimura, M. Watanabe, *Neurosci. Res.* **53**, 376 (2005).
31. N. Klugbauer *et al*., *FEBS Lett.* **470**, 189 (2000).
32. H. Berkefeld *et al*., *Science* **314**, 615 (2006).
33. J. Liu, J. Xia, K. H. Cho, D. E. Clapham, D. Ren, *J. Biol. Chem.* **282**, 18945 (2007).
34. M. S. Nadal *et al*., *Neuron* **37**, 449 (2003).
35. U. Schulte *et al*., *Neuron* **49**, 697 (2006).
36. Materials and methods are available as supporting material on *Science* Online.
37. C. Bokel, S. Dass, M. Wilsch-Brauninger, S. Roth, *Development* **133**, 459 (2006).
38. H. Hoshino *et al*., *Mol. Biol. Cell* **18**, 1143 (2007).
39. S. Roth, F. S. Neuman-Silberberg, G. Barcelo, T. Schupbach, *Cell* **81**, 967 (1995).
40. L. Chen, S. Bao, X. Qiao, R. F. Thompson, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12132 (1999).
41. K. Hashimoto *et al*., *J. Neurosci.* **19**, 6027 (1999).
42. B. Hille, *Ion Channels of Excitable Membranes* (Sinauer Associates, Sunderland, MA, ed. 3, 2001).
43. D. Turetsky, E. Garringer, D. K. Patneau, *J. Neurosci.* **25**, 7438 (2005).
44. C. Antz *et al*., *Nat. Struct. Biol.* **6**, 146 (1999).
45. D. S. Koh, J. R. Geiger, P. Jonas, B. Sakmann, *J. Physiol.* **485**, 383 (1995).
46. D. Colquhoun, P. Jonas, B. Sakmann, *J. Physiol.* **458**, 261 (1992).
47. C. P. Castro, D. Piscopo, T. Nakagawa, R. Derynck, *J. Cell Sci.* **120**, 2454 (2007).
48. Y. Sun *et al*., *Nature* **417**, 245 (2002).
49. B. Barbour, M. Häusser, *Trends Neurosci.* **20**, 377 (1997).
50. A. Rollenhagen *et al*., *J. Neurosci.* **27**, 10434 (2007).
51. We thank J. P. Adelman for insightful comments and critical reading of the manuscript and A. Haupt for help with bioinformatics. This work was supported by grants of the Deutsche Forschungsgemeinschaft to B.F. (SFB 746/TP16, SFB780/TPA3, EXC 294) and to N.K. (SFB780/TPB4).

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5919/1313/DC1
Materials and Methods
Figs. S1 to S6
Table S1
References

29 October 2008; accepted 16 January 2009
10.1126/science.1167852

Brightly Fluorescent Single-Walled Carbon Nanotubes via an Oxygen-Excluding Surfactant Organization

Sang-Yong Ju,[1] William P. Kopcha,[2] Fotios Papadimitrakopoulos[1,2]*

Attaining high photoluminescence quantum yields for single-walled carbon nanotubes (SWNTs) in order to broaden their optoelectronics and sensing applications has been a challenging task. Among various nonradiative pathways, sidewall chemisorption of oxygen provides a known defect for exciton quenching through nanotube hole doping. We found that an aliphatic (dodecyl) analog of flavin mononucleotide, FC12, leads to high dispersion of SWNTs, which tend to aggregate into bundles. Unlike other surfactants, the surface organization of FC12 is sufficiently tight to exclude oxygen from the SWNT surface, which led to quantum yields as high as 20%. Toluene-dispersed, FC12-wrapped nanotubes exhibited an absorption spectrum with ultrasharp peaks (widths of 12 to 25 milli–electron volts) devoid of the characteristic background absorption of most nanotube dispersions.

The ability to readily assign the (*n*,*m*) chirality of semiconducting single-walled carbon nanotubes (SWNTs) by means of photoluminescence excitation (PLE) mapping (*1*), together with their photostability (*2*), holds promise for applications in optoelectronics (*3*), biological imaging (*2*, *4*), and sensing (*4*). Although the optical properties of SWNTs are excitonic in nature (*5*), these structures exhibit low-fluorescence quantum yields. Possible causes include low-lying, nonradiative states (dark excitons) (*6*) or various defects that, as a result of the large exciton diffusion length (~90 nm) in SWNTs, contribute to substantial photoluminescence quenching (*7*, *8*). Oxygen in particular, in the presence of an acid or neutral environment (*9*), can quench photoluminescence through hole doping and subsequent nonradiative Auger recombination (*8*, *10*).

To make matters worse, nanotube bundling (*11*, *12*), along with chemical defects resulting from covalent functionalization (*13*) and nanotube inhomogeneities (*14*), can further decrease or completely quench nanotube luminescence. Individual SWNTs can have photoluminescence quantum yields as high as 8% (*15*), but solution-suspended SWNTs have shown much lower quantum yields [i.e., 1.5% for polyfluorene (PFO)–wrapped SWNTs (*16*), 1.1% for purified DNA-wrapped SWNTs (*11*), and less than 0.1% for surfactant-micellarized nanotubes (*16*, *17*)]. Most SWNT surfactants allow oxygen to interact and dope these nanotubes, and are sufficiently labile that they allow the nanotubes to reform bundles (*9*). Here, we show that a low-molecular-weight, organic-soluble analog of flavin mononucleotide, FC12, imparts considerable individualization in toluene and other aromatic solvents (i.e., *o*-xylene and benzene). In addition, the tight self-organization of FC12 around SWNTs leads to an effective exclusion of oxygen that affords quantum yields as high as 20%.

Flavin mononucleotide (FMN), a common redox cofactor related to vitamin B_2, was recently shown to self-organize around SWNTs through a helical conformation (*18*). Such helical wrapping (Fig. 1, A and B) originates from two sets of self-recognizing H-bonds that "stitch" the neighboring FMN moieties into a continuous helical ribbon (Fig. 1A), the concentric π-π interaction of the isoalloxazine ring with the underlying graphene sidewalls (Fig. 1B), and a soluble *d*-ribityl phosphate side group that imparts effective solubilization in aqueous media. In an effort to broaden flavin-based dispersion in organic solvents, we synthesized an isoalloxazine derivative with an aliphatic (dodecyl) side group, termed FC12. The synthetic route of FC12 involves two facile steps with an overall yield of ~35% (*19*). FC12 dispersions of CoMoCAT (Co-Mo bimetallic catalyst synthesized) SWNTs (*20*) were obtained by sonicating 1 mg of FC12, 1 mg of SWNTs, and 4 ml of various solvents for 4 hours at 300 W. The mixture was centrifuged for 20 min at 10,000*g*, which eliminated visible SWNT bundles in various solvents [i.e., benzene, toluene, *o*-xylene, ethylacetate, tetrahydrofuran (THF), pyridine, acetone, and *N*,*N*-dimethyl formamide (DMF)]. Table 1 summarizes the physical properties of these solvents as a function of dielectric constant (ε). SWNT photoluminescence was observed for only some of these solvents: benzene, toluene, *o*-xylene, ethylacetate, THF, and acetone (see below).

PLE maps for benzene, toluene, ethylacetate, and acetone show that the photoluminescence intensity (~315,000 counts) of FC12–(6,5)-SWNTs in toluene dispersion is 15 to 20 times

[1]Nanomaterials Optoelectronics Laboratory, Polymer Program, University of Connecticut, Storrs, CT 06269, USA. [2]Department of Chemistry, Institute of Materials Science, University of Connecticut, Storrs, CT 06269, USA.

*To whom correspondence should be addressed. E-mail: papadim@mail.ims.uconn.edu

that for other solvents (see maximum intensity in the bottom of each panel in Fig. 2; table S1 lists the relative photoluminescence intensities as a function of nanotube species, diameter, and chiral angle). In all PLE maps, all intensity maxima from the observed first (E_{11}^S) and second (E_{22}^S) electronic transitions of semiconducting SWNTs are similarly redshifted as in the case of FMN (*18*). The similar E_{11}^S and E_{22}^S redshift trend in organic solvents, as in the case of FMN in H_2O (i.e., 15 to 51 nm for E_{11}^S and 23 to 71 nm for E_{22}^S), is in agreement with the tight organization of flavin around SWNTs (*18*). Figure S2 plots the relative photoluminescence intensity of Fig. 2, B to D, as a function of nanotube diameter d_t. With the exception of (7,5)-SWNTs, which are more abundant in toluene, the relative photoluminescence intensities of the observed nanotube species do not change appreciably with d_t. However, the average photoluminescence full width at half-maximum (FWHM) values of all nanotube species in the solvents with low dielectric constant (i.e., 26 and 27 meV for benzene and toluene, respectively) are significantly less than those in the high-dielectric solvents (i.e., 39 and 45 meV for ethylacetate and acetone, respectively) (table S1). The larger FWHM and lower photoluminescence intensities indicate either large inhomogeneity in FC12 functionalization, or a greater degree of bundling than in the benzene and toluene dispersions, or both.

Tan *et al*. (*12*) reported that SWNT aggregation enhances exciton energy transfer (EET) between SWNTs within bundles, where excitation of large band gap tubes leads to emission from SWNTs with smaller band gaps. Such spectral features are easily discernible for species with high concentrations, as in the case of (6,5) and (7,5) for CoMoCAT SWNTs (*20*). The red arrows in Fig. 2, A to D, indicate such an EET feature between (6,5)- and (7,5)-SWNTs. When normalized to the PLE intensity of (7,5)-SWNTs, the (6,5) → (7,5) EET feature from the benzene and toluene is smaller than those seen in ethylacetate and acetone by a factor of 2 to 3 (fig. S3). This result provides an initial indication that SWNT individualization in toluene is one of the reasons for the higher photoluminescence intensity and narrower peaks.

Figure 3A illustrates the corresponding visible–near infrared (Vis-NIR) spectra of FC12-dispersed SWNTs in acetone, ethylacetate, and toluene. The toluene absorption spectrum shows sharper peaks that are shifted to higher energies with respect to those of ethylacetate and acetone, as well as a nearly flat baseline. Note that the three absorption curves in Fig. 3A are not offset with respect to each other; rather, they reside on a power-law background ($a\lambda^{-b}$, where λ is the wavelength and a and b are fitted parameters) (*19*, *21*). In addition, the E_{11}^S/E_{22}^S ratio of (6,5)-SWNTs, ~6, is the highest reported thus far, with DNA-wrapped (*22*), PFO-wrapped (*16*), and sodium dodecyl benzene sulfonate (SDBS)–dispersed (*23*) SWNTs showing E_{11}^S/E_{22}^S ratios of 3, 3.5, and 3.5, respectively for (6,5)- or (7,5)-SWNTs. Moreover, the (6,5) E_{11}^S/E_{22}^S ratio is ~6

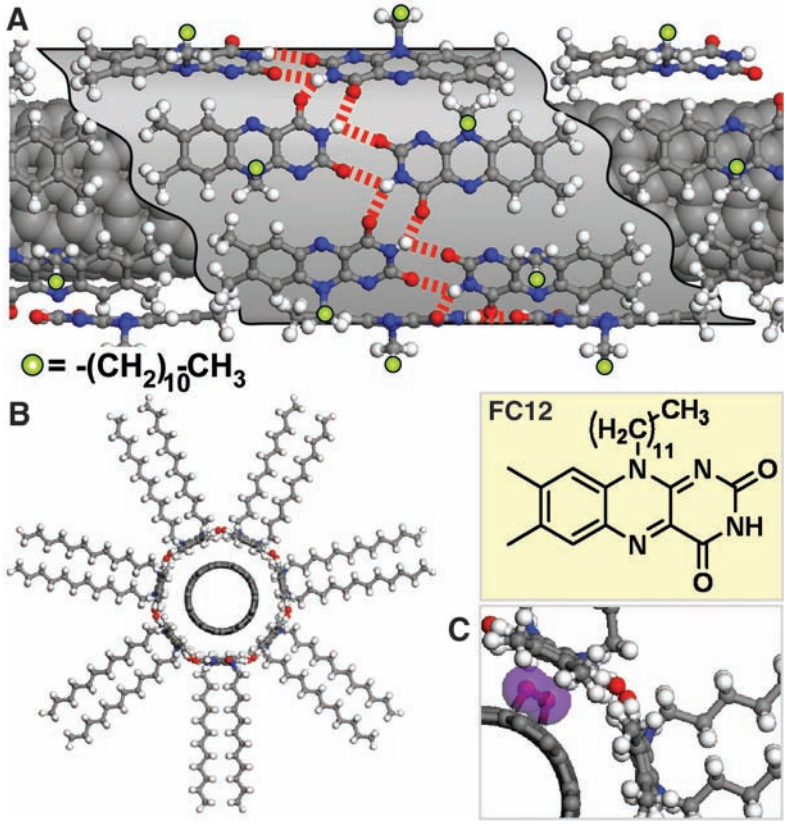


Fig. 1. (**A**) Top view of isoalloxazine moieties of FC12 wrapped in a 7_1 helical pattern. Red dotted lines and green spheres depict the intermolecular H-bonding between neighboring isoalloxazine moieties and the positions of the solubilizing side chains (undecyl), respectively. Color code for atoms: gray, carbon; white, hydrogen; blue, nitrogen; red, oxygen. (**B**) Side view illustrates the tight 7_1 helical wrapping of FC12 around (6,5)-SWNTs, which greatly interferes with the presence of a chemisorbed oxygen species (i.e., 1,4-endoperoxide and its van der Waals radii illustrated in purple) (**C**).

Table 1. Quantum yield values for (6,5)-SWNTs as a function of solvent dielectric constant (ε), H-bonding nature, and solubility values for FC12 and lumiflavin.

Solvent	Photoluminescence activity	Quantum yield of (6,5)-SWNTs (%)		ε	H-bonding capability*	Lumiflavin solubility (μg/ml)	FC12 solubility (mg/ml)	Max. E_{11}^S absorption
		"Sample"	"Individual"					
Benzene	Yes	5	10	2.3	P	4.4	0.31	0.01
Toluene	Yes	11	20	2.4	P	4.3	0.23	0.03
o-Xylene	Yes	8.7	16.9	2.5	P	5.9	0.25	0.02
Ethylacetate	Yes	0.024	0.4	6	M	14.1	0.51	0.075
THF	Yes	0.07	0.1	7	M	46.2	11.03	0.21
Pyridine	No	—	—	12.5	M-S	528.7	70.38	0.3
Acetone	Yes	0.08	0.2	21	M	53.9	1.81	0.14
DMF	No	—	—	39	M-S	316.7	22.87	0.7
D_2O†	Yes	0.08	0.8	79	S	—	—	0.19

*P, poor; M, moderate; M-S, moderate-strong (*32*). †FMN-dispersed SWNTs in D_2O.

only for toluene, dropping to ~3.5 and 4 for ethylacetate and acetone FC12 dispersions, respectively. Because the E_{11}^{S} transitions are more susceptible to doping-induced bleaching than are the E_{22}^{S} transitions (*9*), larger E_{11}^{S}/E_{22}^{S} ratios are indicative of a more pristine nature for the SWNTs.

A close-up of the toluene E_{11}^{S} spectral range, along with peak deconvolution based on the PLE-observed SWNTs, is shown in Fig. 3B. The background-subtracted absorption spectrum (circles) was deconvoluted with a multiplicity of Lorentzian peaks, where their maxima were allowed to vary 1 to 2 nm from the corresponding PLE-derived position (table S1) along with unrestricted FWHM (*19*). The NIR spectra were adequately described by a simple summation of the individual (*n*,*m*)-SWNTs observed in the corresponding PLE map of Fig. 2B. This one-to-one correlation between photoluminescence and absorption spectra indicates profound individualization for all observed nanotubes. With the exception of (7,6)-SWNTs, whose FWHM coincides with H_2O absorption, the FWHM values of all nanotubes range from 12 to 25 meV, with an average of 18 meV (22 meV for weighted average) (table S2). Comparison of these results with previously reported FWHM absorption values for DNA-wrapped nanotubes (18 to 24 meV) (*22*) and PFO-dispersed nanotubes (19 to 20 meV) (*16*) provides additional support for the idea that the helical wrapping of FC12 affords a highly uniform environment around these SWNTs (*24*).

To assess the deconvolution fidelity of Fig. 3B, we compared the percent abundance obtained by absorbance (Fig. 3B) and PLE intensities (table S2). The diameter (d_t), modality [mod($n-m$,3) = 1 or 2], family ($2n+m$ = constant), and (*n*,*m*) chirality of SWNTs all influence their absorptivity and corresponding photoluminescence efficiency (*25*, *26*). Our group (*21*), as well as others (*15*, *27*), have indicated that the theoretically calculated absorption [$I_{cal}^{abs}(n,m)$] and photoluminescence [$I_{cal}^{PL}(n,m)$] intensity factors calculated by Oyama *et al*. (*25*) provide a good framework for correlating (*n*,*m*) abundance between absorbance and photoluminescence results. Columns 4 and 5 of table S2 list the experimentally observed relative (*n*,*m*)-SWNT abundance as obtained by absorbance and PLE results, respectively. When these results are scaled by the Oyama-derived factors (columns 7 and 8), the average divergence between the absorbance- and photoluminescence-derived abundance results (0.12) becomes less than the divergence of the unscaled abundances (0.18).

What is the nature of the significantly greater background absorption for ethylacetate and

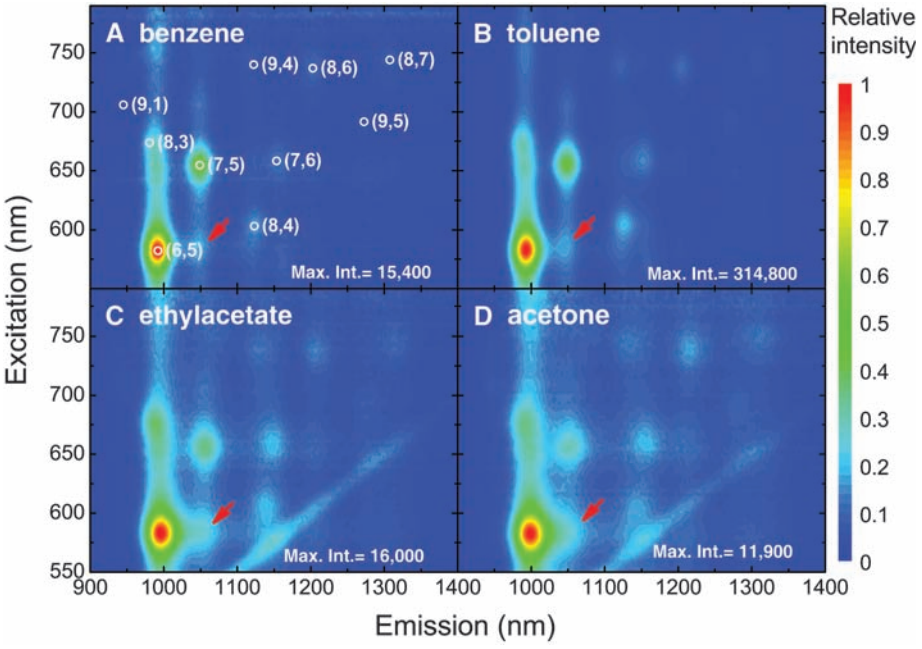


Fig. 2. Solvent-driven influences on the photoluminescence characteristics and brightness (maximum intensity counts) of FC12-dispersed CoMoCAT SWNTs. Numbers in parentheses and white circles indicate the (*n*,*m*) chiral indices and peak position, respectively, of various SWNT species in (**A**) benzene, (**B**) toluene, (**C**) ethylacetate, and (**D**) acetone. The toluene FC12-SWNT dispersion shows a maximum photoluminescence intensity (314,800 counts) for a significantly lesser amount of SWNTs according to Fig. 3A. Red arrows depict EET features from large to small band gap nanotubes.

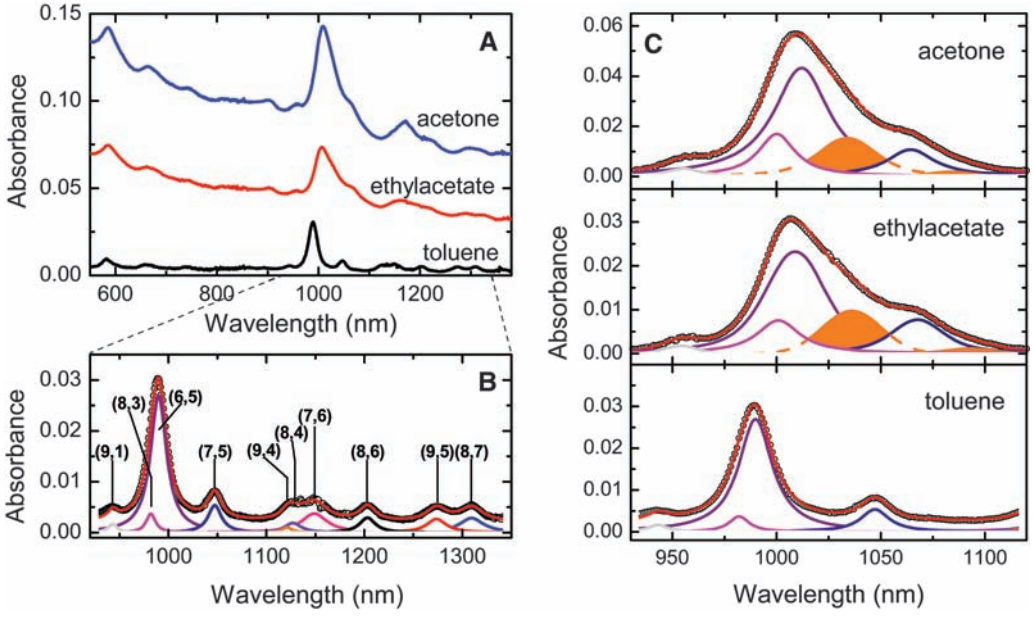


Fig. 3. Aggregation-induced spectral purity of FC12-dispersed SWNTs, as deduced by spectral deconvolution of their electronic absorption in various solvents. (**A**) Nonoffset Vis-NIR absorption spectra of FC12-dispersed SWNTs in acetone, ethylacetate, and toluene. (**B**) Deconvoluted NIR region of the toluene sample, based on the PLE-observed (*n*,*m*)-SWNT species of Fig. 2B. Circles and red curves denote the experimental and reconstructed spectra from the summation of the color-coded Lorentzian peaks, assigned to various (*n*,*m*) nanotubes. The former curves were offset upward by 0.002 absorbance units to facilitate visual comparison. (**C**) Absorption spectra from acetone, ethylacetate, and toluene FC12-SWNT dispersions from 930 to 1120 nm, along with their deconvoluted components depicting the color-coded (9,1)-, (8,3)-, (6,5)-, and (7,5)-SWNT species, in conjunction with two orange-colored Gaussians attributed to aggregate absorptions from the major (6,5) and (7,5) nanotube species.

acetone? Figure 3C illustrates the deconvoluted window from 930 to 1120 nm of Fig. 3A. This region contains the E_{11}^S absorptions of the most abundant (6,5)- and (7,5)-SWNTs in CoMoCAT sample (*20*), along with (9,1)- and (8,3)-SWNTs. After background subtraction (*19*), the pronounced features of (9,1)-, (6,5)-, and (7,5)-SWNTs allowed us to determined an average redshift of 21 meV for all three absorptions. When we redshifted and confined the peak position of all four nanotube absorptions of Fig. 3A and allowed peak height and FWHM optimization for Fig. 3C, the resulting fits (for several attempts) reproduced poorly the 1025 to 1050 nm and the 1075- to 1100-nm regions of ethylacetate and acetone sample, which are not occupied by any nanotubes (fig. S4). To account for this shortcoming, we introduced two Gaussian peaks in these positions (highlighted in orange), and re-deconvolved these spectra, to obtain an optimum fit. These two Gaussian peaks are shifted by an average of ~27 meV compared to the redshifted positions of (6,5) and (7,5)-SWNTs. Substitution of these two Gaussian peaks with Lorentzian shape produced a less optimum fit, indicative of an ensemble rather than an individual nanotube species.

Dynamic light scattering (DLS) was used to qualitatively assess the presence or absence of nanotube bundling in various solvents. In order to minimize laser absorption by both nanotube and the tail-end of FC12 that inadvertently interferes with size-dependent Brownian motion through thermal convection, we used a 633 nm laser line at 5 mW intensity power. Badaire *et al.* (*28*) reported that for laser powers below 150 mW, SWNT-related thermal convection effects become insignificant. The time-dependent autocorrelation function, $C(t)$, for the toluene and ethylacetate dispersed FC12-SWNT samples of Fig. 3A, are shown in fig. S5. After the rapid initial $C(t)$ decay, the toluene dispersion exhibits only a minor fluctuation between 20 and 75 μs. However, the ethylacetate dispersion exhibits a broad fluctuation spanning in excess of 400 μs, indicative of large bundling (*28*).

The presence of small and large SWNT bundles has been previously observed (*11*, *12*, *29*). Density gradient centrifugation has been used to fractionate, according to buoyancy, surfactant-suspended SWNTs in bundles of different diameter (*11*). Small bundles were shown to exhibit a progressively redshifted absorption (as high as 19 meV) and larger FWHM relative to individualized SWNTs. Similarly, their photoluminescence emissions were broader, were redshifted, and had considerably smaller photoluminescence quantum yields. These findings are in good agreement with the 21 meV in average redshift of the main E_{11}^S absorptions in Fig. 3C. This, along with the broader FWHM in both PLE and absorption spectra of ethylacetate and acetone FC12 dispersions, indicates that bundling is a major factor in luminescence quenching.

The profoundly intense PLE signal of toluene-dispersed FC12 dispersion, in particular that of the (6,5)-SWNTs, prompted us to determine their photoluminescence quantum yield using the method in (*11*). The absorbance (Fig. 4A) and emission (Fig. 4B) spectra of Styryl-13 were compared with the E_{22}^S absorption and E_{11}^S emission profiles of SDS- and FC12-dispersed CoMoCAT SWNTs. The 11% quantum yield of Styryl-13 at 3.3×10^{-7} M in methanol was used as a primary reference standard because of the close excitation and emission spectral overlap with (6,5)-SWNTs (*11*). Before determining the quantum yield of toluene-dispersed FC12-SWNTs, we used the aqueous-dispersed SDS-SWNTs as a secondary standard (*16*, *17*). Because the power-law background signal in both SDS and FC12 can influence significantly the nanotube quantum yield figures, we define as "sample" and "individual" quantum yields the values obtained when (6,5) absorption is taken from either zero absorbance units or the power-law fitted line, respectively. Figure S6 illustrates sample quantum yields of SDS-dispersed (6,5)-SWNTs as a function of their E_{22}^S absorption peak intensity. A value of 0.5% for the "sample" quantum yield was obtained at ultimate dilution while keeping the SDS concentration above the critical micelle concentration of ~1 weight percent (*16*, *17*). Progressively higher concentrations decrease the quantum yield down to 0.05% and below, in accordance with (*16*, *17*). The substantially higher photoluminescence intensity of FC12- versus SDS-dispersed SWNTs (Fig. 4B) yields a "sample" quantum yield of ~11% for (6,5)-SWNTs. When the scattering background is removed, the "individual" (6,5) quantum yield increases to ~20%. Table 1 lists the corresponding (6,5) quantum yields for all solvents in this study. *o*-Xylene and benzene demonstrated 17% and 10% "individual" quantum yield for (6,5)-SWNTs, respectively, with the remaining solvents showing progressively lower quantum yield values.

There may be several causes for the surprisingly large quantum yields for (6,5)-SWNTs. Near-armchair nanotubes such as (6,5)-SWNTs have been theorized to exhibit progressively higher quantum yields as their diameter decreases (*25*, *30*), although modality and family dependence also play an important role. Such increases, however, together with recent ~8% quantum yield determination for SDBS-suspended SWNTs (using single-nanotube photometry) (*15*), cannot explain the aforementioned 20% quantum yield value for (6,5)-SWNTs. Similarly, the possibility of selective enrichment of only (6,5)-SWNTs can easily be ruled out on the basis of one-to-one correlation between photoluminescence and NIR absorption, as well as the close resemblance of the relative photoluminescence abundance histograms for FC12- and SDS-dispersed nanotubes in fig. S2.

To pinpoint the nature of such a high quantum yield, we investigated the tight wrapping of FC12 onto nanotubes by time-dependent photoluminescence intensity traces as a function of O_2. Unlike SDS dispersion, which exhibits progressively lower photoluminescence as a function of irradiation (*9*), the photoluminescence intensity of toluene FC12 dispersion remained constant (Fig. 4C). Although the reversible acid-induced p-doping of SWNTs in the presence of oxygen is well documented (*7*–*9*), less attention has been exerted on SWNT interactions with oxygen alone (*8*). Density functional theory calculations by Dukovic *et al.* (*8*) indicate that singlet O_2 chemisorbs on a (5,0)-SWNT and adopts a 1,4-endoperoxide structure (Fig. 1C). The calculated activation energy for 1,4-endoperoxide desorp-

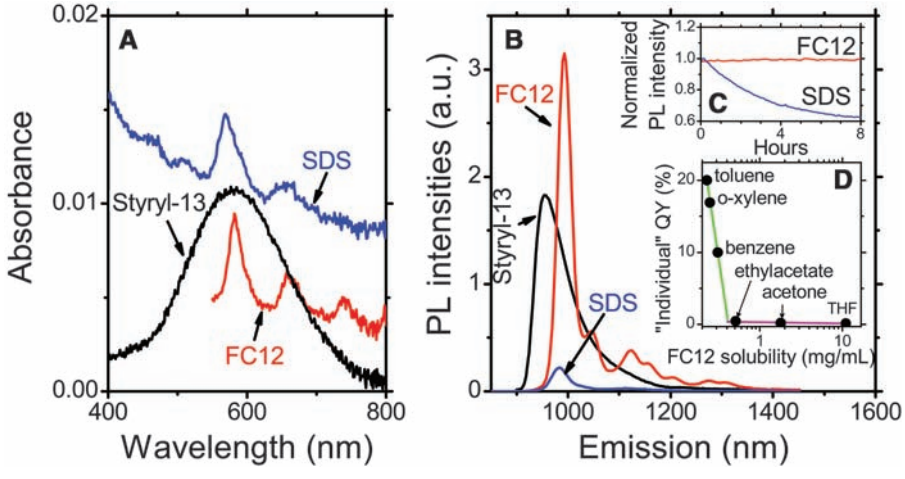


Fig. 4. Quantum yield determination and photo-oxidation stability of SDS- and FC12-dispersed SWNTs. Quantum yield values for (6,5)-SWNTs were determined by comparing their electronic absorption (**A**) and photoluminescence emission (**B**) relative to a Styryl-13 reference standard (*11*). (**C**) Normalized photoluminescence intensity behavior of (6,5)-SWNTs as a function of irradiation time in their respective D_2O and toluene media. (**D**) "Individual" quantum yield dependence of (6,5)-SWNTs versus FC12 solubility in various solvents.

tion (~1 eV) (*8*) is about half of the binding energy of a single isoalloxazine moiety (~2.2 eV) (*31*), which suggests that chemisorbed O_2 should be displaced by FC12 organization around the nanotubes. As illustrated in Fig. 1C, the seamless isoalloxazine wrapping is severely prohibited in the presence of a chemisorbed 1,4-endoperoxide because of van der Waals repulsions. In contrast to FC12, the loose organization of other surfactants (such as SDS and SDBS) around SWNTs permits the inclusion of these 1,4-endoperoxide defects. We note that PFO-wrapped SWNTs, in which the surfactant is more organized, have a greater nanotube "sample" photoluminescence quantum yield (1.5%) (*16*) versus 0.1 to 0.5% for SDS-dispersed SWNTs (*16*, *17*).

The origin of the large solvent variation in the observed photoluminescence quantum yield may originate in the relative solubility differences of the two FC12 submoieties (i.e., isoalloxazine ring and dodecyl side chain). We procured 10-methyl isoalloxazine (lumiflavin), the closest analog to the isoalloxazine ring, and investigated its solubility characteristics versus those of FC12. Table 1 lists the solubilities of lumiflavin and FC12 as a function of dielectric constant (ε) of various solvents. Both lumiflavin and FC12 exhibit significantly reduced solubilities in nonpolar solvents (i.e., benzene, toluene, and *o*-xylene), with toluene the lowest. As solvent polarity increases (i.e., ethylacetate, THF, and acetone), their solubilities increase accordingly. Pyridine and DMF exhibit the highest solubility values for both lumiflavin and FC12, although SWNT photoluminescence activity is absent. Such behavior would originate from the moderate to strong H-bonding ability of pyridine and DMF, which would dissociate the H-bonded FC12 ribbon responsible for nanotube dispersion.

The "individual" quantum yield of (6,5)-SWNTs is shown in Fig. 4D as a function of FC12 solubility in each solvent (Table 1). High quantum yields (green line) were seen for low-polarity solvents, and low quantum yields (magenta line) were seen for medium-polarity solvents. For both regimes, the increase in FC12 solubility was followed by a decrease in quantum yield. Because H-bonding is responsible for both helix stability and FC12 (or lumiflavin) dissolution, increasing solvent polarity is expected to increase the helix dissociation constant and render FC12-wrapped nanotubes more prone to bundling and less able to desorb 1,4-endoperoxide defects. The proportionality of the helix dissociation constant to the FC12 solubility would account for the linear dependence of 1,4-endoperoxide defect removal along the one-dimensional SWNT structure. The uniform and defect-free environment offered by the flavin organization, which is needed as a result of the large exciton diffusion length (~90 nm) (*7*) in SWNTs, opens an array of new frontiers in SWNT photophysics. Moreover, the flavin organization is also compliant with the hierarchical assembly of nanotubes for device manipulation.

References and Notes

1. S. M. Bachilo *et al.*, *Science* **298**, 2361 (2002); published online 29 November 2002 (10.1126/science.1078727).
2. D. A. Heller, S. Baik, T. E. Eurell, M. J. Strano, *Adv. Mater.* **17**, 2793 (2005).
3. J. Chen *et al.*, *Science* **310**, 1171 (2005).
4. P. W. Barone, S. Baik, D. A. Heller, M. S. Strano, *Nat. Mater.* **4**, 86 (2005).
5. F. Wang, G. Dukovic, L. E. Brus, T. F. Heinz, *Science* **308**, 838 (2005).
6. V. Perebeinos, J. Tersoff, P. Avouris, *Nano Lett.* **5**, 2495 (2005).
7. L. Cognet *et al.*, *Science* **316**, 1465 (2007).
8. G. Dukovic *et al.*, *J. Am. Chem. Soc.* **126**, 15269 (2004).
9. M. S. Strano *et al.*, *J. Phys. Chem. B* **107**, 6979 (2003).
10. F. Wang, G. Dukovic, E. Knoesel, L. E. Brus, T. F. Heinz, *Phys. Rev. B* **70**, 241403 (2004).
11. J. Crochet, M. Clemens, T. Hertel, *J. Am. Chem. Soc.* **129**, 8058 (2007).
12. P. H. Tan *et al.*, *Phys. Rev. Lett.* **99**, 137402 (2007).
13. M. S. Strano *et al.*, *Science* **301**, 1519 (2003).
14. Y. Miyamoto, A. Rubio, S. Berber, M. Yoon, D. Tomanek, *Phys. Rev. B* **69**, 121413 (2004).
15. D. A. Tsyboulski, J. D. R. Rocha, S. M. Bachilo, L. Cognet, R. B. Weisman, *Nano Lett.* **7**, 3080 (2007).
16. A. Nish, J.-Y. Hwang, J. Doig, R. J. Nicholas, *Nat. Nanotechnol.* **2**, 640 (2007).
17. M. J. O'Connell *et al.*, *Science* **297**, 593 (2002).
18. S.-Y. Ju, J. Doll, I. Sharma, F. Papadimitrakopoulos, *Nat. Nanotechnol.* **3**, 356 (2008).
19. See supporting material on *Science* Online.
20. S. M. Bachilo *et al.*, *J. Am. Chem. Soc.* **125**, 11186 (2003).
21. Z. Luo, L. D. Pfefferle, G. L. Haller, F. Papadimitrakopoulos, *J. Am. Chem. Soc.* **128**, 15511 (2006).
22. M. Zheng, E. D. Semke, *J. Am. Chem. Soc.* **129**, 6084 (2007).
23. Y. Tan, D. E. Resasco, *J. Phys. Chem. B* **109**, 14454 (2005).
24. O. Kiowski *et al.*, *Phys. Rev. B* **75**, 075421 (2007).
25. Y. Oyama *et al.*, *Carbon* **44**, 873 (2006).
26. C. Fantini *et al.*, *Phys. Rev. Lett.* **93**, 147406 (2004).
27. B. Wang *et al.*, *J. Am. Chem. Soc.* **129**, 9014 (2007).
28. S. Badaire, P. Poulin, M. Maugey, C. Zakri, *Langmuir* **20**, 10367 (2004).
29. A. Hagen, T. Hertel, *Nano Lett.* **3**, 383 (2003).
30. S. Reich, C. Thomsen, J. Robertson, *Phys. Rev. Lett.* **95**, 077402 (2005).
31. C. S. Lin, R. Q. Zhang, T. A. Niehaus, T. Frauenheim, *J. Phys. Chem. C* **111**, 4069 (2007).
32. D. W. V. Krevelen, *Properties of Polymers* (Elsevier, New York, ed. 3, 1990).

33. We thank S. Vaddiraju, J. Doll, D. Abanulo, A. Nish, and R. J. Nicholas for useful discussions. Supported mainly by Air Force Office of Scientific Research grant FA9550-06-1-0030, and in part by NSF grants NIRT DMI-0422724 and CBET-0828771/0828824 and NIH grant ES013557.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5919/1319/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 and S2
References

23 September 2008; accepted 21 January 2009
10.1126/science.1166265

Probing the Angular Momentum Character of the Valence Orbitals of Free Sodium Nanoclusters

C. Bartels, C. Hock, J. Huwer,* R. Kuhnen, J. Schwöbel,† B. von Issendorff‡

Although many properties of polyatomic metal clusters have been rationalized by an electron shell model resembling that used for free atoms, it remained unclear how reliable this analogy is with respect to the angular momentum eigenstate character of the electronic wave functions. We studied free size-selected negatively charged clusters of sodium atoms (Na_n^-) of approximately spherical shape (n = 19, 40, 55, 58, 147) by angle-resolved photoelectron spectroscopy over a broad range of photon energies (1.5 to 5 electron volts). Highly anisotropic, state- and energy-dependent angular distributions emerged for all sizes. Well-defined classes of energy dependence related to the approximate angular momenta of the bound-state orbitals indicate that the overall character of the valence electron wave functions is not appreciably influenced by the interaction with the ion background. The measured distributions nevertheless deviate strongly from the predictions of single-electron models, hinting at a distinct role of correlated multielectron effects in the photoemission process.

Under the free-electron model, a cluster of metal atoms can be seen as a realization of the textbook case of a spherical box potential filled with a well-defined number of electrons, and therefore as an ideal model system to study the structure and dynamics of a finite-size Fermi system. In such a system, the electrons occupy angular momentum eigenstates, which leads to a highly discretized density of states—the electron shell structure (*1*, *2*). By analogy with atomic orbital filling, the clusters can thus be construed as "artificial atoms" (*3*). Sodium is the best representative of a free-electron metal (*4*), and indeed it was for sodium clusters that an influence of this shell structure was initially observed (*5*). This result inspired a wealth of studies on simple metal clusters, and shell effects have been found in many cluster properties, for example, in the size dependences of binding energies, ionization potentials, and absorption spectra (*1*). Nevertheless, photoelectron spectroscopy, which in principle allows a direct imaging of

the electron density of states, revealed a somewhat more complicated situation than in the simplest models (*6*–*9*). The interaction with the ion background lifts the degeneracy of the angular momentum states and broadens the shells into bandlike groups of levels with considerable overall widths. Strictly speaking, in the presence of an ion background without spherical symmetry, the angular momentum l is no longer a good quantum number. It is therefore an important question how much simple model character remains in the true electronic states of a real cluster. To date, no experiment has directly targeted the angular momentum character of the valence orbitals of simple metal clusters. Angle-resolved photoelectron spectroscopy in principle offers a way to do so, as we will briefly explain.

The photoelectron angular distribution (PAD) of a sample of randomly oriented molecules for single-photon absorption is given by $1 + \beta\,(3/2\,\cos^2\theta - 1/2)$, where in the case of linearly polarized light, θ is the electron emission angle with respect to the electric field vector (*10*). The distribution is thus fully described by the anisotropy parameter β, which assumes values between –1 and +2, corresponding to perfectly perpendicular and perfectly parallel distributions, respectively (Fig. 1). For spherical systems, bound and continuum states have well-defined angular momenta l, and in the case of linearly polarized light, the transition selection rules $\Delta l = \pm 1$ and $\Delta m = 0$ hold.

The emission of an electron from an s state ($l = 0$) therefore always results in an outgoing wave with $l' = 1$ and $m = 0$, corresponding to $\beta = +2$. All other cases give rise to two partial waves with $l' = l \pm 1$, each of which is modestly peaked in the direction of the light polarization (β converges to 1/2 for large l). The interference of the two waves, however, can lead to any value of β between −1 and +2. In particular, negative values of β can only be produced by such an interference effect. As Bethe has shown for one-electron systems (*11*) and Cooper and Zare for many-electron systems (*12*), β can be analytically calculated from the angular momentum l of the bound state and the amplitudes and phases of the two outgoing partial waves (see supporting text). PADs therefore carry information about the angular momentum character of the orbitals from which the electrons emerge.

In our experiment (*13*), negatively charged clusters of sodium atoms (Na_n^-) were produced in a gas-aggregation cluster source. The cluster ions were stored for about 1 s in an ion trap cooled to 6 K, where they were thermalized by collisions with helium buffer gas. Cooling is essential for this type of experiment; it strongly enhances the quality of the photoelectron spectra as it brings the clusters close to their vibrational ground state. After extraction from the trap, the clusters were size-selected in a double-reflectron time-of-flight mass spectrometer and irradiated by linearly polarized pulses from a dye laser (pulse width ~ 10 ns, peak intensity < 10^5 W/cm^2). The photon energies mostly used were 2.48 and 4.02 eV; for selected cluster sizes (n = 19, 55, and 147), the photon energy was additionally varied between 1.5 and 5.0 eV in small steps. The photoelectrons were detected in an optimized velocity-map imaging spectrometer (*14*–*16*). The measured projections were transformed into angle-resolved photoelectron spectra by a slightly modified version of the pbasex algorithm (*17*); β parameters for individual transitions were extracted from these spectra by least-squares fits to the observed peaks (*13*).

Examples are shown in Fig. 2, where angle-integrated spectra (upper panels) and angle-resolved spectra (lower panels) are presented for clusters comprising 40, 58, and 147 atoms. These clusters are close to either electronic or geometrical (atomic layer) shell closings and have a well-defined electron shell structure (*18*). Each of them can furthermore be assumed to adopt predominantly a single geometrical structure, because the presence of several structural isomers in the experiment usually leads to a blurring of the photoelectron spectra (*18*).

Neutral Na_{40} is an electronically closed shell cluster; its anion has the formal electronic configuration $1s^2\ 1p^6\ 1d^{10}\ 2s^2\ 1f^{14}\ 2p^6\ 1g^1$. The uppermost three of these shells can be identified in the angle-integrated photoelectron spectrum in Fig. 2A. The electron shells are far from being narrow peaks; owing to the interaction with the ionic background, they are broadened into bands with a width comparable to their separation energy. Nevertheless, the corresponding angle-resolved spectra are surprisingly simple: Whereas the angular distribution is essentially isotropic for electrons detached from the 1g shell, it is parallel for 2p electrons and perpendicular for 1f electrons. So, here all substates of a given shell exhibit very similar behavior, whereas different shells exhibit different PADs.

Neutral Na_{58} also is an electronically closed shell cluster; the anion formally has a singly occupied 2d orbital above the completely filled 1g shell. In Na_{55}^-, which (except for a small Jahn-Teller distortion) has perfect icosahedral symmetry (*18*), this 1g band is split into two well-defined peaks (Fig. 3B), whereas in the lower-symmetry Na_{58}^- it is broadened into a band with a width of almost 400 meV (Fig. 2B). Nevertheless, the angular distributions are very similar for all substates within the 1g band, and substantially different from those of the 2p and the 2d bands. Again, the PADs of electrons detached from different shells are surprisingly different.

Na_{147}^- has icosahedral geometry and therefore a highly structured electronic density of states (*18*). For such a large cluster size, each major band consists of several approximate angular momentum states. This characteristic is directly reflected in the angular distributions: The 2g shell shows a uniform behavior, but the 1i/2f/3p and the 1h/2d/3s bands consist of several components with clearly different PADs. The most important point, however, is that one can observe such strongly structured spectra at all for a cluster of this size. In particular, the occurrence of negative values of the β parameter ($\beta = -0.68 \pm 0.02$ for the 2g shell at 530 nm, and $\beta = -0.87 \pm 0.11$ for the 1i shell at 348 nm), which can only be produced by destructive interference of the outgoing partial waves in the direction of the light polarization, demonstrates that the emission pro-

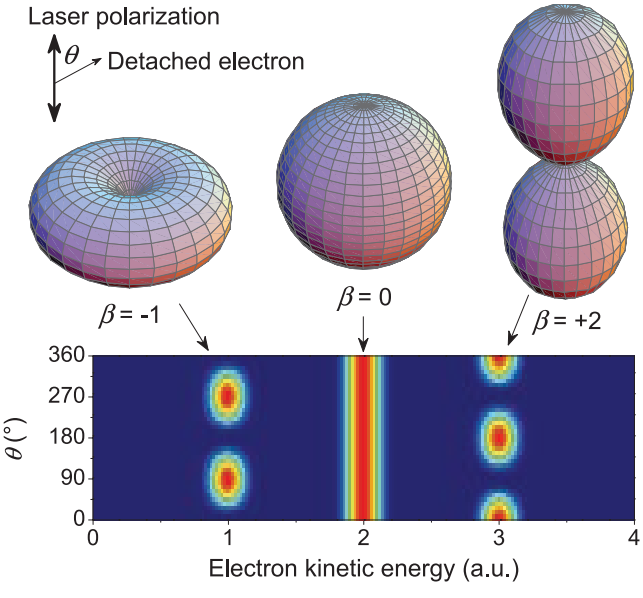


Fig. 1. Schematic angle-resolved spectrum of three hypothetical electronic states exhibiting perpendicular, isotropic, and parallel distributions (β parameters –1, 0, and +2, respectively). The angular distributions are visualized as three-dimensional polar plots above.

Department of Physics and Freiburg Materials Research Center, University of Freiburg, Stefan-Meier-Straße 21, 79104 Freiburg, Germany.

*Present address: ICFO–Institut de Ciències Fotòniques, Mediterranean Technology Park, 08860 Castelldefels (Barcelona), Spain.

†Present address: Institut für Angewandte Physik und Zentrum für Mikrostrukturforschung, Jungiusstraße 9a, 20355 Hamburg, Germany.

‡To whom correspondence should be addressed. E-mail: bernd.von.issendorff@uni-freiburg.de

cess is highly coherent. Apparently, potentially coherence-destroying processes like electron-electron or electron-phonon scattering do not yet play a major role in this size region.

Because PADs generally have a state-specific photon energy dependence (*19*), two different states may produce identical distributions at a given photon energy. A full characterization of the bound-state orbitals can therefore only be obtained by studying the PADs as a function of photon energy.

We have performed such experiments for several cluster sizes. Here we present the results for the electronically closed shell Na_{19}^-, which has an almost spherical but low-symmetry structure (*20*), and for Na_{55}^-, which has icosahedral symmetry but is two electrons short of an electronic shell closing (*18*).

Examples of spectra measured at different wavelengths are shown in Fig. 3. As expected, the angular distributions show strong variations with photon energy. The only exception is the uppermost level of Na_{19}^-, which is a 2s state. Electron emission from this orbital should always result in a p_z wave ($l' = 1$, $m = 0$), and indeed the corresponding angular distributions are parallel for all photon energies investigated. In contrast, a pronounced evolution of the angular distributions with photon energy is observed for electrons from the 1d and 1p states of Na_{19}^-. For Na_{55}^-, the angular distributions of all of the states quite drastically change with energy. Moreover, the angular distributions for the different substates of an angular momentum eigenstate are not always alike. At 550 nm, for example, the lower-energy component of the 1g state has an isotropic distribution, whereas the higher-energy component is still perpendicular. This difference indicates that the correlation between the initial angular momentum and the PAD is more complex than suggested above.

The evolution of the PADs is most clearly seen by plotting the anisotropy parameter β as evaluated for the different states as a function of photon energy *hν* (Fig. 4, A and B). The resulting

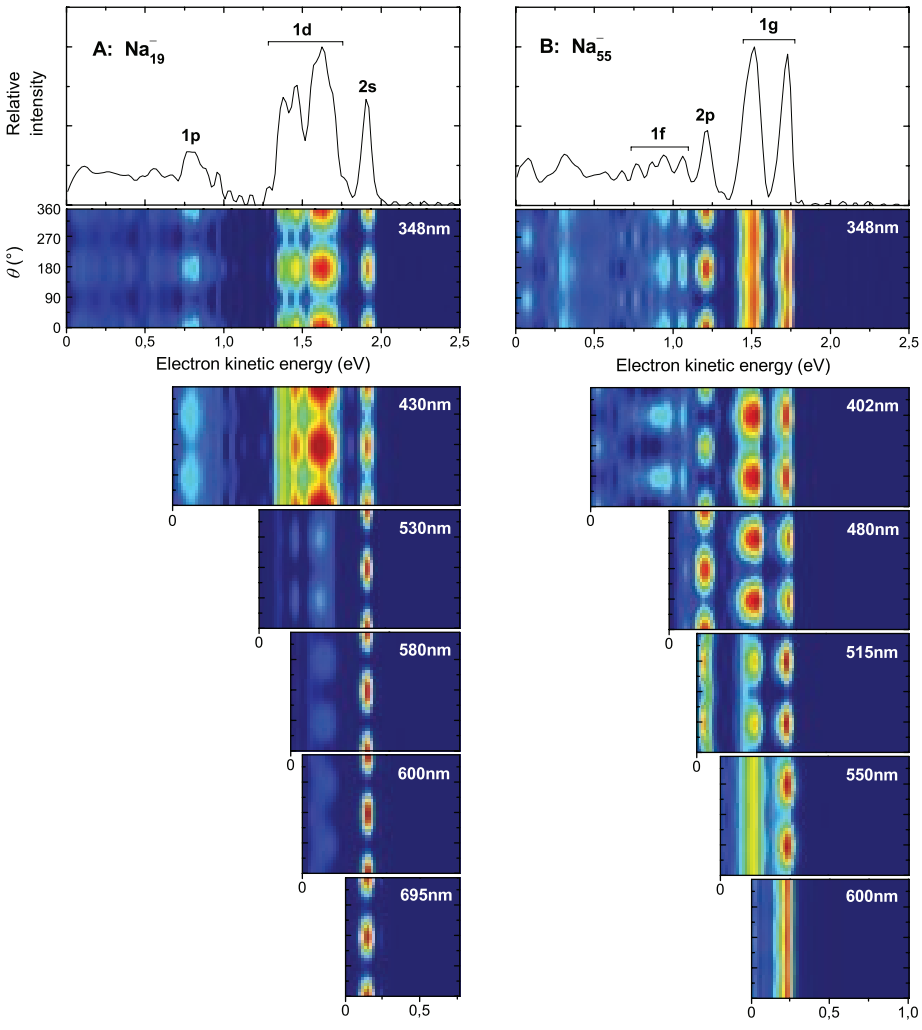


Fig. 3. (**A** and **B**) The evolution of angle-resolved photoelectron spectra for the cluster sizes 19 and 55. (A) The distributions for the 2s electrons of Na_{19}^- are always parallel, whereas the distribution of the 1d electrons evolves from isotropic at threshold (600 nm) to perpendicular (580, 530 nm) and then parallel (430, 348 nm). (B) The angular distributions of the 1g electrons of Na_{55}^- are isotropic at threshold, then perpendicular in the wavelength range of 550 to 402 nm, then isotropic again. Different peaks corresponding to the same approximate angular momentum can exhibit clearly different angular distributions, as seen for the 1g electrons from Na_{55}^- at 550 nm.

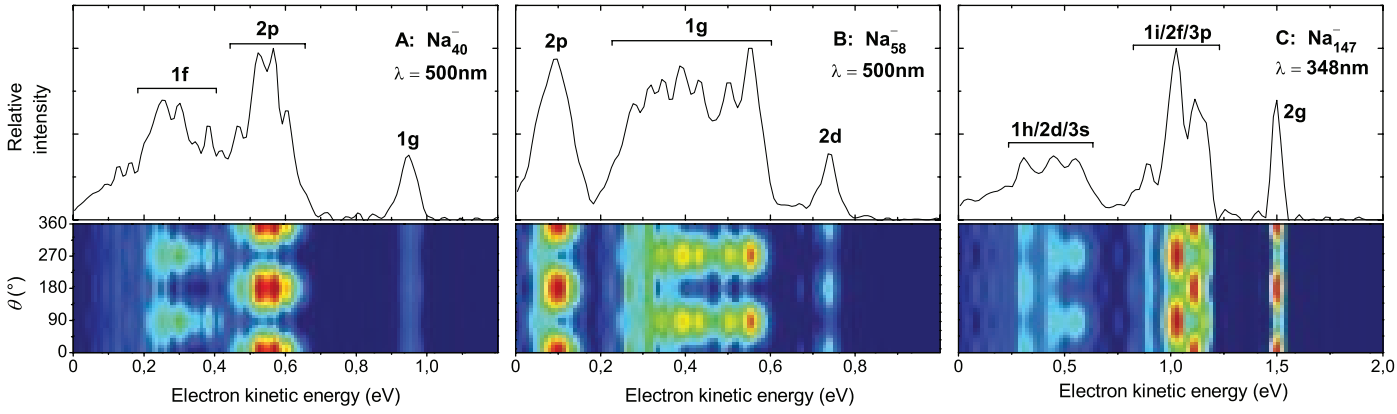


Fig. 2. Typical spectra for the cluster sizes 40 (**A**), 58 (**B**), and 147 (**C**). The lower panels show the energy- and angle-resolved spectra, with the emission angle θ measured with respect to the laser polarization. The upper panels show the corresponding angle-integrated spectra, together with the assignment of shell-model quantum numbers.

curves exhibit a smooth but notable energy dependence and cover essentially all of the accessible range between –1 and +2.

Closer inspection shows that all substates of a given approximate angular momentum eigenstate have a very similar photon-energy dependence, but with a certain shift in energy (for example, the substates of the 1d shell of Na_{19}^- plotted in green, or the substates of 1f of Na_{55}^- plotted in blue). The reason for this shift is that the photon energy is not really the relevant parameter in this case; it is the kinetic energy of the detached electron that determines the wave function of the continuum state and thereby the radial transition matrix elements and phase shifts. Indeed, when plotting the β parameters as a function of the photoelectron kinetic energy, practically identical behavior is observed for all substates of a given shell (Fig. 4, C and D). This finding demonstrates that the interaction with the ionic background evidently modifies only the energies of the different states, not the overall character of the wave functions. Each single-particle state seems to retain a dominant contribution from just one angular momentum eigenstate, which can be seen as an ex post facto justification of the use of the shell model.

Nevertheless, some deviations from such simple behavior are apparent. The 2s state of Na_{19}^-, which in principle should have β = +2 for all energies, exhibits a broad dip in its β curve at about 2.8 eV. Because this value is in the expected energy range of the collective plasmon excitation of Na_{19}^- (*21*, *22*), one could speculate that here the competition between plasmon excitation and electron emission alters the angular distribution, but calculations are necessary before definite statements can be made.

In general, the angular distributions of the different shells seem to possess a characteristic and rather simple energy dependence. This observation suggests that it might be possible to describe them within a single-electron central-potential model. We have performed such calculations using a "wine bottle" potential as obtained from self-consistent jellium theory (*23*), with a modified long-range part that incorporates the interaction of the outgoing electron with the neutral but polarizable cluster (*24*, *25*). Numerically solving the Schrödinger equation for this potential yielded the bound and continuum wave functions, which allowed the calculation of β using the Bethe-Cooper-Zare formula (figs. S2 and S3 and supporting text). To check the sensitivity of the calculations to the precise shape of the potential, we varied the potential depth by ± 200 meV around the value that gave the best agreement between the calculated and measured electron binding energies.

As seen in Fig. 4, E and F, the resulting curves show strong variations of the anisotropy parameter within a few electron volts above the detachment threshold and, except for the 1g state of Na_{55}^-, qualitatively resemble the evolution of the experimental values; the quantitative agreement, however, is poor. Changing the potential depth introduces some changes in the resulting curves but leaves the overall trend unaffected. Further tests were performed with square-well and Woods-Saxon potentials (fig. S4). For the Woods-Saxon potential, the parameters for radius, depth, and surface thickness were systematically varied over a broad range of values, but no match with the experimental data was obtained even for unphysical choices of parameters (fig. S5).

It is unlikely that the neglect of the ionic background structure is responsible for this failure of the model to reproduce the measurements. The substates of the 1g band in Na_{58}^- exhibit a very similar behavior. Actually, all substates of the 1g band in all sizes from Na_{53}^- to Na_{57}^- have very similar PADs; the same can be observed for other bands and in other size ranges. This demonstrates that details of the wave functions that are related to the ionic background cannnot have a strong influence on the angular distributions and leads us to the conclusion that it might be the single-active-electron assumption that is wrong.

The calculation of the initial bound and the final continuum wave functions with the same effective single-particle potential ignores any correlated interaction of the emitted electron with the other electrons. Though this is expected to be a good approximation for electrons with high kinetic energies, it is questionable for relatively slow electrons that spend a substantial time inside and close to the residual system (*26*). For the experiments discussed here, the electron kinetic energy is on the order of 1 eV, meaning that the electron leaves the cluster within a time of about 1 fs. The time scale for the response of the electronic system is given by the inverse bulk plasmon frequency (*27*), which is also ~1 fs. Consequently, the emitted electron interacts with neither a frozen electronic system nor a fully adiabatically responding one, but experiences an intermediate situation. The observed angular distributions therefore are probably the result of

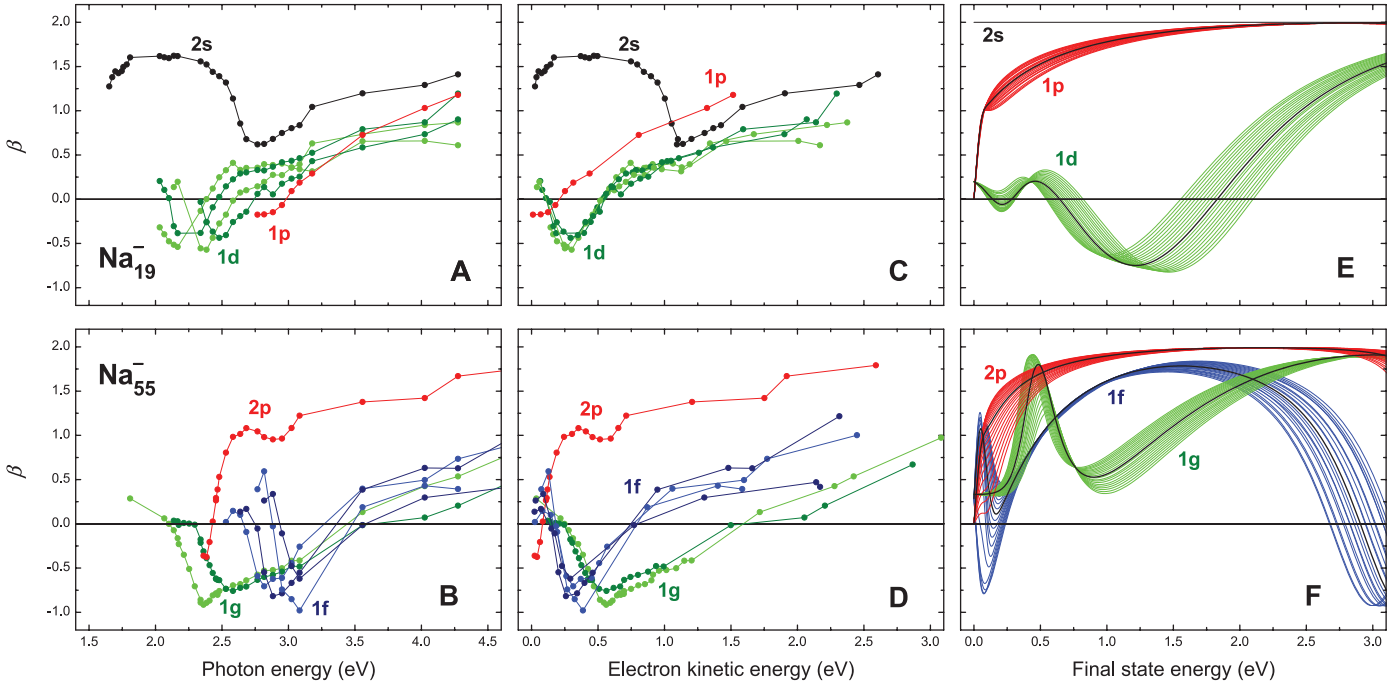


Fig. 4. The anisotropy parameters β as a function of photon energy (**A** and **B**) and as a function of electron kinetic energy (**C** and **D**) compared to the curves calculated by the model described in the text (**E** and **F**). The calculated curves correspond to different potential depths, varied in the range of ±200 meV around the optimum value.

correlated multiparticle dynamics and could be used, in combination with theory, to characterize these interactions in detail.

References and Notes

1. W. A. de Heer, *Rev. Mod. Phys.* **65**, 611 (1993).
2. M. Brack, *Rev. Mod. Phys.* **65**, 677 (1993).
3. P. L. McEuen, *Science* **278**, 1729 (1997).
4. N. W. Ashcroft, N. D. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976).
5. W. D. Knight *et al*., *Phys. Rev. Lett.* **52**, 2141 (1984).
6. C. L. Pettiette *et al*., *J. Chem. Phys.* **88**, 5377 (1988).
7. J. G. Eaton, L. H. Kidder, H. W. Sarkas, K. M. McHugh, K. H. Bowen, in *Nuclear Physics Concepts in the Study of Atomic Cluster Physics*, R. Schmidt, H. O. Lutz, R. Dreizler, Eds. (Springer, Berlin, 1992), vol. 404 of *Lecture Notes in Physics*, pp. 291–304.
8. H. Handschuh, C.-Y. Cha, P. S. Bechthold, G. Ganteför, W. Eberhardt, *J. Chem. Phys.* **102**, 6406 (1995).
9. G. Wrigge, M. Astruc Hoffmann, B. v. Issendorff. *Phys. Rev. A* **65**, 063201 (2002).
10. C. N. Yang, *Phys. Rev.* **74**, 764 (1948).
11. H. Bethe, in *Handbuch der Physik*, H. Geiger, K. Scheel, Eds. (Springer, Berlin, 1933), chap. 3, pp. 475–484.
12. J. Cooper, R. N. Zare, *J. Chem. Phys.* **48**, 942 (1968).
13. Materials and methods are available as supporting material on *Science* Online.
14. H. Helm, N. Bjerre, M. J. Dyer, D. L. Huestis, M. Saeed, *Phys. Rev. Lett.* **70**, 3221 (1993).
15. A. J. R. Heck, D. W. Chandler, *Annu. Rev. Phys. Chem.* **46**, 335 (1995).
16. A. T. J. B. Eppink, D. H. Parker, *Rev. Sci. Instrum.* **68**, 3477 (1997).
17. G. A. Garcia, L. Nahon, I. Powis, *Rev. Sci. Instrum.* **75**, 4989 (2004).
18. O. Kostko, B. Huber, M. Moseler, B. von Issendorff, *Phys. Rev. Lett.* **98**, 043401 (2007).
19. J. Berkowitz, *Photoabsorption, Photoionization, and Photoelectron Spectroscopy* (Academic Press, New York, 1979).
20. M. Moseler *et al*., *Phys. Rev. B* **68**, 165413 (2003).
21. M. Schmidt, H. Haberland, *Eur. Phys. J. D* **6**, 109 (1999).
22. C. Yannouleas, *Chem. Phys. Lett.* **193**, 587 (1992).
23. W. Ekardt, *Phys. Rev. B* **29**, 1558 (1984).
24. W. D. Hall, J. C. Zorn, *Phys. Rev. A* **10**, 1141 (1974).
25. W. D. Knight, K. Clemenger, W. A. de Heer, W. A. Saunders, *Phys. Rev. B* **31**, 2539 (1985).
26. J. W. Cooper, *Phys. Rev.* **128**, 681 (1962).
27. R. Huber *et al*., *Nature* **414**, 286 (2001).

28. This work was supported by the Deutsche Forschungsgemeinschaft.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5919/1323/DC1
Materials and Methods
SOM Text
Figs. S1 to S7
References

4 November 2008; accepted 16 January 2009
10.1126/science.1168080

Implications of a VLBI Distance to the Double Pulsar J0737-3039A/B

A. T. Deller,[1]* M. Bailes,[1] S. J. Tingay[2]

The double pulsar J0737-3039A/B is a unique system with which to test gravitational theories in the strong-field regime. However, the accuracy of such tests will be limited by knowledge of the distance and relative motion of the system. Here we present very long baseline interferometry (VLBI) observations which reveal that the distance to PSR J0737-3039A/B is 1150^{+220}_{-160} parsecs, more than double previous estimates, and confirm its low transverse velocity (~9 kilometers per second). Combined with a decade of pulsar timing, these results will allow tests of gravitational radiation emission theories at the 0.01% uncertainty level, putting stringent constraints on theories that predict dipolar gravitational radiation. They also allow insight into the system's formation and the source of its high-energy emission.

The double pulsar system PSR J0737-3039A/B (*1*, *2*) is one of eight known double neutron star (DNS) systems, and the only system in which both neutron stars are visible as pulsars. The "A" pulsar, which has been spun-up by accretion ("recycled") to a period P = 22.7 ms, was discovered first; the nonrecycled "B" pulsar (P = 2.77 s) was found during follow-up timing observations. PSR J0737-3039A/B is the most relativistic known DNS system. It has an orbital period of 2.5 hours and a coalescence time (due to orbital energy loss to gravitational radiation) of 85 million years. Compared to most DNS systems, it has a low eccentricity (0.08) and is thought to possess a low transverse velocity [10 km s^{-1}; (*3*)], which is difficult to explain in standard models of pulsar formation. Pulsar timing of PSR J0737-3039A/B produces results consistent with the theory of general relativity (GR) at the 0.05% level (*3*).

The distance to PSR J0737-3039A/B is estimated using the pulsar dispersion measure (DM) and models of the ionized component of the Galaxy (*4*, *5*). However, distances derived in this manner have been shown to be in error by a factor of 2 or more for individual systems [e.g., PSR B0656+14; (*6*)]. A more accurate distance to the PSR J0737-3039A/B system is needed to determine the contribution of kinematic effects to pulsar timing, the uncertainty of which would otherwise limit the precision GR tests could achieve; to establish the source of the x-ray emission from the system (*7*, *8*); and to refine the estimated radio luminosity.

We used the Australian Long Baseline Array to obtain a model-independent distance to PSR J0737-3039A/B, through the direct measurement of its annual geometric parallax. We have made seven very long baseline interferometry (VLBI) observations of PSR J0737-3039A/B over an 18-month period between August 2006 and February 2008, which allowed us to fit the system's position, proper motion, and parallax [for more details see (*9*), Table 1, and Fig. 1; all errors are 1σ unless otherwise stated].

The distance to PSR J0737-3039A/B is 1150^{+220}_{-160} pc, which is inconsistent with previous DM-based estimates of 570 pc (*4*) and 480 pc (*5*). Our measurement is more significant than the distance of 333^{+667}_{-133} pc obtained through a marginal timing parallax detection (*3*), which differed by ~1σ from our result. The revised value of electron density along the line of sight to PSR J0737-3039A/B is 0.043 cm^{-3}, suggesting that the influence of the Gum nebula (a large, ionized hydrogen region between the Earth and PSR J0737-3039A/B) along this line of sight is less than originally thought. The proper motion of PSR J0737-3039A/B is 4.37 ± 0.65 milli–arc sec $year^{-1}$, consistent with a previous timing measurement of 4.2 ± 0.6 milli–arc sec $year^{-1}$ (*3*).

We used the revised distance to PSR J0737-3039A/B to estimate the precision attainable for testing predictions of gravitational wave emission

Table 1. Fitted VLBI results for PSR J0737–3039A/B.

Parameter	Value
Right ascension α (RA; J2000)*	07:37:51.248419(26)
Declination δ (Dec; J2000)*	−30:39:40.71431(10)
Proper motion in RA ($\mu_\alpha \times \cos\delta$; milli–arc sec $year^{-1}$)	−3.82(62)
Proper motion in declination (μ_δ; milli–arc sec $year^{-1}$)	2.13(23)
Parallax (milli–arc sec)	0.87(14)
Distance (pc)	1150^{+220}_{-160}
Transverse velocity (km s^{-1})	24^{+9}_{-6}
Transverse velocity in LSR† (km s^{-1})	9^{+6}_{-3}
Reference epoch (MJD)‡	54,100

*As discussed in (*9*), the actual error in the pulsar position is dominated by the alignment of the barycentric reference frame used for pulsar timing and the quasi–inertial frame used for VLBI, and is approximately an order of magnitude greater than the formal fit error shown here. †Local standard of rest. ‡Modified Julian day.

[1]Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Mail H39, Post Office Box 218, Hawthorn, VIC 3122, Australia. [2]Curtin Institute of Radio Astronomy, Curtin University of Technology, Bentley, WA, Australia.

*To whom correspondence should be addressed. E-mail: adeller@astro.swin.edu.au

from the system. To compare the rate of change of orbital period ($\dot{P}_b$) due to the loss of energy to gravitational radiation with the GR prediction, the observed $\dot{P}_b$ must be measured as accurately as possible [the current significance is 70σ (*3*)], and contributing factors to $\dot{P}_b$ other than GR must be estimated and subtracted as accurately as possible. For PSR J0737-3039A/B, the major contributing factors are differential Galactic rotation ($\dot{P}_b^{rot}$), acceleration toward the plane of the Galaxy ($\dot{P}_b^{z}$), and the apparent acceleration caused by the transverse motion of the system [the Shklovskii effect $\dot{P}_b^{Shk}$; (*10*)], to which we will refer collectively as Galactic and kinematic contributions ($\dot{P}_b^{gk}$). For the newly calculated distance of 1150 pc, the magnitude of these effects is calculated using Eqs. 2.12 and 2.28 from (*11*) as

$$\frac{\dot{P}_b^{rot}}{P_b} = -\frac{v_0^2}{cR_0} \times \left(\cos l + \frac{\beta}{\sin^2 l + \beta^2}\right) \quad (1)$$

$$\frac{\dot{P}_b^{z}}{P_b} = -\frac{K_z}{c}\sin b \quad (2)$$

$$\frac{\dot{P}_b^{Shk}}{P_b} = \frac{\mu^2 d}{c} \quad (3)$$

where l, b, and z are Galactic longitude, latitude, and height, respectively; d and μ are pulsar distance and proper motion; R_0 and v_0 are the Galactic radius and speed of the solar system [taken to be 7.5 kpc and 195 km s^{-1}, respectively (*12*)]; K_z is the vertical gravitational potential of the Galaxy [taken from (*13*) as 0.45 km^2 s^{-2} pc^{-1} at the height of PSR J0737-3039A/B]; and c is the speed of light. The dominant uncertainties are in d (16%); μ (15%); and R_0, v_0, and K_z, whose errors are estimated at ~10%.

Equations 1 to 3 give $\dot{P}_b^{rot}/P_b = (-4.3 \pm 0.7) \times 10^{-20}$ s^{-1}, $\dot{P}_b^{z}/P_b = (3.8 \pm 0.8) \times 10^{-21}$ s^{-1}, and $\dot{P}_b^{Shk}/P_b = (5.3 \pm 1.8) \times 10^{-20}$ s^{-1}. Combining these terms gives $\dot{P}_b^{gk}/P_b = (1.5 \pm 2.1) \times 10^{-20}$ s^{-1}, and multiplying by the observed orbital period P_b = 8834.5 s (*3*) yields the net effect of these terms on the observed orbital period derivative: $\dot{P}_b^{gk} = (1.3 \pm 1.8) \times 10^{-16}$.

These contributions to the orbital period derivative are four orders of magnitude below the GR contribution, and two orders of magnitude below the current measurement error [$\dot{P}_b^{obs} = (-1.252 \pm 0.017) \times 10^{-12}$; (*3*)]. Thus, with the current accuracy in the measurement of distance and transverse velocity, GR tests can be made to the 0.01% level with PSR J0737-3039A/B using $\dot{P}_b$. However, about 10 years of further precision timing will be required to reach this point. Measuring $\dot{P}_b$ at this level will place stringent requirements on the class of gravitational theories that predict substantial amounts of dipolar gravitational radiation, exceeding the best solar system tests (*3*). Measuring the moment of inertia of pulsar A, however, would require another order of magnitude improvement in the measurement precision of $\dot{P}_b$ (*3*). In the near future, additional VLBI and/or timing measurements can be expected to reduce the error in both the distance and velocity of PSR J0737-3039A/B below 10%; however, even with negligible error in these parameters, the existing accuracy of measurements of R_0, v_0, and K_z would limit the accuracy of $\dot{P}_b$ measurements in this system to 0.004%. To attain the 10^{-5} precision necessary to measure the neutron star moment of inertia, the constants R_0 and v_0 must be measured to a precision approaching 1%.

X-ray observations of PSR J0737-3039A/B show that most of the x-ray emission from the system is modulated at the spin period of the A pulsar (*7*, *8*, *14*, *15*), but there has been considerable debate over where and how the x-rays are generated. Normally, pulsar x-ray emission is thought to have a magnetospheric or thermal origin (*16*). However, the small binary separation and interaction of the pulsar wind of the A pulsar with the magnetosphere of the B pulsar (*17*) provides alternate x-ray–generation mechanisms.

The spin-down luminosity of pulsar A is more than three orders of magnitude greater than that of pulsar B. Thus, pulsar A is likely to dominate any magnetospheric x-ray emission from the system.

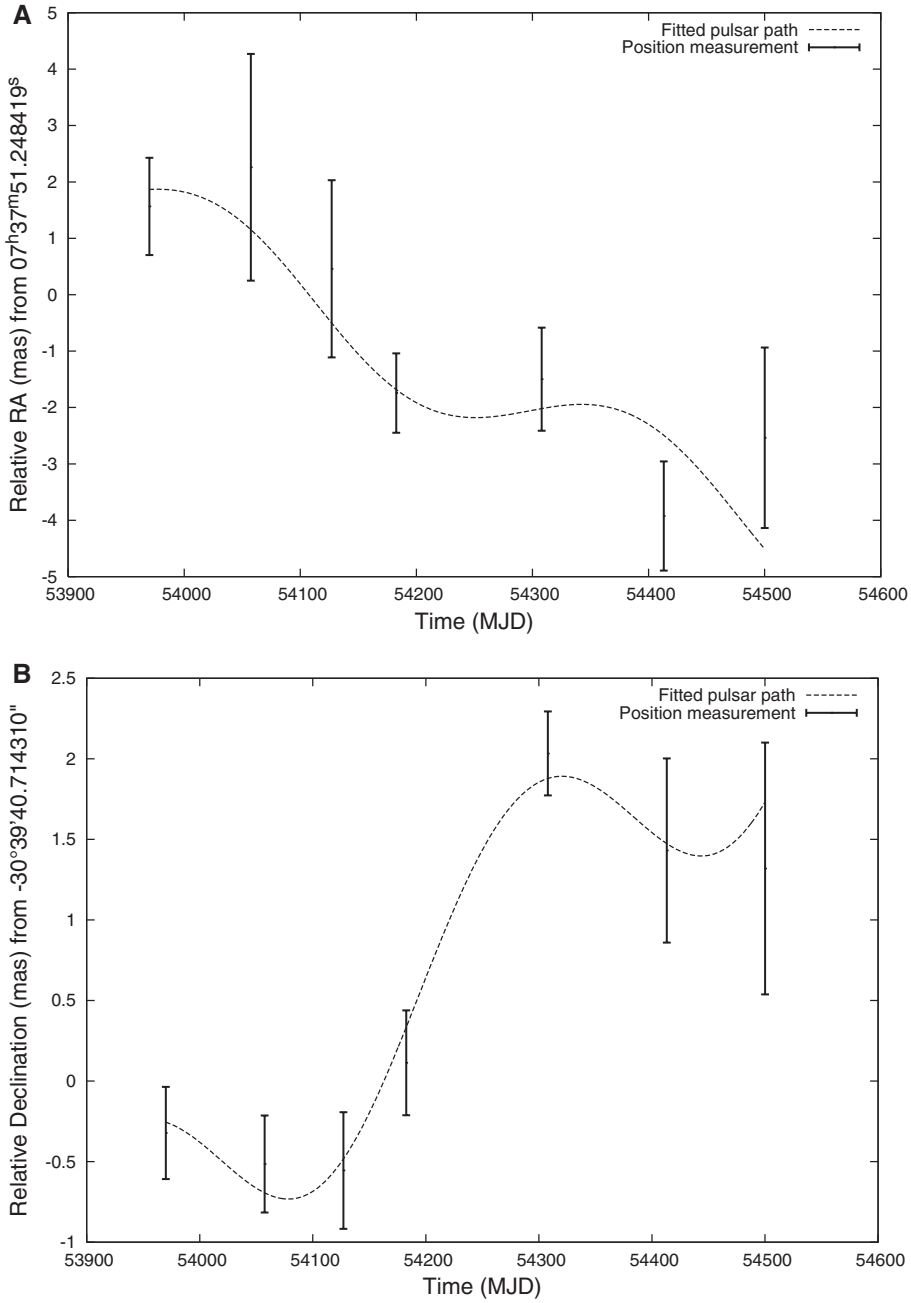


Fig. 1. Motion of PSR J0737-3039A/B plotted against time. (**A**) Offset in right ascension; (**B**) offset in declination. The best fit (dashed line) is overlaid on the measured positions. The reduced chi-squared of the fit was 0.79, implying that the measurement errors (and thus the errors on fitted parameters) might be overestimated and hence conservative.

Although a wholly magnetospheric (power-law) origin for the observed x-ray emission is plausible, the neutral hydrogen column density [$N_H \sim 1.5 \times 10^{21}$ cm^{-2}; (*8*)] implied by the model is higher than expected from the pulsar's previously assumed location in the Gum nebula. The Gum nebula is believed to be ~500 pc distant, with a depth of several hundred pc (*5*).

However, the value of N_H is consistent with the usual average of 10 neutral hydrogen atoms for every free electron along the line of sight. Our revised distance estimate places PSR J0737-3039A/B beyond the Gum nebula, implying that the measured value for N_H is not discrepant. It also increases the estimated x-ray luminosity by a factor of 5, but the revised value for a power-law fit (1.2×10^{31} erg s^{-1}) remains consistent with known relations between pulsar spin-down luminosity and x-ray luminosity (*18*). Hence, our result supports a power-law model of magnetospheric origin (from pulsar A) for the bulk of the x-ray emission from PSR J0737-3039A/B.

The discovery of PSR J0737-3039A led to a marked upward revision in the estimated Galactic merger rate of DNS systems (*19*), although uncertainty over the characteristics of recycled pulsars means that the true value of the merger rate remains poorly constrained. Specifically, the distribution of recycled pulsar luminosities is generally extrapolated from the entire pulsar population (*19*) even though it (along with the distributions of pulse shape and beaming fraction) appears to differ from the distribution for slower pulsars (*20*). Our revised distance shows that the radio luminosity of PSR J0737-3039A is a factor of 5 greater than previously assumed. If this revision were to markedly influence the recycled pulsar luminosity function, then the assumed space density of DNS systems would be reduced, with a corresponding impact on DNS merger rates estimations.

Finally, we used the measured transverse velocity for PSR J0737-3039A/B (24^{+9}_{-6} km s^{-1}) to constrain models of the formation of the system. After subtracting estimates of the peculiar motion of the solar system and Galactic rotation (*21*), we measure a transverse velocity in local standard of rest of 9^{+6}_{-3} km s^{-1}. This is comparable to the unadjusted value of 10 km s^{-1} presented in (*3*), and is within the range of transverse velocities expected for the massive stars that are DNS progenitors [~20 km s^{-1}; (*22*)]. Because the transverse velocity of PSR J0737-3039A/B is so low, if the system received a large velocity kick at birth, it must have a large radial velocity. However, there are no observational methods available to determine the radial velocity in a DNS system.

Because of the accurate measurement of its Shapiro delay, PSR J0737-3039A/B is known to lie edge-on (*3*). If the only kick it received was provided by the loss of binding energy during the supernova explosion, the resultant three-dimensional space velocity should be on the order of ~ 50 km s^{-1}, estimated from the system's observed eccentricity and orbital velocity (*23*). This space velocity would be constrained to the plane of the orbit. From simple geometry, the probability of observing a transverse velocity less than 10 km s^{-1} is about one in eight, which is small, but not unreasonable. Conversely, if the double pulsar had received a large kick (*24*), the odds of observing such a low transverse velocity become increasingly remote. Not only would the radial velocity have to be increasingly large, but the inclination angle of the system must not be altered by the kick. Hence, our transverse velocity results reinforce those of (*3*) and are consistent with the interpretation of (*25*), who argue for almost no mass loss and kick in the case of PSR J0737-3039A/B.

The implication of low kick velocities in PSR J0737-3039A/B–like systems offers a possible, albeit speculative, explanation for the formation of PSR J1903+0327, a heavy, highly recycled millisecond pulsar (mass 1.8 solar masses, period 2.15 ms) with a main-sequence companion of 1 solar mass (*26*). The orbit of such a pulsar should have been highly circularized during the mass-transfer phase (*27*). However, PSR J1903+0327 possesses an intermediate orbital eccentricity ($e = 0.44$).

A formation mechanism for PSR J1903+0327 has been suggested in which a triple system experiences a white dwarf–neutron star coalescence (*28*). However, a coalescing DNS system such as PSR J0737-3039A/B could also create a PSR J1903+0327–like pulsar. Thus, given the low velocity of PSR J0737-3039A/B, an alternative formation mechanism for PSR J1903+0327 involves a triple system containing a close DNS binary and a main-sequence star.

References and Notes

1. M. Burgay *et al.*, *Nature* **426**, 531 (2003).
2. A. G. Lyne *et al.*, *Science* **303**, 1153 (2004).
3. M. Kramer *et al.*, *Science* **314**, 97 (2006).
4. J. H. Taylor, J. M. Cordes, *Astrophys. J.* **411**, 674 (1993).
5. J. M. Cordes, T. J. W. Lazio, http://arxiv.org/abs/astro-ph/0207156 (2002).
6. W. F. Brisken, S. E. Thorsett, A. Golden, W. M. Goss, *Astrophys. J.* **593**, L89 (2003).
7. M. A. McLaughlin *et al.*, *Astrophys. J.* **605**, L41 (2004).
8. A. Possenti *et al.*, *Astrophys. J.* **680**, 654 (2008).
9. Information on materials and methods is available on *Science* Online.
10. I. S. Shklovskii, *Sov. Astron.* **13**, 562 (1970).
11. T. Damour, J. H. Taylor, *Astrophys. J.* **366**, 501 (1991).
12. W. S. Dias, J. R. D. Lépine, *Astrophys. J.* **629**, 825 (2005).
13. J. Holmberg, C. Flynn, *Mon. Not. R. Astron. Soc.* **352**, 440 (2004).
14. S. Chatterjee, B. M. Gaensler, A. Melatos, W. F. Brisken, B. W. Stappers, *Astrophys. J.* **670**, 1301 (2007).
15. A. Pellizzoni, A. Tiengo, A. De Luca, P. Esposito, S. Mereghetti, *Astrophys. J.* **679**, 664 (2008).
16. W. Becker, in *Highly Energetic Physical Processes and Mechanisms for Emission from Astrophysical Plasmas*, P. C. H. Martens, S. Tsuruta, M. A. Weber, Eds., vol. 195 of *IAU Symposium* (2000), p. 49.
17. M. Lyutikov, *Mon. Not. R. Astron. Soc.* **353**, 1095 (2004).
18. J. E. Grindlay *et al.*, *Astrophys. J.* **581**, 470 (2002).
19. V. Kalogera *et al.*, *Astrophys. J.* **601**, L179 (2004).
20. M. Kramer *et al.*, *Astrophys. J.* **501**, 270 (1998).
21. F. Mignard, *Astron. Astrophys.* **354**, 522 (2000).
22. M. W. Feast, M. Shuttleworth, *Mon. Not. R. Astron. Soc.* **130**, 245 (1965).
23. A. Blaauw, *Bull. Astron. Inst. Neth.* **15**, 265 (1961).
24. B. Willems, J. Kaplan, T. Fragos, V. Kalogera, K. Belczynski, *Phys. Rev. D Part. Fields Gravit. Cosmol.* **74**, 043003 (2006).
25. T. Piran, N. J. Shaviv, *Phys. Rev. Lett.* **94**, 051102 (2005).
26. D. J. Champion *et al.*, *Science* **320**, 1309 (2008).
27. M. A. Alpar, A. F. Cheng, M. A. Ruderman, J. Shaham, *Nature* **300**, 728 (1982).
28. E. P. J. van den Heuvel, *Science* **320**, 1298 (2008).

29. We acknowledge discussions with M. Kramer and E. van den Heuvel. A.T.D. is supported via a Swinburne University of Technology Chancellor's Research Scholarship and a Commonwealth Scientific and Industrial Research Organisation (CSIRO) postgraduate scholarship. The Long Baseline Array is part of the Australia Telescope, which is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1167969/DC1
Materials and Methods
References and Notes

31 October 2008; accepted 20 January 2009
Published online 5 February 2009;
10.1126/science.1167969
Include this information when citing this paper.

Phase-Sensitive Observation of a Spin-Orbital Mott State in Sr_2IrO_4

B. J. Kim,[1,2]* H. Ohsumi,[3] T. Komesu,[3] S. Sakai,[3,4] T. Morita,[3,5] H. Takagi,[1,2]* T. Arima[3,6]

Measurement of the quantum-mechanical phase in quantum matter provides the most direct manifestation of the underlying abstract physics. We used resonant x-ray scattering to probe the relative phases of constituent atomic orbitals in an electronic wave function, which uncovers the unconventional Mott insulating state induced by relativistic spin-orbit coupling in the layered 5*d* transition metal oxide Sr_2IrO_4. A selection rule based on intra-atomic interference effects establishes a complex spin-orbital state represented by an effective total angular momentum = 1/2 quantum number, the phase of which can lead to a quantum topological state of matter.

Transition metal oxides (TMOs) with perovskite structure are hosts to many fascinating phenomena, including high-temperature superconductivity (*1*) and colossal magnetoresistance (*2*), in which the valence *d*-electron states are described in terms of crystal-field (CF) states: triply degenerate t_{2g} states (xy, yz, zx) and doubly degenerate e_g states ($x^2 - y^2$, $3z^2 - r^2$). These CF states are all pure real functions, so that when the degeneracy is re-

moved (e.g., by the Jahn-Teller effect), the orbital angular momentum is totally quenched. However, when the spin-orbit coupling (SOC) becomes effective, the CF states are mixed with complex phases, which may partially restore the orbital angular momentum in the t_{2g} manifold. This effect is particularly pronounced in TMOs with heavy 5*d* elements, where SOC is at least an order of magnitude larger than those of TMOs with 3*d* elements and can sometimes give rise to unconventional electronic states.

5*d* TMO Sr_2IrO_4 is a layered perovskite with low-spin d^5 configuration, in which five electrons are accommodated in almost triply degenerate t_{2g} orbitals. Metallic ground states are expected in 5*d* TMOs because of their characteristic wide bands and small Coulomb interactions as compared with those of 3*d* TMOs. Sr_2IrO_4, however, is known to be a magnetic insulator (*3*, *4*). A recent study has shown that the strong SOC inherent to 5*d* TMOs can induce a Mott instability even in such a weakly correlated electron system (*5*), resulting in a localized state very different from the well-known spin $S = 1/2$ state for conventional Mott insulators, proposed to be an effective total angular momentum $J_{\mathrm{eff}} = 1/2$ state in the strong SOC limit expressed as

$$|J_{\mathrm{eff}} = 1/2, m_{J_{\mathrm{eff}}} = \pm 1/2\rangle = \frac{1}{\sqrt{3}}(|xy,\mp\sigma\rangle \mp |yz,\pm\sigma\rangle + i|zx,\pm\sigma\rangle) \quad (1)$$

where m is the component of J_{eff} along the quantization axis and σ denotes the spin state. This state derives from the addition of $S = 1/2$ to the effective orbital angular momentum $L_{\mathrm{eff}} = 1$, which consists of triply degenerate t_{2g} states but acts like the atomic $L = 1$ state with a minus sign; that is, $L_{\mathrm{eff}} = -L$. As a result, $J_{\mathrm{eff}} = 1/2$ has orbital moment parallel to spin (*6*). Note the characteristic equal mixture of *xy*, *yz*, and *zx* orbitals with complex number *i* involved in one of the factors and the mixed up-and-down spin states (*7*).

This realization of a Mott insulator with $J_{\mathrm{eff}} = 1/2$ moment provides a new playground for correlated electron phenomena, because emergent physical properties that arise from it can be drastically different from those of the conventional Mott insulators. A prime example is when $J_{\mathrm{eff}} = 1/2$ is realized in a honeycomb lattice structure where electrons hopping between $J_{\mathrm{eff}} = 1/2$ states acquire complex phase; it generates a Berry phase leading to the recent prediction of quantum spin-Hall effect at room temperature (*8*), and it also leads to the low-energy Hamiltonian of Kitaev model relevant for quantum computing (*9*). Experimental establishment of the $J_{\mathrm{eff}} = 1/2$ state is thus an important step toward these physics, and the direct probe of complex phase in the wave function has been awaited. However, it is usually difficult to retrieve the phase information experimentally, because it is always the intensity, the square modulus of the wave function, that is measured; and thus a reference, with which the state under measurement can interfere, is required.

The resonant x-ray scattering (RXS) technique uses resonance effects at an x-ray absorption edge to selectively enhance the signal of

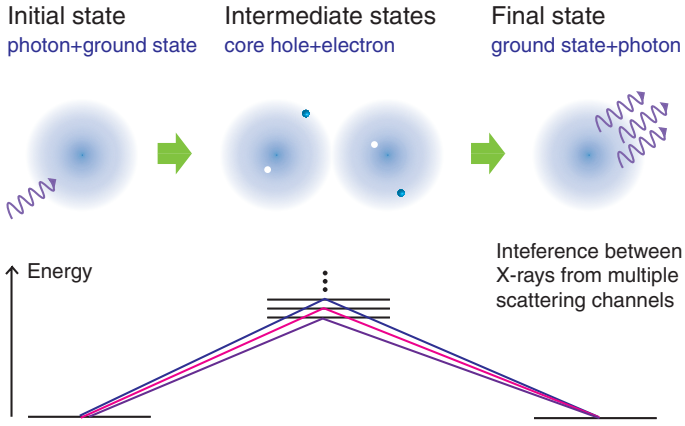
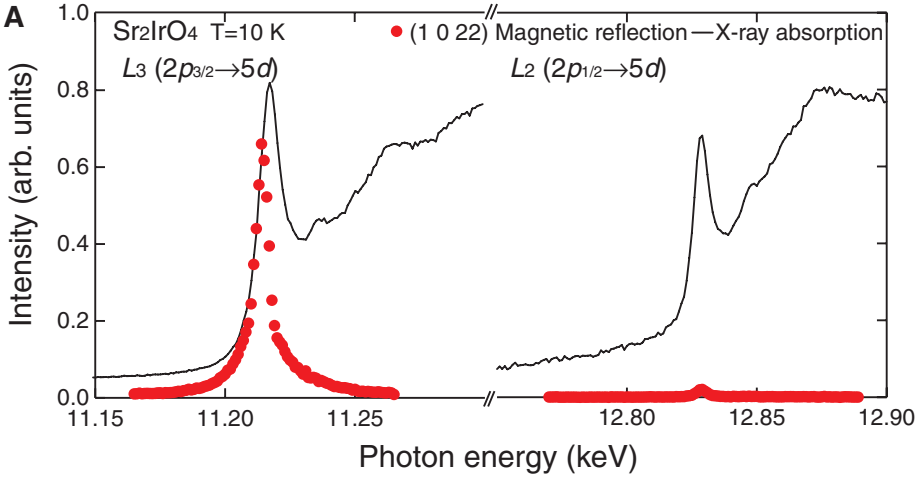


Fig. 1. Schematic diagram of the RXS process. The electron makes a trip from the initial to the final state via multiple paths of intermediate states and thereby scatters a photon with initial and final polarization of α and β, respectively. The presence of multiple scattering paths can give rise to interferences among them, which is reflected in the intensity of the scattered photon.



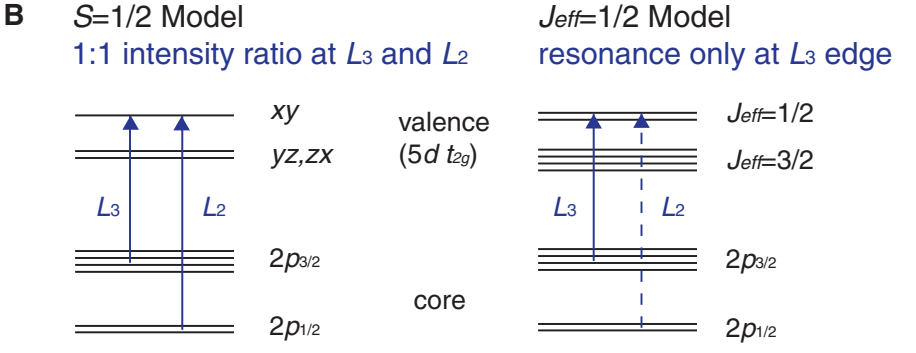


Fig. 2. Resonant enhancement of the magnetic reflection (1 0 22) at the *L* edge. (**A**) Solid lines are x-ray absorption spectra indicating the presence of Ir L_3 ($2p_{3/2}$) and L_2 ($2p_{1/2}$) edges around 11.22 and 12.83 keV. The dotted red lines represent the intensity of the magnetic (1 0 22) peak (Fig. 3C). Miller indices are defined with respect to the unit cell in Fig. 3A. (**B**) Calculation of x-ray scattering matrix elements expects equal resonant scattering intensities at L_3 and L_2 for the $S = 1/2$ model. For the $J_{\mathrm{eff}} = 1/2$ model, in contrast, the resonant enhancement occurs only for the L_3 edge, and zero enhancement is expected at the L_2 edge.

[1]Department of Advanced Materials, University of Tokyo, Kashiwa 277-8561, Japan. [2]Magnetic Materials Laboratory, RIKEN Advanced Science Institute, Wako 351-0198, Japan. [3]RIKEN SPring-8 Center, Sayo 679-5148, Japan. [4]Department of Physical Science, Hiroshima University, Higashi-Hiroshima 739-8526, Japan. [5]Department of Physics, Kwansei-Gakuin University, Sanda 669-1337, Japan. [6]Institute of Multidisciplinary Research for Adavanced Materials, Tohoku University, Sendai 980-8577, Japan.

*To whom correspondence should be addressed. E-mail: bjkim6@gmail.com (B.J.K.); htakagi@k.u-tokyo.ac.jp (H.T.)

interest, and has become a powerful tool for investigating ordering phenomena (*10*, *11*). So far, the emphasis has been seen only in the amplification of the signal. However, the RXS signal contains important information about the phase of the wave function for valence electrons, because RXS results from quantum interference between different scattering paths via intermediate states of a single site. The RXS process is described by the second-order process of electron-photon coupling perturbation, as schematically shown in Fig. 1, and its scattering amplitude $f_{\alpha\beta}$ from a single site is expressed under dipole approximation by

$$f_{\alpha\beta} = \sum_m \frac{m_e \omega_{im}^3}{\omega} \frac{\langle i|R_\beta|m\rangle\langle m|R_\alpha|i\rangle}{\hbar\omega - \hbar\omega_{im} + i\Gamma/2} \quad (2)$$

In this process, a photon with energy $(\hbar)\omega$ is scattered by being virtually absorbed and emitted with polarizations α and β, respectively; and in the course of the process, an electron of mass m_e makes dipole transitions through position operators R_α and R_β from and to the initial state i, via all possible intermediate states m, collecting the phase factors associated with the intermediate states, weighted by some factors involving energy differences between the initial and intermediate states $(\hbar)\omega_{im}$ and the lifetime broadening energy Γ. The interference between various scattering paths is directly reflected in the scattering intensities of the photon, and in this way the valence electronic states can be detected with phase sensitivity. This process can be contrasted with that in x-ray absorption spectroscopy (XAS), which is a first-order process and measures only the amplitudes of the individual paths, or transition probabilities to various valence states.

We have applied this technique to explore unconventional electronic states produced by the strong SOC in Sr_2IrO_4. Sr_2IrO_4 is an ideal system in which to fully use this technique. The magnetic Bragg diffraction in magnetically ordered Sr_2IrO_4 comes essentially from scattering by Ir t_{2g} electrons, to which RXS using the L edge ($2p \rightarrow 5d$) can be applied to examine the electronic states. The wavelength at the L edge of $5d$ Ir is as short as ~1 Å, in marked contrast to >10 Å for $3d$ elements. This short wavelength makes the detection of RXS signals much easier than in $3d$ TMOs, because there exists essentially no constraint from the wavelength in detecting the magnetic Bragg signal. Moreover, the low-spin $5d^5$ configuration, a one-hole state, greatly reduces the number of intermediate states and makes the calculation of scattering matrix elements tractable. The excitation to the t_{2g} state completely fills the manifold, and the remaining degrees of freedom reside only in the $2p$ core holes. Because the intermediate states are all degenerate in this case, the denominator factors involving energies and lifetimes of the intermediate states in Eq. 2 can drop out. A careful analysis of the scattering intensity can show that the wave function given by Eq. 1 represents the ground state in Sr_2IrO_4 (*4*).

Figure 2A shows the resonance enhancement of the magnetic reflection (1 0 22) at the L edge of a Sr_2IrO_4 single crystal (*4*), overlaid with XAS spectra to show the resonant edges. Whereas there is a huge enhancement of the magnetic reflection by a factor of ~10^2 at the L_3 edge, the resonance at L_2 is small, showing less than 1% of the intensity at L_3. The constructive interference at L_3 gives a large signal that allows the study of magnetic structure, whereas the destructive interference at the L_2 edge hardly contributes to the resonant enhancement.

To find out the necessary conditions for the hole state leading to the destructive interference at the L_2 edge, we calculate the scattering amplitudes. The most general wave function for the hole state in the t_{2g} manifold involves six basis states, which can be reduced by block-diagonalizing the spin-orbit Hamiltonian as

$$c_1|xy,+\sigma\rangle + c_2|yz,-\sigma\rangle + c_3|zx,-\sigma\rangle \quad (3)$$

With its time-reversed pair, they fully span the t_{2g} subspace. We neglect higher-order corrections such as small residual coupling between t_{2g} and e_g manifolds. In the limit of the tetragonal crystal field [$Q \equiv E(d_{xy}) - E(d_{yz,zx})$] due to the elongation of octahedra much larger than SOC (λ_{SO}) (that is, $Q >> \lambda_{SO}$), the ground state will approach $c_1 = 1$ and $c_2 = c_3 = 0$ and become a $S = 1/2$ Mott insulator, whereas in the other limit of strong SOC, $Q << \lambda_{SO}$, c_i's will all be equal in magnitude, with c_1, c_2 pure real and c_3

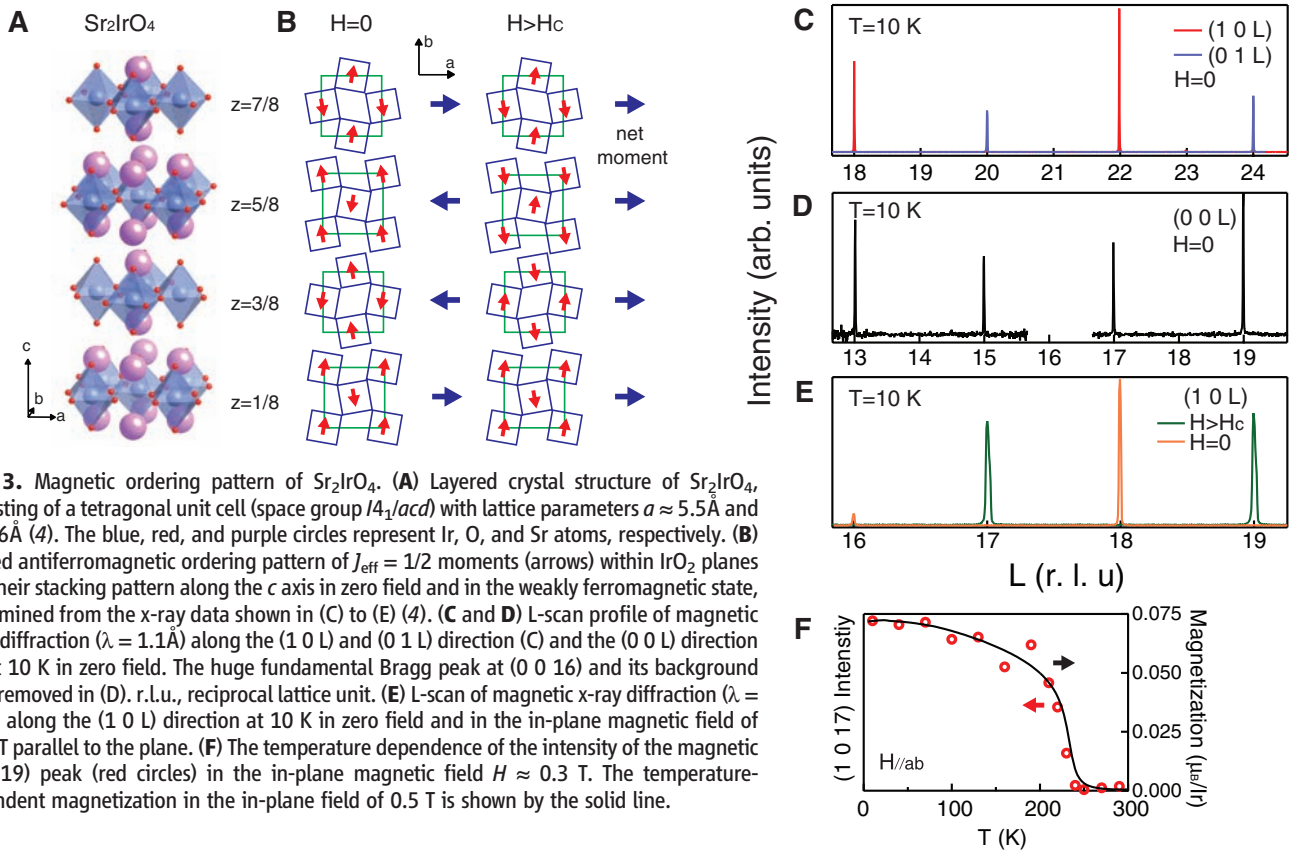


Fig. 3. Magnetic ordering pattern of Sr_2IrO_4. (**A**) Layered crystal structure of Sr_2IrO_4, consisting of a tetragonal unit cell (space group $I4_1/acd$) with lattice parameters $a \approx 5.5$Å and $c \approx 26$Å (*4*). The blue, red, and purple circles represent Ir, O, and Sr atoms, respectively. (**B**) Canted antiferromagnetic ordering pattern of $J_{eff} = 1/2$ moments (arrows) within IrO_2 planes and their stacking pattern along the c axis in zero field and in the weakly ferromagnetic state, determined from the x-ray data shown in (C) to (E) (*4*). (**C** and **D**) L-scan profile of magnetic x-ray diffraction (λ = 1.1Å) along the (1 0 L) and (0 1 L) direction (C) and the (0 0 L) direction (D) at 10 K in zero field. The huge fundamental Bragg peak at (0 0 16) and its background were removed in (D). r.l.u., reciprocal lattice unit. (**E**) L-scan of magnetic x-ray diffraction (λ = 1.1Å) along the (1 0 L) direction at 10 K in zero field and in the in-plane magnetic field of ≈0.3 T parallel to the plane. (**F**) The temperature dependence of the intensity of the magnetic (1 0 19) peak (red circles) in the in-plane magnetic field $H \approx 0.3$ T. The temperature-dependent magnetization in the in-plane field of 0.5 T is shown by the solid line.

pure imaginary as in Eq. 1. Calculation of $f_{\alpha\beta}$ for the L_2 edge using Eq. 5 gives

$$f_{\alpha\beta} = \begin{pmatrix} (c_1+ic_3)(c_1^*-ic_3^*) & (ic_1^*+c_3^*)(c_1-c_2) & 0 \\ (-ic_1+c_3)(c_1^*-c_2^*) & (c_1-c_2)(c_1^*-c_2^*) & 0 \\ 0 & 0 & (c_2+ic_3)(c_2^*-ic_3^*) \end{pmatrix} \quad (4)$$

Because the magnetic signal comes from the imaginary part of off-diagonal elements (*12*), the necessary condition for the vanishing intensity is

$$c_1 = c_2 \qquad \text{or} \qquad c_3 = ic_1 \quad (5)$$

This condition places very stringent constraints on the allowed hole state and rules out all single-orbital $S = 1/2$ models. For the $S = 1/2$ case, the imaginary part of the off-diagonal elements does not vanish and equal resonant intensities are expected at the L_2 and L_3 edges (Fig. 2B). Given the constraints in Eq. 5, the scattering intensities at L_3 edge are calculated to be

$$I_{L_3} = \frac{1}{4}(Im(c_1c_3^*))^2 \qquad \text{or}$$

$$I_{L_3} = \frac{1}{4}(Re(c_1^*c_2))^2 \quad (6)$$

respectively. Thus, taking the phase convention $c_1 = 1$ without loss of generality, the resonant intensity at L_3 measures the imaginary part of c_3 relative to the real part of c_2, which is also a measure of orbital angular momentum. The large enhancement at L_3 necessarily implies that the *yz* orbital is out of phase with the *zx* orbital, the contrast between L_3 and L_2 being maximal when the relative phase is $\pi/2$. Taking the constraints of Eqs. 5 and 6 together, we conclude that the ground state is very close to the $J_{\text{eff}} = 1/2$ limit (c_1:c_2:c_3 = 1:1:*i*), and the minute enhancement at L_2 edge shows the smallness of the deviation from the $J_{\text{eff}} = 1/2$ limit coming from the factors not taken into account (*13*). The wave function in Eq. 2, representing $J_{\text{eff}} = 1/2$, indeed gives zero off-diagonal elements in $f_{\alpha\beta}$ for L_2 and nonzero elements for the L_3 edge. This is a direct measurement of phase and provides evidence for the $J_{\text{eff}} = 1/2$ state in Sr_2IrO_4.

Having identified the nature of the local moment, we now look at the global magnetic structure using the enhanced signal due to the resonance at the L_3 edge. Sr_2IrO_4 shows a metamagnetic transition below 240 K and, above the metamagnetic critical field H_C (≈0.2 T well below 240 K), shows weak ferromagnetism with a saturation moment of ≈0.1 μ_B/Ir (*4*). The origin of this field-induced weak ferromagnetism has remained unidentified, because the neutron diffraction data did not show any detectable indication of magnetic ordering (*14*). Our RXS results indicate that the magnetic structure of Sr_2IrO_4 is canted antiferromagnetic.

Figure 3A shows the crystal structure containing four IrO_2 layers in a unit cell, enlarged by superstructure from the rotational distortion of octahedra (*14*). Figure 3B shows the magnetic ordering pattern determined from the experiment shown in Fig. 3, C to E. The arrows in Fig. 3B do not represent spins but $J_{\text{eff}} = 1/2$ moments. In zero field, the magnetic reflections are observed at (1 0 4*n*+2) and (0 1 4*n*), which implies that the moments are aligned antiferromagnetically within a layer and the symmetry changes from tetragonal to orthorhombic (Fig. 3C). The canting of the moments yields a nonzero net moment within a layer, which orders in the up-down-down-up antiferromagnetic pattern along the *c* axis. This is evidenced by the presence of (0 0 *odd*) peaks shown (Fig. 3D). The width of the peak gives an estimate of interlayer correlation length of 100 *c* or 400 IrO_2 layers. When the magnetic field greater than H_C is applied, the peaks at (1 0 4*n*+2) disappear and new peaks show up at (1 0 *odd*) (Fig. 3E), which implies that the net moments in the planes are aligned ferromagnetically to produce a macroscopic field. The temperature dependence of the scattering intensity in the weakly ferromagnetic state above H_C, shown in Fig. 3F, scales very well with that of the magnetization and confirms again the magnetic nature of the peaks.

Our study demonstrates that x-rays can be extended to a new level to probe even finer details of magnetic structure. Until now, only the intersite interference effects were used to study ordering phenomena over a length scale of many lattice sites. A quantitative analysis on the interference effects within a single site provides phase information on the constituent wave function of the electron responsible for the magnetism. This technique should find important applications in systems where complex phases give rise to novel physics.

References and Notes

1. J. G. Bednorz, K. A. Muller, *Z. Phys. B* **64**, 189 (1986).
2. For a review, see *Colossal Magnetoresistive Oxides*, Y. Tokura, Ed. (Gordon & Breach Science, New York, 2000).
3. G. Cao, J. Bolivar, S. McCall, J. E. Crow, R. P. Guertin, *Phys. Rev. B* **57**, R11039 (1998).
4. Materials and methods are available as supporting material on *Science* Online.
5. B. J. Kim *et al*., *Phys. Rev. Lett.* **101**, 076402 (2008).
6. The parallel spin and orbital configuration are in accordance with Hund's third rule in the case of more than half-filled shell, despite its apparent contradiction implied by the $J_{\text{eff}} = 1/2$ quantum number [see (*5*)].
7. This results in an orbital moment twice as large as that of spin,: $< L_z > = 2/3$ and $2 < S_z > = 1/3$, with $< L_z > + 2 < S_z > = 1$.
8. A. Shitade, H. Katsura, X.-L. Qi, S.-C. Zhang, N. Nagaosa, http://arxiv.org/abs/0809.1317.
9. G. Jackeli, G. Khaliullin, http://arxiv.org/abs/0809.4658.
10. Y. Murakami *et al*., *Phys. Rev. Lett.* **81**, 582 (1998).
11. P. Abbamonte *et al*., *Science* **297**, 581 (2002).
12. M. Blume, *Resonant Anomalous X-ray Scattering, Theory and Applications* (Elsevier Science, Amsterdam, 1994).
13. The small enhancement at L_2 can be explained by finite 10*Dq*, which results in a small admixture of *xy* and $x^2 - y^2$ and small tetragonal splitting. It is not a priori obvious that $J_{\text{eff}} = 1/2$ can be stabilized in the presence of the tetragonal crystal field, but the rotational distortion (Fig. 3A) acts in the opposite direction to lower the *xy* state, and this delicate balance may result in the $J_{\text{eff}} = 1/2$ state (*15*).
14. M. K. Crawford *et al*., *Phys. Rev. B* **49**, 9198 (1994).
15. B. J. Kim *et al*., *Phys. Rev. Lett.* **97**, 106401 (2006).

16. The authors gratefully acknowledge discussions with C. Kim, J.-H. Park, S. Fujiyama, J. Matsuno, B. Smith, and G. Khaliullin; preliminary x-ray analysis by D. Hashizume; and continuous encouragement by M. Takata and T. Ishikawa. This work was partly supported by a Ministry of Education, Culture, Sports, Science and Technology (MEXT) Grant-in-Aid for Scientific Research (S) (19104008) and Grant-in-Aid for Scientific Research on Priority Areas "Novel States of Matter Induced by Frustration" (19052001 and 19052008).

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5919/1329/DC1
Materials and Methods
Figs. S1 and S2
References

13 October 2008; accepted 16 January 2009
10.1126/science.1167106

The Earliest Horse Harnessing and Milking

Alan K. Outram,[1]* Natalie A. Stear,[2] Robin Bendrey,[3,7] Sandra Olsen,[4] Alexei Kasparov,[1,5] Victor Zaibert,[6] Nick Thorpe,[7] Richard P. Evershed[2]

Horse domestication revolutionized transport, communications, and warfare in prehistory, yet the identification of early domestication processes has been problematic. Here, we present three independent lines of evidence demonstrating domestication in the Eneolithic Botai Culture of Kazakhstan, dating to about 3500 B.C.E. Metrical analysis of horse metacarpals shows that Botai horses resemble Bronze Age domestic horses rather than Paleolithic wild horses from the same region. Pathological characteristics indicate that some Botai horses were bridled, perhaps ridden. Organic residue analysis, using $\delta^{13}C$ and δD values of fatty acids, reveals processing of mare's milk and carcass products in ceramics, indicating a developed domestic economy encompassing secondary products.

The domestication of the horse is associated with the spread of Indo-European languages and culture, bronze metallurgy, and specialized forms of warfare (*1*–*3*). Genetic studies of modern domestic horse breeds (*Equus caballus*) (*4*, *5*) imply either multiple domestication events or domestic stallions from a single original lineage being bred with captured local juvenile wild mares (*6*, *7*), but fail to clearly identify when and where horse domestication first took place. A prime candidate for this locus is the Eurasian steppe, specifically the Botai culture, northern

Kazakhstan, in the mid–fourth millennium B.C.E., where faunal assemblages consist almost entirely of horse remains (*1*, *6*–*9*). The case for horse herding within the Botai culture includes a semi-sedentary settlement structure, incompatible with hunting mobile wild herds, and skeletal element abundances lacking differential transport patterns associated with large quarries. Tools that were probably used for hide working and producing leather straps predominate over projectile points and other hunting equipment. However, age structures within horse herds at Botai do not clearly indicate a husbanded rather than hunted population (*6*, *7*, *9*, *10*). Indirect evidence for domesticated horses is strongly suggestive but inconclusive. Here, we discuss three new lines of direct evidence to confirm early domestication of horses in the Botai culture.

Horse metapodia are useful in archaeozoological metrical analyses because of their load-bearing function and proclivity to undergo morphological changes relating to breed and differing physical activities. Specimens were selected from four sites in northern/central Kazakhstan, including Botai (during the 2005–2006 seasons) (*11*); the Tersek culture sites of Kozhai and Kumkeshu [supporting online material (SOM)] (*12*); and the large settlement of Kent (*13*), dating to the late Bronze Age (circa 1300 to 900 B.C.E.), by which time the horses are clearly domestic. Twelve key measurements were taken (*14*), and principal components analysis (PCA) revealed patterning between sites, with the loading plot indicating that most differentiation related to the ratio of greatest length (GL) against four key width measurements. The ratios between GL and greatest breadth of proximal epiphysis (Bp), smallest width of diaphysis (SD), smallest depth of diaphysis (SDD), and greatest breadth of distal epiphysis (Bd) are indices of general limb slenderness, rather than overall size, and have been used in differentiating equid species (*15*). All four ratios show exactly the same pattern. Specimens from the two Tersek sites, Kozhai and Kumkeshu, show considerable similarity, whereas the domestic horses from Kent are appreciably more slender. The Botai horses are also significantly more slender than those of the Tersek sites, with the distribution of ratios at Kent and Botai being very similar (no significant difference at 95% confidence interval in a Student's *t* test of difference between means of the four ratios). However, the Botai specimens showed significant differences in the means of the SD/GL, SDD/GL, and Bd/GL ratios of Kumkeshu and Kozhai at high confidence levels (well above 95%), whereas the Bp/GL test fell just short of a 90% confidence level. Figure 1 shows a scatter plot of two of the mean ratios (SD/GL and Bd/GL) for the sites discussed above and two other published populations: Kuznetsk (*16*) late Pleistocene horses (deriving from Palaeolithic sites in Novokuznetsk, southern Siberia; being the geographically closest wild horses for which the appropriate measurements are available) and modern Mongolian domestic horses (*17*). The Botai horses cluster very closely with the Bronze Age domestic horses from Kent and modern Mongolian domestic horses. The Kuznetsk Paleolithic horses appear to be much less slender, and the Tersek population displays intermediate morphology. The domestic populations are clearly more slender and, most significantly, the Botai horses plot with the modern Mongolian and Bronze Age domestic specimens, providing evidence that the Botai horses were domesticated.

We examined evidence for bitting damage resulting from harnessing with a bridle or similar restraint (this may refer to a range of possible mouthpieces, including leather thong bridles), in which damage to the skeletal tissues of the mouth occurs when a horse is ridden or driven with a bit/bridle. A macroscopic method that quantifies bitting damage to the mesial or anterior edge of mandibular second premolars (P_2s) was applied (*18*). When a horse is bitted, the bit is placed in the mouth on the mandibular diastema (the bridge of bone between the anterior and cheek teeth), where it can come into contact with the mesial edge of P_2s and cause a recognizable vertical strip of wear through the cementum to expose the enamel. More severe wear exposes dentine below the enamel. However, not all enamel exposure on the anterior border of P_2s is

[1]Department of Archaeology, University of Exeter, Exeter, EX4 4QE, UK. [2]School of Chemistry, University of Bristol, Bristol, BS8 1TS, UK. [3]CNRS UNR 5197, Muséum Nationale d'Histoire Naturelle, Paris 75005, France. [4]Carnegie Museum of Natural History, Pittsburgh, PA 15206–3706, USA. [5]Institute for the History of Material Culture, RAS, St. Petersburg 191186, Russia. [6]Kokshetau University, 020000 Kokshetau, Akmolinsk Oblast, Kazakhstan. [7]Department of Archaeology, University of Winchester, Winchester SO22 4NR, UK.

*To whom correspondence should be addressed. E-mail: a.k.outram@ex.ac.uk

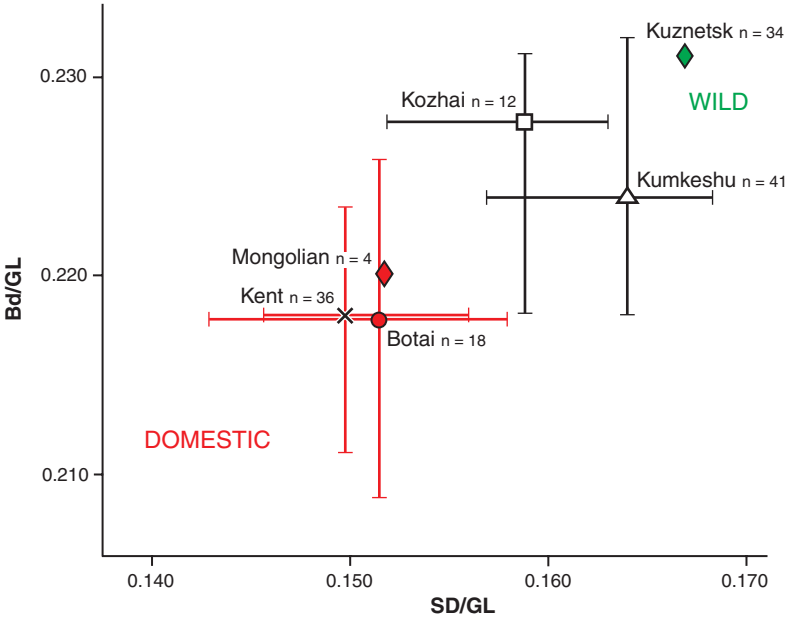


Fig. 1. Scatter plot of mean ratios of measurements Bd/GL and SD/GL on horse metacarpals from different ancient and modern populations. Botai horses plot in association with other ancient and modern domestic populations. Bars are mean ± interquartile ranges (not available from published data for Kuznetsk or Mongolian populations).

Fig. 2. A Botai stallion's lower second premolar (mesial edge), displaying a clear parallel-sided band of bit wear that penetrates through the cementum and enamel. This morphology and depth of wear occur only in bridled animals [figures 2 and 4 of (*18*)].

due to bitting, and other forms of damage can occur, such as dietary wear, so interpretations must consider the size and shape of the area exposed. The bit also comes into contact with the upper surface of the diastema, which can lead to periostitis at this site (*19*), and repetitive contact can result in the deposition of pathological new bone or destruction of bone (*18*). The criteria used in this investigation are based on studies of modern animals (bitted and unbitted) with known life histories (*18*). From the 2005–2006 Botai excavations, 15 P_2s (table S1) displaying full occlusal wear (that is, in excess of 4 years old) and mandibular diastemata were examined. Of nine measurable P_2s, specimen no. 7 exhibited changes to the anterior edge that could be unambiguously attributed to bitting damage, as evidenced by wear that penetrated through the enamel to expose the dentine (Fig. 2). Mammalian tooth enamel is very hard (about 300 to 400 Vickers hardness number), and although enamel exposure occurs in both bitted and unbitted equids (cementum can also be removed through dietary wear), dentine exposure in this area occurs only in bitted/bridled animals (*18*). Two further P_2s exhibited bands of enamel exposure that were possibly caused by bit wear, but the form of this is not a clear parallel-sided band, so it is not unequivocal. Application of the mandibular diastema scoring system (*18*) identified four cases of new bone formation at a level indicating bitting/bridling. The clearly bitted P_2 (no. 7) was dated by accelerator mass spectrometry to 4658 ± 33 years before the present (3521 to 3363 calendar years B.C.E., 94.6% probability; Oxford Radiocarbon Accelerator Unit reference OxA-18383), which is consistent with the Botai culture. Thus, 5 out of 15 mandibles studied provided evidence of bitting damage.

The traditional economies of modern Kazakhstan exploit horses for both meat and milk. Degraded animal fat survives in archaeological pottery (*20*),

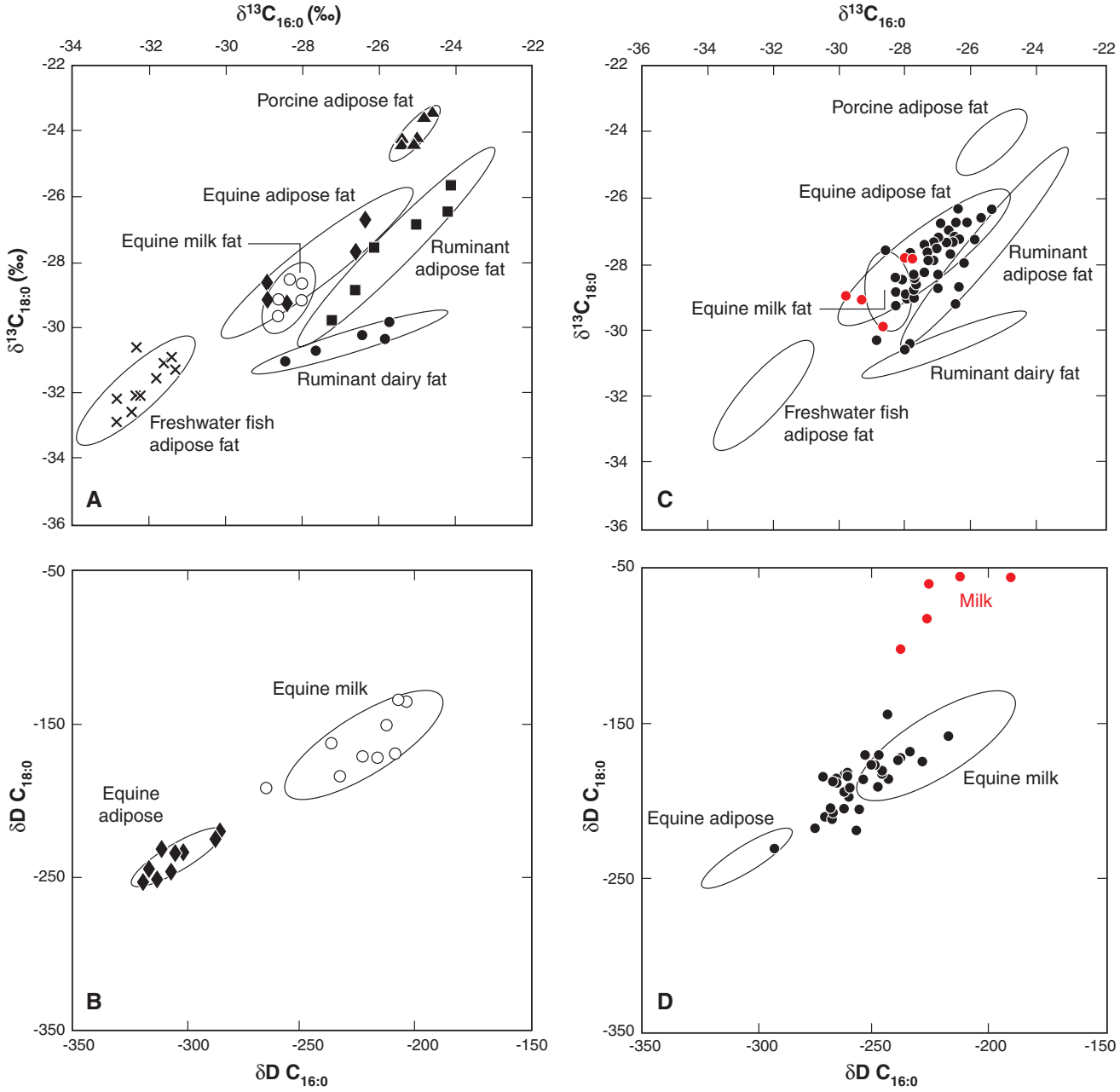


Fig. 3. Scatter plots of $\delta^{13}C$ and δD values of the $C_{18:0}$ and $C_{16:0}$ fatty acid animal fats of modern reference fats (**A** and **B**, respectively) and of organic residues from archaeological potsherds (**C** and **D**, respectively) from Kazakhstan. (C) shows the $\delta^{13}C$ values of $C_{18:0}$ and $C_{16:0}$ fatty acids of 50 analyzable lipid residues from 89 Botai potsherds sampled, and (D) shows the δD values of the $C_{18:0}$ and $C_{16:0}$ fatty acids of the residues from potsherds assigned as equine fats. All confidence ellipses are mean ± SD and correspond to the values exhibited by modern reference fats. The residues highlighted in red correspond to archaeological equine milk fats.

and its sources can be classified on the basis of the $\delta^{13}C$ values of the major *n*-alkanoic acids, palmitic ($C_{16:0}$) and stearic ($C_{18:0}$) acid, which allows nonruminant and ruminant carcass and ruminant dairy fats to be distinguished (*21*, *22*). Ruminant dairying developed relatively quickly after the domestication of cattle, sheep, and goats (*23*); direct identification of mares' milk in pottery vessels would be clear evidence of horse domestication. Sampling of modern animal fats, including equine adipose and milk fats from Kazakh animals fed on the natural steppe vegetation, was undertaken. Although horse adipose and milk fats were resolved from the fats of ruminant animals, their $\delta^{13}C$ values overlap (Fig. 3A); hence, although equine fats can be detected with this approach, equine milk cannot be unambiguously identified.

To achieve separation, we used compound-specific deuterium isotope (δD) analysis of the major *n*-alkanoic acids, exploiting the phenomenon that in midcontinental regions, such as the Eurasian steppe, the δD values of summer and winter precipitation consistently differ by >100 per mil (*24*). Tissue lipid integrates both the water and dietary deuterium signal (*25*), hence their adipose fat integrates the annual δD signal. However, summer milk fat records only the summer δD signal; thus, the δD values of fatty acids in summer milk and adipose fats will differ because of the large difference in the δD values of summer versus mean annual precipitation. Figure 3B confirms that the δD values of the modern reference horse fats exhibit the predicted difference between adipose and summer milk fats.

The $\delta^{13}C$ values from the major fatty acid components of the Botai cooking vessels confirm the preponderance of horse fat residues (Fig. 3C), mirroring the dominance of horse bones at the site. A few residues fall into the ruminant reference distribution, which may well indicate the presence of small numbers of hunted cervids or bovids. Most significantly, the δD values show two distinct clusters. The red points in Fig. 3, C and D, correspond to the respective $\delta^{13}C$ and δD values of the same five potsherds. All the δD values of the fatty acid components of these residues exhibit significantly elevated δD values. They very likely derive from mare's milk because of their relative displacement from the major cluster of carcass fats (Fig. 3D and SOM). The relatively higher δD values in the archaeological fats are consistent with increased aridity during this period of prehistory (*26*, *27*).

Although existing archaeological evidence for horse domestication at Botai is inconclusive (*10*), our new skeletal evidence, based on metacarpal metrics, supports the presence of a proportion of domesticated horses in the Botai herds. Moreover, our bitting damage evidence indicates the use of bridles to control working animals and supports assertions that finds of leather thong–producing tools are consistent with horse domestication (*6*, *7*). Finally, evidence for extensive horse carcass product processing in pottery vessels provides direct evidence for their exploitation as a dietary staple. The demonstration of mares' milk processing confirms that at least some of the mares at Botai were domesticated. The fact that horse milking existed in a region remote from the locus of ruminant domestication in the "Fertile Crescent" and in an area seemingly devoid of domestic ruminants indicates that the evolution of strategies for exploiting animals for their milk was not contingent on the adoption of the conventional "agricultural package," as it appears to have developed independently in the Botai region.

References and Notes

1. D. W. Anthony, *The Horse, the Wheel, and Language* (Princeton Univ. Press, Princeton, NJ, 2007).
2. J. P. Mallory, *In Search of the Indo-Europeans: Language, Archaeology and Myth* (Thames and Hudson, London, 1996).
3. S. Piggott, *Wagon, Chariot and Carriage* (Thames & Hudson, London, 1992).
4. T. Jansen *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10905 (2002).
5. C. Vilà *et al*., *Science* **291**, 474 (2001).
6. S. L. Olsen, in *Horses and Humans: The Evolution of Human-Equine Relationships*, S. L. Olsen, S. Grant, A. M. Choyke, L. Bartosiewicz, Eds. (Archaeopress, Oxford, 2006), pp. 81–113.
7. S. L. Olsen, in *Documenting Domestication*, M. A.Zeder, D .G. Bradley, E. Emshwiller, B. D. Smith, Eds. (Univ. of California Press, Berkeley, CA, 2006), pp. 245–269.
8. M. Levine, in *Late Prehistoric Exploitation of the Eurasian Steppe*, M. Levine, Y. Rassamakin, A. Kislenko, N. Tatarintseva, Eds. (McDonald Institute, Cambridge, 1999), pp. 5–58.
9. S. Olsen, B. Bradley, D. Maki, A. Outram, in *Beyond the Steppe and the Sown*, D. L. Peterson, L. M. Popova, A. T. Smith, Eds. (Brill, Leiden, Netherlands, 2006), pp. 89–111.
10. M. Levine, in *Prehistoric Steppe Adaptation and the Horse*, M. Levine, C. Renfrew, K. Boyle, Eds. (McDonald Institute of Archaeological Research, Cambridge, 2003), pp. 1–7.
11. V. F. Zaibert, A. Tyulevaev, A. V. Zadorozhnyj, U. Kulakov, *Tajny Drevnyej Stepi: Issledovaniya Poseleniya Botaj (2004-2006)* (Kokshetauskogo Universiteta, Kokshetau, Kazakhstan, 2007).
12. S. S. Kalieva, V. N. Logvin, *Skotovody Turgaya v Tret'em Tysyacheletii do Nashej Ehry* (Kustanajckij Gosudarstvennyj Universitet, Kustanaj, Kazakhstan, 1997).
13. V. V. Yevdokimov, V. V. Varfolomeev, *Ehpoxa Bronzy Central'nogo i Cevernogo Kazaxstana* (Karagandinskij Gosudarstvennyj Universitet, Karaganda, Kazakhstan, 2002).
14. V. Eisenmann, in *Equids in the Ancient World*, R. H. Meadow, H.-P. Uerpmann, Eds. (Dr. Ludwig Reichert Verlag, Wiesbaden, Germany, 1986), pp. 67–116.
15. A. S. Gilbert, in *Equids in the Ancient World, Volume II*, R. H. Meadow, H.-P. Uerpmann, Eds. (Dr. Ludwig Reichert Verlag, Wiesbaden, Germany, 1991), pp. 75–122.
16. I. V. Foronova, in *Equids in Time and Space*, M. Mashkour, Ed. (Oxbow Books, Oxford, 2006,) pp. 20–30.
17. I. E. Grebnev, S. K. Vasil'ev, in *Steregushchie Zoloto Grify (Ak-Alakhinskie Kurgany)*, N. V. Polos'mak, Ed. (Nauka, Novosibirsk, Russia, 1994), pp. 106–111.
18. R. Bendrey, *J. Archaeol. Sci.* **34**, 1036 (2007).
19. D. G. Bennett, in *Equine Dentistry*, G. J. Baker, J. Easley, Eds. (Elsevier Saunders, London, 2005), pp. 9–22.
20. R. P. Evershed *et al*., *Acc. Chem. Res.* **35**, 660 (2002).
21. S. N. Dudd, R. P. Evershed, *Science* **282**, 1478 (1998).
22. M. S. Copley *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1524 (2003).
23. R. P. Evershed *et al*., *Nature* **455**, 528 (2008).
24. K. Rozanski, L. Araguas-Araguas, R. Gonfiantini, in *Climate Change in Continental Records*, P. K. Swart, K. C. Lohmann, J. McKenzie, S. Savin, Eds. (American Geophysical Union, Washington, DC, 1993), pp. 1–36.
25. K. A. Hobson, L. Atwell, L. I. Wassenar, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8003 (1999).
26. C. V. Kremenetski, P. E. Tarasov, A. E. Cherkinsky, *Quat. Int.* **41-42**, 125 (1997).
27. P. E. Tarasov, D. Jolly, J. O. Kaplan, *Palaeogeog. Palaeocl.* **136**, 281 (1997).

28. Thanks to E. Usmanova, V. Varfolomeev, V. Loman, V. Logvin, S. Kalieva, L. Gayduchenko, D. Chivall, R. Berstan, B. Bradley, S. Babich, and E. Isayev. We are grateful for the following grants: Natural Environment Research Council grant NE/B504506/1, British Academy grant SG-35540, and NSF grants BS9816476 and BCS0415441.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5919/1332/DC1
SOM Text
Table S1
References

17 November 2008; accepted 14 January 2009
10.1126/science.1168594

Promoting Intellectual Discovery: Patents Versus Markets

Debrah Meloso,[1]* Jernej Copic,[2] Peter Bossaerts[3]

Because they provide exclusive property rights, patents are generally considered to be an effective way to promote intellectual discovery. Here, we propose a different compensation scheme, in which everyone holds shares in the components of potential discoveries and can trade those shares in an anonymous market. In it, incentives to invent are indirect, through changes in share prices. In a series of experiments, we used the knapsack problem (in which participants have to determine the most valuable subset of objects that can fit in a knapsack of fixed volume) as a typical representation of intellectual discovery problems. We found that our "markets system" performed better than the patent system.

In a patenting system, the first to discover the solution to a problem receives a prize in the form of exclusive property rights to the fruits from the discovery. Patents are generally viewed as a superior way to promote intellectual discovery because they provide strong incentives to invest in effort, the cost of which can be recuperated from the earnings generated by applications of the discovery.

The patent system has been criticized on various grounds. First, there is the obvious fair-

ness issue: Only the winner is compensated for effort. Second, ownership may become fragmented, which inhibits further discovery if such requires input from and coordination between the owners of prior inventions (*1*). Third, intellectual-property rights defy precise definition, and hence their scope will be subject to constant costly challenge (*2*). Lastly, patents imply monopoly rights and exploitation of monopoly rights leads to suboptimal production, sometimes no production whatsoever (*3*); the distortion becomes even more pronounced when the monopoly right covers downstream licensing (*4*).

Here, we propose an alternative, markets-based system to organize compensation of intellectual discovery. The key feature is to introduce markets for all the items that can potentially become crucial components of or inputs to implementations of a future discovery. Everyone can then own shares (securities) of these items, and these shares can be traded in an anonymous market before discovery. This way, agents who posit that a certain item is more likely to be part of a new discovery are induced to invest in it, while selling shares in items that show less potential. Incentives are indirect: Discoverers are compensated when they are the first to realize which items can solve an outstanding problem; compensation is in the form of share price increases; indeed, one can expect the value of items in useful discoveries to increase, whereas items not part of them will become cheaper. Inventors also have a strong incentive to reveal their discovery: in order for share prices to adjust in their favor as soon as possible. This should speed up development of applications based on new discoveries.

For example, concrete is made of several elements, such as gypsum, lime, water, etc. The Romans realized at one point that the addition of volcanic ash made concrete far more durable and waterproof (*5*). In a patent system, the person who first discovers (and files a patent) to add volcanic ash to concrete would be given monopoly rights to sell ash-based concrete. In the proposed markets-based system, all agents who discover that volcanic ash was a useful additive to concrete could take positions in shares whose value would increase with the price of volcanic ash. After announcement of the discovery (which the inventors would rush to make), the demand for volcanic ash would increase as use of the new additive becomes widespread, and hence the price of volcanic ash increases. The inventors are compensated indirectly for their discovery, through an increase in the value of their shareholding.

Our proposed "markets system" shares some commonalities with the "cost book system" in 19th century Cornwall that led to major advances in steam technology despite an absence of patent right filings. Key to the success of this historical system was the ability of "adventurers" (investors) to freely take positions in and trade shares of several mines, each experimenting with different technologies. Historians (*6*) have described how the cost book system not only channeled much-needed money to the most promising ideas but also fueled a period of extraordinary technological advance that was in sharp contrast with the decades before, when high royalties resulting from Watt and Boulton's patent stifled innovation. Adventurers bet directly on possible solutions (technologies). We instead propose that inventors trade claims in potential components of the best solution.

In a series of experiments, we compared the performance of a patentlike prize system against that of the proposed markets system. We used the knapsack problem (KP) as a typical representation of intellectual discovery problems. In the KP, the

[1]Bocconi University, 20136 Milan, Italy. [2]University of California Los Angeles, Los Angeles, CA 90095, USA. [3]California Institute of Technology (Caltech), Pasadena, CA 91125, USA, and Ecole Polytechnique Fédérale Lausanne, 1015 Lausanne, Switzerland.

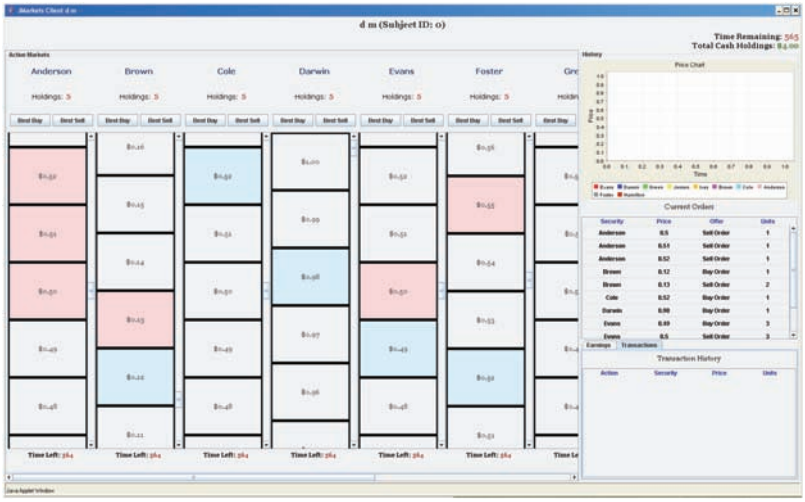
*To whom correspondence should be addressed. E-mail: debrah.meloso@unibocconi.it

A

Weight limit: 1500

	A	B	C	D	E	F	G	H	I	J
Values	37	72	106	32	45	71	23	44	85	62
Weights	50	820	700	46	220	530	107	180	435	360
Solution										

B



C

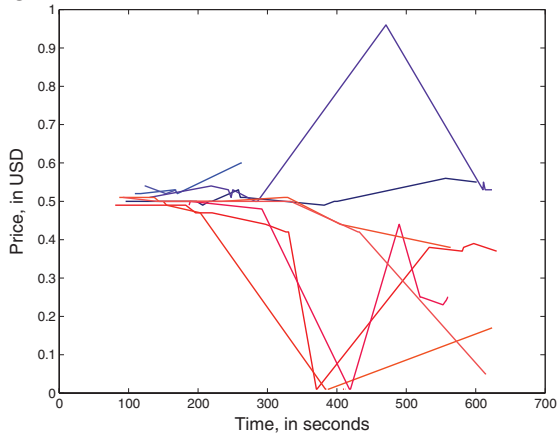


Fig. 1. (**A**) Presentation of an instance of the KP (instance IV), as seen by experiment participants. The first row lists the values of the items and the second one, their weights; the third row is left blank for participants to enter their suggested solution. The weight limit of the container is indicated on top of the table. The goal is to find the combination of the items that maximizes total value and is feasible (i.e., fits within the weight limit of the container). Objects are indivisible. (**B**) Snapshot of a typical trading interface in the markets treatment. Each column on the left-hand side provides price levels at which orders to buy (indicated in blue) and orders to sell (in red) can be entered. Orders are entered by a simple click of the mouse. Trades take place automatically when an order to buy is entered at a price above the best available sell order or when an order to sell is entered at a price below the best available buy order. Participants can see their current holdings above the columns. A list of the participant's current standing orders is provided to the right; above it is a chart with past transaction prices; below it is a summary of earnings in past trading rounds. (**C**) Evolution of transaction prices during a typical experimental period. Lines in blue connect trade prices for each of the securities that represent positions in in items (items that are part of the optimal solution); lines in red connect trade prices of securities corresponding to out items. Although initially equal, trade prices of in securities tend to increase, whereas those of out securities tend to decrease. USD indicates U.S. dollars.

challenge is to find the optimal combination of a number of components (items). Specifically (Fig. 1A), participants were asked to fill containers with at most 10 to 12 items; containers had a weight constraint, which meant that generally not all items fit; items had different values; the goal was to maximize total value of items that could fit.

At the heart of the KP is the task to devise the right combination of a potentially large number of inputs. In the economics literature (*7*, *8*), intellectual discovery is generally modeled as an exercise of information aggregation, in which all disparate pieces of information (signals) need to be merely added. The KP is different from information aggregation because it involves deciding which inputs to combine. This is a computationally far more demanding exercise (*9*). For instance, one can never be sure whether an item is in the optimal solution until one derives the entire optimal solution. We claim that the KP better reflects intellectual discovery because it too requires one to put things together in original ways rather than just adding things up. To quote (*4*), "Two half-baked ideas do not equal one fully-baked idea."

In our experiment, we used eight instances (variations) of the KP. The instances varied in degree of difficulty, as measured by an index based on Sahni's heuristics (*10*, *11*). This would enable us to determine whether the success of a particular system was uniform or applied only to a limited range of problems. In our prize system, we rewarded the first participant to discover the optimal solution with \$66 (all values are given in U.S. dollars). Although we imposed a time constraint (420 s), it was binding in only one instance (no one found the right solution within the allowed time). To best reflect real-world patent systems, discovery was immediately and publicly announced. However, the nature of the solution was not revealed, and participants could continue work on the KP instance at hand. After time elapsed, participants handed in their suggested solutions. However, only the person who first discovered the right solution was compensated.

In the markets system, participants were given an equal number of shares in each of the items of the particular KP, as well as cash. They could trade these shares in an anonymous, electronic exchange platform during a preset amount of time (840 s). The allowed time was double that of the prize system to compensate for the fact that subjects needed to perform two tasks: to solve the KP and to trade (to exploit the knowledge they gained from solving the KP). The platform was organized as a continuous double-sided open book (Fig. 1B), like most purely electronic stock markets in the world. The accumulation of orders generated the first transactions after about 100 s. Thereafter, trading remained brisk in virtually all markets (Fig. 1C). After markets closed, each share in an item that was in the optimal solution paid a liquidating dividend of \$1; shares corresponding to items not in the optimal solution expired worthless.

In the markets system, we arranged initial allocations such that we paid on average a total of \$60 to \$68 in dividends per instance, depending on the number of participants. These numbers were deliberately made as close as possible to the amount we paid to the single participant who discovered the right solution in the prize system (\$66) to ensure that incentives were strictly less in the markets system. In the markets system, the total payments were distributed across participants depending on the number of shares of each item they were holding at the close of the markets. As such, a participant who knew the optimal solution could only obtain the same compensation as in the prize system if she or he were able to buy all shares for items in the solution with the proceeds from selling her or his remaining shares, a highly unlikely situation. We also collected participants' suggested solutions after markets closed; we did not compensate for correctness of these suggestions to avoid introducing prizelike features.

In each experiment, participants solved four instances of the KP under the prize system and four instances under the markets system. We repeated the experiment four times, so that each instance was solved twice under both systems. In one session, there were 17 participants; in all other, 15 participants. The experiments had the approval of the Caltech Institute Review Board for the protection of human participants.

The correct solution was found under the markets system whenever this was the case under the prize system. Therefore, if the concern is to design a system that produces the optimal solution, the markets and prize systems are equivalent. In one important respect, however, the markets system outperformed: Significantly more participants reported the correct solution than under the prize system (Fig. 2A). For both systems, the fraction of participants who reported the correct solution declined with problem difficulty (Fig. 2B). The fraction may seem to decline faster for the markets treatment, but the difference in slopes was not significant. An outlier influenced the fits: Nobody ever solved the most difficult problem (difficulty = 6). It was solved in follow-up experiments [ran to check for robustness (*11*)], but only with the markets system, further corroborating its superiority.

In the prize system, only the first to find the optimal solution is compensated, which may discourage many from spending effort. In the markets system, everyone could be compensated in principle, which may be sufficient to explain why more participants find the optimal solution. Alternatively, prices may convey information that facilitates problem solving for participants who would never find the optimal knapsack on their own. Figure 3A shows that prices indeed do provide a potential channel of communication: Prices of shares of items that were part of the optimal knapsack ("in" items) tended to be higher than shares of items that were not part of it ("out" items); the mean transaction price of in items was significantly higher than that of out items ($P < 0.01$). Figures S1 to S8 show how prices of the in and out items diverge over time. It would have been interesting to explore the impact of this divergence on the ability and

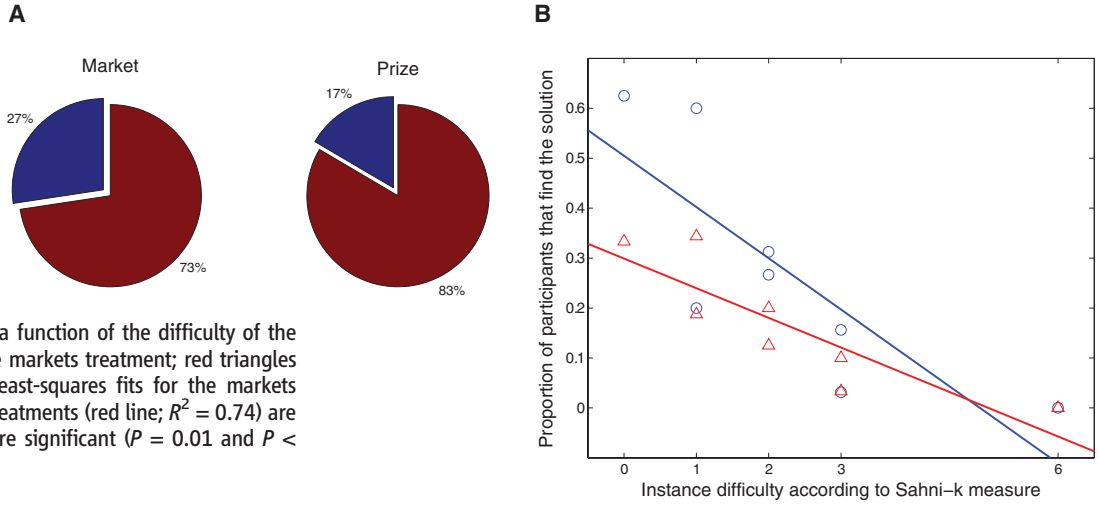


Fig. 2. (**A**) Overall proportion of participants that reported the correct solution (blue) and the incorrect one (red) for the markets and the prize treatments. The proportions are significantly different across treatments (χ^2 statistic of equality of two binomial distributions = 6.28; $P = 0.01$). (**B**) Proportion of participants who reported the correct solution of each instance of KP as a function of the difficulty of the problem. Blue circles indicate the markets treatment; red triangles represent the prize treatment. Least-squares fits for the markets (blue line; $R^2 = 0.64$) and price treatments (red line; $R^2 = 0.74$) are shown. Slopes of the two lines are significant ($P = 0.01$ and $P < 0.01$, respectively).

speed with which participants discover the optimal solution. However, this would have required us to introduce incentives for participants to submit suggested solutions every minute or so. The resulting side payments would have blurred the comparison with the prize system.

In the markets system, subjects are compensated indirectly for superior problem solving skills: Those who figure out the right solution buy "in" securities and sell "out" securities, to eventually be paid higher total dividends. As a result, we expect to find more divergence in final holdings of shares than in initial holdings. Because we started all participants with equal number of shares in each of the items (namely, five), any dispersion in final holdings would be consistent with this expectation. Final holdings for both in and out securities displayed substantial dispersion (Fig. 3B). Final holdings of out are more frequently below initial holdings than those of in. Because total frequencies have to add up to 100%, the higher frequencies of low final holdings of out needed to be offset somehow. Figure 3B shows that the offset came from (i) lower frequency of final holdings equal to initial endowments and (ii) higher frequencies of extremely high final holdings (the percentage outcomes for out beyond 15 was almost four times larger than for in: 1.66% against 0.43%; the maximum for out holdings, at 37 units, was substantially above that for in holdings, at 28).

Despite the absence of direct incentives, the experimental results indicate that our markets system performs better than a patentlike prize system. Referring to the five criticisms of the patent system, (i) our market system is fairer in the sense that anyone who finds the optimal solution can potentially win by taking the right positions in the markets; (ii) it does not lead to fragmented ownership of intellectual property rights because the optimal solution is in the public domain; (iii) defining the scope of intellectual property rights is not a problem because they do not exist in the first place; (iv) no monopoly rights result from finding the optimal solution; and likewise (v) downstream licensing is not needed: anyone can use the optimal solution without having to directly compensate the discoverer.

It may be argued that more people in the market system discovering the same (solution) is inefficient. What ultimately counts, however, is total cost to obtain the discovery, which in our experiments was identical across the two systems. In addition, repeated discovery should be beneficial in situations where there is a discovery learning curve (i.e., when successful future discovery depends critically on full experience with the process that led to past discoveries).

The patent system has been motivated by the fact that the information embedded in a discovery is a nonexcludable and nonrival good. Economic theory argues that markets for such goods will fail (*8*). Our data suggest that this claim is false. The results underscore the necessity of experimentation when deciding between public institutions.

In certain cases, real-world implementation of our markets system already exists. Take, for instance, fuel cell technology. Successful development of this technology will depend, among others, on finding the right catalyst. This is putting tremendous pressure on prices of all catalysts, starting with platinum. Markets in platinum exist already. Those who think that platinum will provide the best catalyst will profit, if their belief is right, from buying platinum shares. New markets in competing catalysts, such as carbon silk, are bound to emerge. Positions in the various markets implement bets on different fuel cell technologies. These positions could go both ways: If an inventor thinks platinum is not the

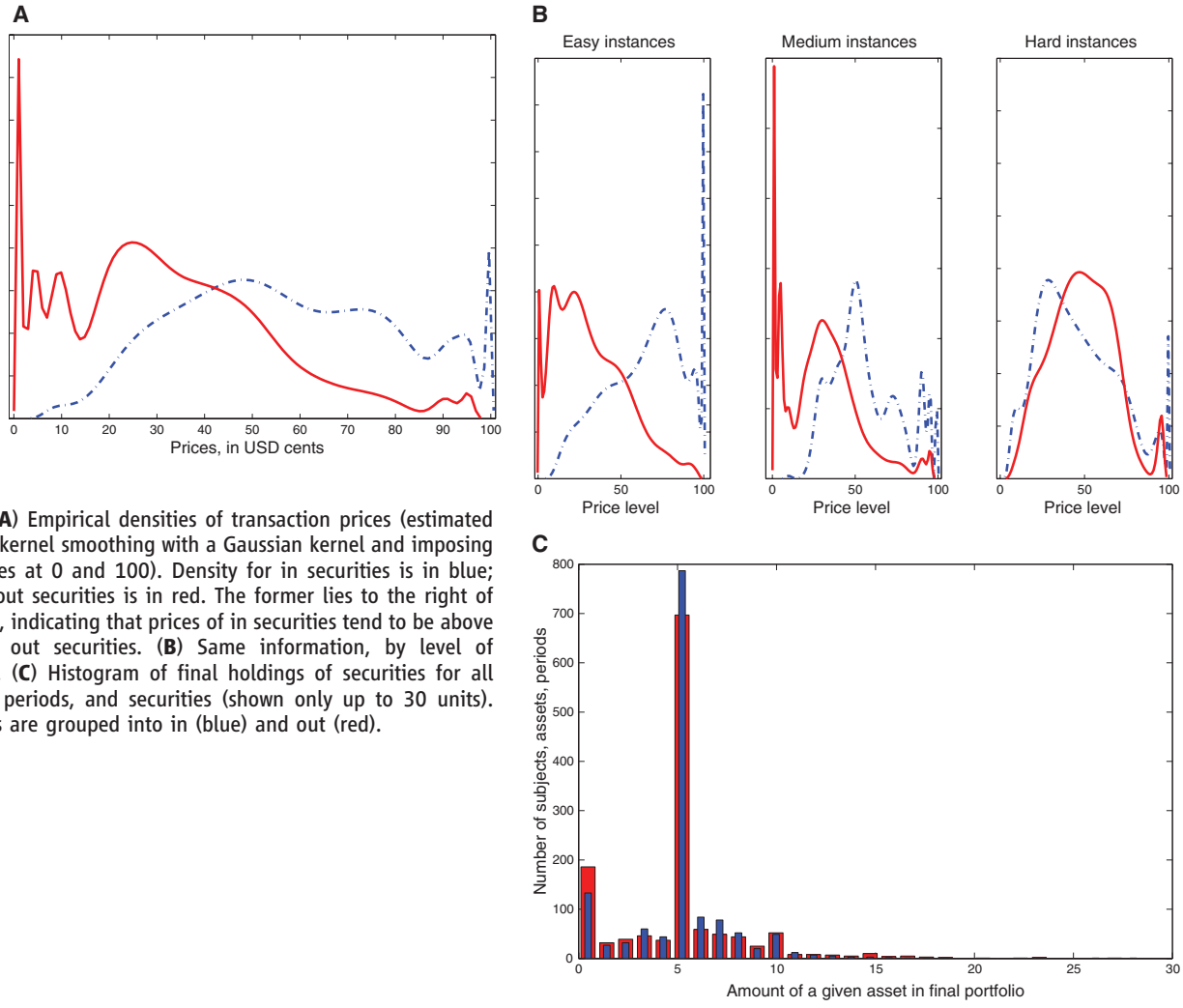


Fig. 3. (**A**) Empirical densities of transaction prices (estimated by using kernel smoothing with a Gaussian kernel and imposing boundaries at 0 and 100). Density for in securities is in blue; that for out securities is in red. The former lies to the right of the latter, indicating that prices of in securities tend to be above those of out securities. (**B**) Same information, by level of difficulty. (**C**) Histogram of final holdings of securities for all subjects, periods, and securities (shown only up to 30 units). Securities are grouped into in (blue) and out (red).

answer, then she or he could short-sell this metal to profit from the expected reduction of the platinum price. Of course, the prices of catalysts are affected not only by advances in fuel cell technology but also by usage for other purposes. Still, financial engineers nowadays are capable of tailoring positions to isolate one specific source of risk. The fuel cell example illustrates how inventors at present may actually be able to profit from a discovery twice: through the monopoly granted by the patent and through positions in markets for catalysts. We propose to eliminate the former, leaving inventors only with the incentives provided by the latter. Our experiments show that this is sufficient to promote intellectual discovery.

In our experiments, subjects were paid as a function of the very best solution. We consider this to be an idealized situation afforded by the experimental setting, which allows for a clear interpretation of performance numbers. In practice, inventors will be compensated for second-best solutions as long as the best one has not been attained yet. This is precisely what happened in 19th century Cornwall: Engineers never really found the best way of building steam engines (we know this because many improvements were made afterward), but they were rewarded for providing better solutions.

The traditional patent system helps people with an idea even if they have no resources, because those who have the resources (usually venture capitalists) provide inventors with cash in return for a share in the intellectual property rights. In our markets system, resources would be generated in a similar way. The people with the ideas but no money could approach investors (such as fund managers) and inform them. These investors could then take positions to exploit expected changes in valuations from adoption of the new technology. They should be eager to pay for the information, and this payment will provide the inventors with the necessary cash for development.

We do not claim that our markets system will work under all circumstances. We envisage that it could replace the patent system whenever a technology builds on goods and services with economic rents, which means that their cost of provision is below market value. Such rents obtain for a variety of reasons. One is limited supply (volcanic ash in the concrete example, platinum in the fuel-cell case, artemisinin in the case of medication against drug-resistant malaria, or the claims to the items in the knapsacks in our experiments). Other important reasons are first-mover advantage (this seems to have been the case with steam engine technology in the mines of Cornwall in the mid-19th century) and lead in the learning curve [studied extensively in the economics literature; see (*7*)].

The success of our markets system relies on willingness to trade; without trading, those who make progress toward finding the optimal solution cannot exploit their acquired knowledge. In our setting, participants never know whether they have the optimal solution (because they would need much more time to check all possible solutions). Thus, there is always the possibility that one is trading with a counterparty who knows better. Why, then, would participants trade? We conjecture that they trade because they tend to be too confident that they are closer to solving the problem than others. Overconfidence is indeed an important human trait, best illustrated by the fact that more than 50% of people usually think they are better than the median (*12*). Other, nonpecuniary incentives for trade (such as a taste for being "right") may play a role. Further research is needed to identify the origin of the success of the markets system in promoting intellectual discovery.

Our proposal relies on anonymous, two-way markets. Until recently, setting up new markets required time. Modern technology, however, has enabled quick design, ready deployment, and low-cost management of markets. With the open-source software we developed for our experiments, jMarkets (*13*), setting up and launching of (online) markets can be done in a matter of hours.

Our experimental findings suggest that the patent system is not a universally superior way to incentivize intellectual discovery. We propose a markets-based system that we found to work better. Its main features are that the compensation for inventions is shared and that, because discovery remains in the public domain, it avoids both distortion in the provision of newly invented products and stifling of future discovery.

References and Notes

1. M. A. Heller, R. S. Eisenberg, *Science* **280**, 698 (1998).
2. A. B. Jaffe, *Res. Policy* **29**, 531 (2000).
3. R. J. Gilbert, D. M. G. Newbery, *Am. Econ. Rev.* **72**, 514 (1982).
4. M. Boldrin, D. Levine, *Am. Econ. Rev.* **92**, 209 (2002).
5. R. Siddall, *Geol. Soc. London Spec. Publ.* **171**, 339 (2000).
6. A. Nuvolari, *Camb. J. Econ.* **28**, 347 (2004).
7. K. Arrow, *Rev. Econ. Stud.* **29**, 155 (1962).
8. N. Gallini, S. Scotchmer, in *Innovation Policy and the Economy*, A. Jaffe, J. Lerner, S. Stern, Eds. (MIT Press, Cambridge, MA, 2002), vol. 2, pp. 51–78.
9. H. Kellerer, U. Pferschy, D. Pisinger, *Knapsack Problems* (Springer-Verlag, Heidelberg, 2004).
10. S. Sahni, *J. Assoc. Comput. Mach.* **22**, 115 (1975).
11. Materials and methods are available as supporting material on *Science* Online.
12. B. Biais, D. Hilton, K. Mazurier, S. Pouget, *Rev. Econ. Stud.* **72**, 287 (2005).
13. http://jmarkets.ssel.caltech.edu.

14. This study was partly funded by the U.S. NSF (grants SES-0616431 and SES-0317715) and the Swiss Finance Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5919/1335/DC1
Materials and Methods
SOM Text
Figs. S1 to S11
Tables S1 to S5
References

3 April 2008; accepted 5 January 2009
10.1126/science.1158624

Molecular and Evolutionary History of Melanism in North American Gray Wolves

Tovi M. Anderson,[1] Bridgett M. vonHoldt,[2] Sophie I. Candille,[1] Marco Musiani,[3] Claudia Greco,[4] Daniel R. Stahler,[2,5] Douglas W. Smith,[5] Badri Padhukasahasram,[6] Ettore Randi,[4] Jennifer A. Leonard,[7] Carlos D. Bustamante,[6] Elaine A. Ostrander,[8] Hua Tang,[1] Robert K. Wayne,[2] Gregory S. Barsh[1]*

Morphological diversity within closely related species is an essential aspect of evolution and adaptation. Mutations in the *Melanocortin 1 receptor* (*Mc1r*) gene contribute to pigmentary diversity in natural populations of fish, birds, and many mammals. However, melanism in the gray wolf, *Canis lupus*, is caused by a different melanocortin pathway component, the *K* locus, that encodes a beta-defensin protein that acts as an alternative ligand for Mc1r. We show that the melanistic *K* locus mutation in North American wolves derives from past hybridization with domestic dogs, has risen to high frequency in forested habitats, and exhibits a molecular signature of positive selection. The same mutation also causes melanism in the coyote, *Canis latrans*, and in Italian gray wolves, and hence our results demonstrate how traits selected in domesticated species can influence the morphological diversity of their wild relatives.

The correspondence between coat color and habitat is often attributed to natural selection, but rarely is supporting evidence provided at the molecular level. In North American gray wolves, coat color frequencies differ between wolves of forested and open habitats throughout western North America (*1*), including Denali National Park (*2*) and the Kenai Peninsula in Alaska (*3*), and much of the Canadian Arctic (*4*, *5*). These differences are especially dramatic between wolves of the high tundra that are migratory and follow barren-ground caribou to their breeding areas, and wolves that are year-round residents in the neighboring boreal forest and hunt nonmigratory prey.

Dark-colored wolves are extremely rare in the tundra but increase in frequency along a southwest cline toward forested areas (Fig. 1A). The potential selective value of dark versus light coat

[1]Departments of Genetics and Pediatrics, Stanford University, Stanford, CA 94305, USA. [2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 91302, USA. [3]Faculty of Environmental Design, University of Calgary, Calgary, AB T2N 1N4, Canada. [4]Istituto Nazionale per la Fauna Selvatica, 40064 Ozzano Emilia (BO), Italy. [5]Yellowstone Center for Resources, National Park Service, Yellowstone National Park, WY 82190, USA. [6]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA. [7]Department of Evolutionary Biology, Uppsala University, 75236 Uppsala, Sweden. [8]National Human Genome Research Institute, Bethesda, MD 20892, USA.

*To whom correspondence should be addressed. E-mail: gbarsh@stanford.edu

Table 1. Distribution of *CBD103* alleles in wolves and coyotes. N/A, not applicable.

Animal and location		Phenotype†		
		White	Gray	Black
Forest wolves*	Total no.	12	2	7
	No. carrying K^B	0	1	7
Tundra/taiga wolves*	Total no.	10	8	2
	No. carrying K^B	0	5	2
Yellowstone wolves	Total no.	0	120	104
	No. carrying K^B	N/A	0	102
Coyotes‡	Total no.	0	61	6
	No. carrying K^B	N/A	0	6

*Forest and tundra/taiga wolves are from the Canadian Arctic (Fig. 1A). The overall frequency of dark (gray or black) wolves is 62 and 7% in the forest and tundra/taiga, respectively (*4*), and the genotype distributions shown do not represent population-based frequencies. All forest and tundra/taiga wolves carrying K^B were K^B/k^y; in the Yellowstone population, 10 were K^B/K^B and 92 were K^B/k^y. †This categorical designation of phenotypes, as defined at sample collection, does not fully capture the spectrum of normal coat color variation as indicated in Fig. 1B. ‡Gray coyotes surveyed were from Nebraska (30) or West Virginia (30); black coyotes were from Minnesota (2) or West Virginia (4).

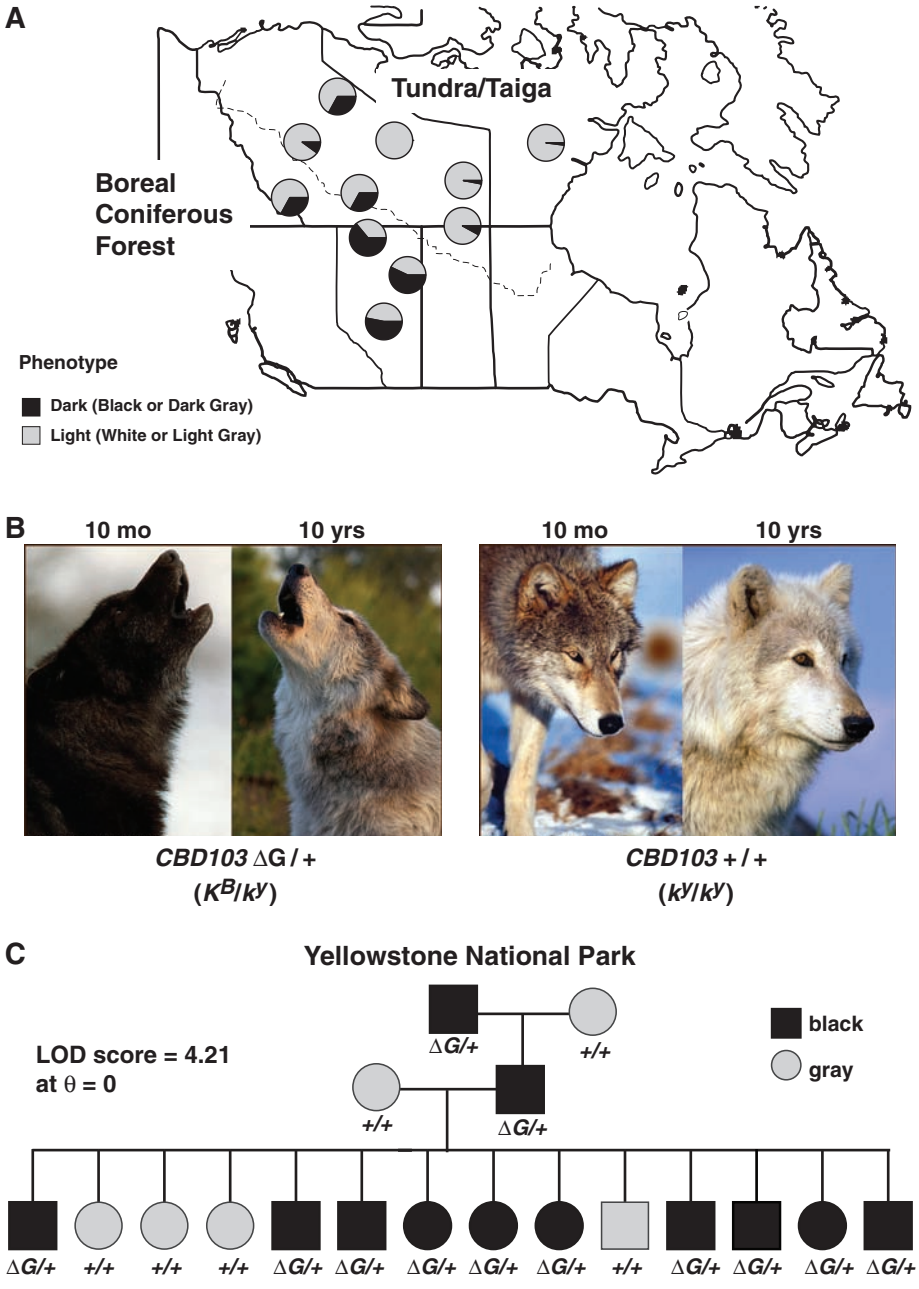


Fig. 1. Distribution of melanism and *K* locus genotypes in North American gray wolves. (**A**) Location and coat color phenotype of Canadian samples used here and as described (*4*). (**B**) Age-related graying and the associated difficulty of inferring genotype from phenotype in gray animals. Each pair of photos shows the same individual at different ages (10 months and 10 years) and documents an increasingly gray appearance at 10 years, reflecting the dilution of eumelanin in the K^B/k^y individual (left pair of images) and dilution of both eumelanin and pheomelanin in the k^y/k^y individual (right pair of images). [Images courtesy of Monty Sloan, Wolf Park, Battle Ground, Indiana] (**C**) Co-segregation of K^B and black coat color in a three-generation pedigree from the Leopold pack in Yellowstone National Park (*17*). ΔG indicates the dominant K^B allele, whereas + indicates the wild-type allele, k^y.

color has been suggested to include concealment during predation and/or indirect effects due to pleiotropy, but remains unresolved because the underlying gene(s) have not been identified (*5*–*7*).

In many vertebrates, natural pigmentary variation is controlled by the agouti– melanocortin 1 receptor (Mc1r) pathway, a ligand receptor pair that modulates the amount and type of pigment—red/yellow pheomelanin or brown/black eumelanin—produced by melanocytes in skin, hair, or feathers. Gain-of-function *Mc1r* mutations are well-recognized causes of melanism in many domestic and laboratory animal species (*8*, *9*), as well as in several natural populations of birds (*10*), rodents (*11*, *12*), and canids (*13*). Recently, we found that pigment type-switching in domestic dogs involves an additional component of the melanocortin pathway, the *K* locus, which encodes a beta-defensin protein, CBD103 (*14*, *15*).

Coat color in Canadian wolves is genetically complex, with phenotypes ranging from white to gray to black, and is also confounded by an independent effect of graying with age (Fig. 1B). However, in Yellowstone National Park, where a small number of founder animals from Canada were recently reintroduced (*16*, *17*), gray and black coat colors segregate as a Mendelian trait. We surveyed molecular variation in *Agouti*, *Mc1r*, and *CBD103* in wolves from North America and identified several *Mc1r* and *Agouti* polymorphisms. However, none of these were predicted to affect gene function and did not associate with black coat color (table S1). In contrast, in a 14-member, three-generation kindred from Yellowstone, we observed complete co-segregation between black coat color and markers at the *K* locus [logarithm of the odds ratio for linkage (lod) score = 4.21 at the maximum likelihood estimate of recombination fraction (θ) = 0, Fig. 1C], which is unlinked and lies on a different chromosome from *Agouti* and *Mc1r*.

In dogs, the ancestral *CBD103* allele (k^y) confers normal *Agouti* and *Mc1r* gene action, whereas a 3–base pair (bp) deletion ($CBD103^{\Delta G23}$ or K^B) suppresses *Agouti* gene action, leading to dominant inheritance of a black coat (*14*, *15*). We observed the same 3-bp deletion in 102 out of 104 black-colored wolves from Yellowstone and 9 out of 9 from the Canadian Arctic. Conversely, $CBD103^{\Delta G23}$ was absent from 120 of 120 gray-colored wolves from Yellowstone and from 22 of 22 white-colored wolves from the Canadian Arctic (Table 1). We also found $CBD103^{\Delta G23}$ in 6 of 10 gray-colored wolves from the Canadian Arctic, suggesting that gray coat color can result either from the absence of $CBD103^{\Delta G23}$ and a modified agouti phenotype (in which individual hairs contain both cream-colored pheomelanin and dark eumelanin) or from secondary factors such as age that dilute the pigmentation of hairs that contain only eumelanin. [Additional genealogy studies of the

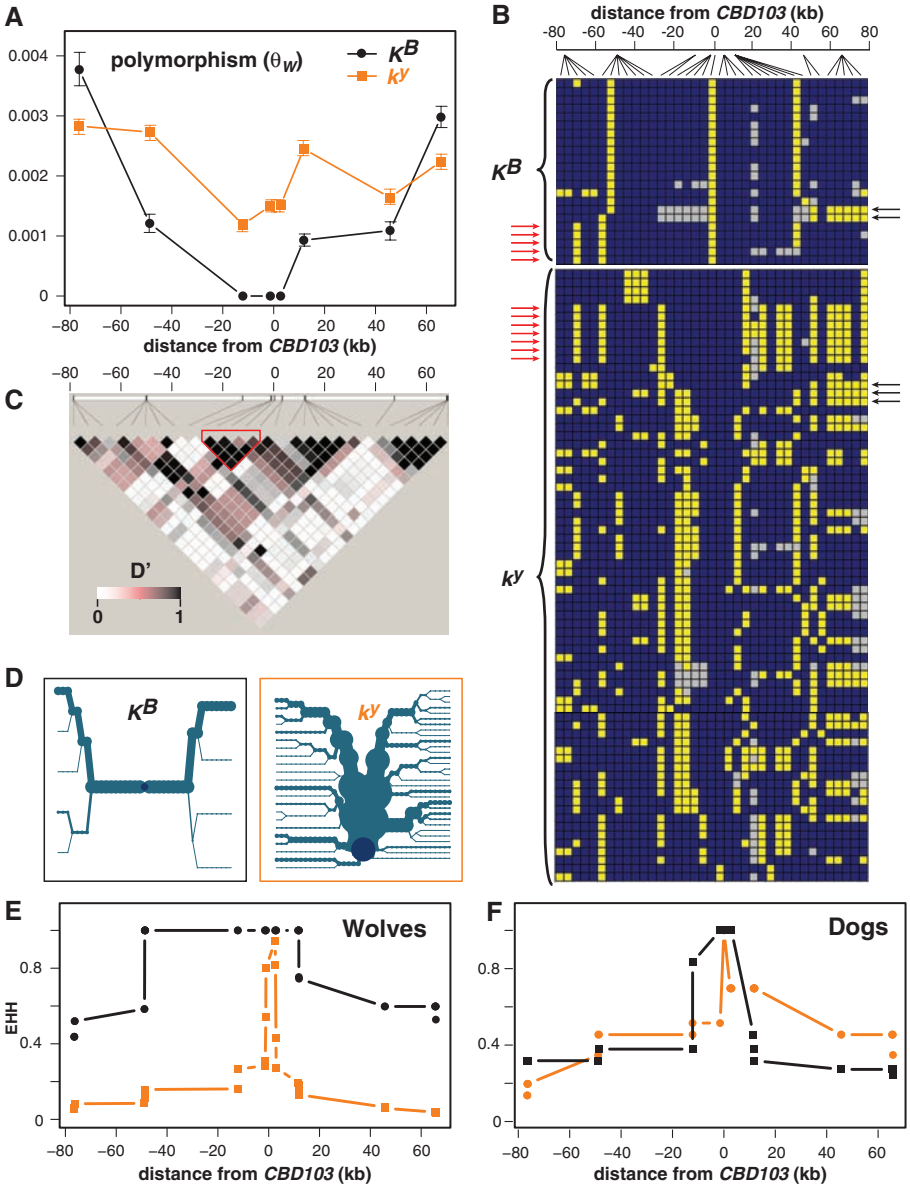


Fig. 2. Polymorphism and haplotype structure of the *K* locus in North American gray wolves [(A) to (E), 1 K^B/K^B, 20 K^B/k^y, and 26 k^y/k^y] and domestic dogs [(F), 6 K^B/K^B and 6 k^y/k^y]. (**A**) Polymorphism (θ_W, ±SD) as a function of distance from *CBD103*. (**B**) Wolf haplotype structure was inferred on the basis of 36 SNPs; each row represents a K^B- or k^y-bearing chromosome; blue and yellow squares represent the major and minor alleles, respectively; and the gray squares represent missing data. Red and black arrows indicate examples of haplotypes likely to represent historical recombination between K^B- and k^y-bearing chromosomes at the 5′ and 3′ ends of the locus, respectively. (**C**) Pairwise LD values (expressed as D') for all wolf chromosomes; the red outline indicates a core region (as in Fig. 3) unlikely to have undergone historical recombination. (**D**) Haplotype bifurcation diagrams for K^B- or k^y-bearing chromosomes, in which the central dark blue dot represents *CBD103*, branches represent haplotype divergence, and the thickness of the lines is proportional to the number of chromosomes. (**E** and **F**) EHH for K^B- or k^y-bearing chromosomes in wolves (E) and dogs (F) as a function of distance from $CBD103^{\Delta G23}$.

Yellowstone population (*17*) together with the paucity of *Mc1r* variation in wolves (table S1) suggest that black coat color reported for the two k^y/k^y Yellowstone wolves is likely to reflect phenotypic ambiguity or misclassification at the time of sampling.] Allele frequencies for $CBD103^{\Delta G23}$ in tundra and forest wolves overall were estimated at 0.02 and 0.19, corresponding to phenotype frequencies of 2 to 33% and 33 to 64% for dark wolves in tundra and forest populations, respectively (Fig. 1A) (*4*).

To investigate the evolutionary history of the melanistic *K* allele, we sequenced eight single-copy noncoding segments distributed across an ~150-kb region centered on *CBD103* in 32 Arctic and 15 unrelated Yellowstone wolves, as well as in 12 domestic dogs: 6 k^y/k^y (akita, basenji, boxer, bulldog, Doberman pinscher, and great dane) and 6 K^B/K^B (curly-coated retriever, Dalmatian, great dane, Labrador retriever, poodle, and Portuguese water dog). We identified 52 biallelic polymorphisms across all canids (36 in wolves) and estimated haplotype structure (tables S3 and S4, Fig. 2B, and fig. S2). The rate of polymorphism among all wolf amplicons was one single-nucleotide polymorphism (SNP) per 510 bp (Watterson's estimator, $\theta_W = 1.96 \times 10^{-3}$), which is similar to genome-wide measurements of polymorphism between the boxer and the gray wolf (1 out of 580 bp) and the coyote (1 out of 420 bp) (*18*). However, partitioning our data according to *K* locus genotype and proximity to *CBD103* revealed little or no polymorphism among K^B-bearing chromosomes close to *CBD103*, rising to levels at or above those observed in k^y-bearing chromosomes in the 75 kb spanning either side of the locus (Fig. 2A). This pattern, and the analogous one for nucleotide diversity (π, fig. S1), is also reflected in a significant difference in haplotype diversity between K^B (8 unique of 22 total) and k^y (59 unique of 72 total) chromosomes ($\chi^2 = 14.2$, $P < 0.001$). Together with the correlations between coat color and habitat (*2*–*5*), the combination of low diversity and high frequency suggests that K^B has been under positive selection in North American forest wolves.

Overall, the patterns of linkage disequilibrium (LD) across 150 kb surrounding the *K* locus were similar to comparisons between different breeds of domestic dogs (*18*), with relatively small haplotype blocks, including an ~4-kb *CBD103* core region within which there is no evidence for historical recombination (Fig. 2C). However, different evolutionary histories for the Arctic wolf K^B and k^y alleles were apparent when the SNP patterns (Fig. 2B) were depicted as haplotype bifurcation diagrams (Fig. 2D), which highlight a central region of ~60 kb devoid of polymorphism among wolf K^B haplotypes. This characteristic, and the corresponding difference between K^B and k^y chromosomes, were represented quantitatively by the extended haplotype homozygosity (EHH) statistic (*19*), which is the empirical probability that two chromosomes chosen at random remain identical at progressively increasing distances from *CBD103*. As depicted in Fig. 2, E and F, the distribution of EHH was considerably broader for K^B as compared to k^y chromosomes in wolves, whereas the

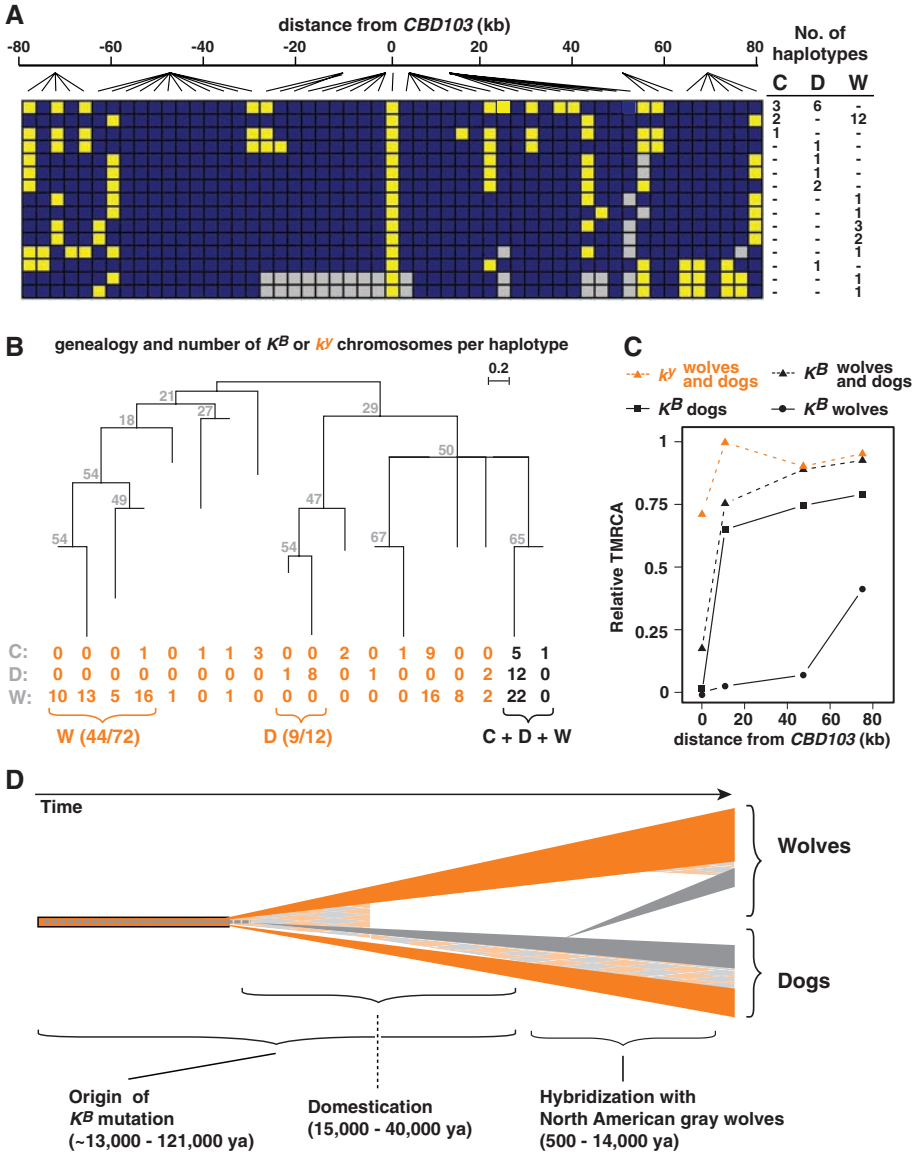


Fig. 3. Evolutionary relationships and history of the *K* locus in canids. (**A**) K^B haplotype structure in wolflike canids based on genotypes defined by 52 SNPs. Each row represents a K^B-bearing haplotype found in coyotes (C), dogs (D), or wolves (W) listed with their respective frequencies on the right and colored as in Fig. 2B. (**B**) Inferred genealogical relationships of the core region (Fig. 2C) haplotypes (with bootstrap values from 500 replicates shown next to branches). Each branch represents 1 of 18 different haplotypes, with the number of chromosomes for each haplotype indicated underneath according to species. (**C**) TMRCA estimates for indicated chromosome subsets calculated according to a molecular clock (*22*) and expressed as a fraction of the divergence time for all wolflike canids. Individual points represent sets of chromosome segments whose relative TMRCA increases as a function of distance from *CBD103*, presumably due to ancient hybridization and recombination. (**D**) Timeline scenario for *K* locus evolution in dogs and wolves, in which ancestral k^y chromosomes are indicated in orange, derivative K^B chromosomes in gray, and recombinant chromosomes as an orange-gray checkered pattern. The k^y-to-K^B mutation may have overlapped or even predated domestication, but the introgression of K^B into North American gray wolves is more recent.

distributions were nearly identical for K^B as compared to k^y chromosomes in dogs. Together with additional analyses of genome-wide SNP data [supporting online material (SOM) text and fig. S3], these observations suggest that K^B has risen to high frequency by a selective sweep.

As in black dogs and melanistic wolves, $CBD103^{\Delta G23}$ was associated with coat color in 67 coyotes (6 black and 61 gray, Table 1 and table S2). These findings suggest three possible evolutionary histories. First, the 3-bp deletion may be relatively old, having occurred in a canid ancestor more than 1 million years ago before the divergence of coyotes from wolves. Second, the 3-bp deletion may have occurred more recently in one of the species, followed by introgression into the others. Finally, the 3-bp deletion may represent a mutational hotspot, having recurred independently in coyotes, wolves, and dogs. To distinguish among these possibilities, we ascertained and compared coyote haplotypes (6 K^B and 18 k^y) with those from the North American wolf and dog.

The pattern of haplotype diversity for all three canids was similar to that observed in wolves alone and showed significantly less diversity among K^B (15 unique of 40 total) relative to k^y (66 unique of 102 total) chromosomes ($\chi^2 = 9.7$, $P = 0.003$). Of the 15 unique K^B haplotypes, 1 haplotype was observed in three coyotes and six dogs, and a second haplotype was observed in two coyotes and 12 wolves (Fig. 3A). However, none of the 66 unique k^y haplotypes were observed in more than one species (fig. S2).

Reconstruction of a phylogenetic network for the entire 150-kb region is complicated by historical recombination between extant K^B and k^y chromosomes (arrows in Fig. 2B) and the lack of a suitable approach for inferring accurate gene genealogies in the presence of recombination (*20*). However, by focusing on the 4-kb *CBD103* core region (Fig. 2C), a simple neighbor-joining tree was constructed for 18 core region haplotypes representing 142 (94 wolf, 24 dog, and 24 coyote) chromosomes (Fig. 3B). In this tree, all the K^B chromosomes define a 2-haplotype cluster, whereas the remaining 16 haplotypes (which represent all the k^y chromosomes) are more dispersed. Furthermore, many of the k^y chromosomes cluster by species (9 out of 12 of the dogs and 44 out of 72 of the wolves), unlike the K^B chromosomes. This contrasting phylogenetic pattern suggests that the K^B mutation occurred in a single species and was later distributed among dogs, wolves, and coyotes by interspecific hybridization. [The 24 k^y haplotypes from coyotes are no closer to each other than to k^y haplotypes from wolves or dogs (Fig. 3B), which is consistent with their history of hybridization with other canids (*21*)].

To gain additional insight into how *K* locus variation in dogs and wolves arose, we estimated coalescent time to the most recent common ancestor (TMRCA) as a function of cumulative distance from *CBD103* for k^y and K^B chromosomes from wolves, dogs, and both groups together. We applied a molecular clock approach to sequencing data from individual amplicons across the entire 150-kb region (Fig. 2), which assumes that mutations occur at the same constant rate at all sites in wolves and dogs and integrates the effects of both recombination and demography (*22*). Close to *CBD103*, TMRCA estimates were near zero for all K^B subsets (Fig. 3C) because there is little or no polymorphism in this region (Fig. 3A). However, at greater distances from *CBD103* (10 to 50 kb), estimates for dog chromosomes are similar to those of dog and wolf chromosomes considered together, regardless of genotype. This suggests that K^B in dogs is sufficiently old to have undergone extensive recombination with k^y chromosomes, and that the recombination history includes hybridization between dogs and wolves. However, in the same 10- to 50-kb range, TMRCA estimates for wolf K^B chromosomes were considerably less than those from dog K^B chromosomes (or from dog and wolf K^B chromosomes considered together), suggesting that K^B was introduced into North American wolves from dogs, not vice versa.

Introgression of K^B from dogs into North American wolves is also supported by geographical and ecological considerations. K^B is widely distributed among domestic dogs, including ancient breeds originating in Asia and Africa. In wolves, however, melanism has been reported outside North America only in Italy, where it is associated with molecular and/or morphologic evidence of recent hybridization with free-ranging dogs (*23*). Indeed, we also examined 22 samples from the Italian Apennines and observed K^B in six of seven black "wolves" (including one previously classified to be a dog-wolf hybrid) but 0 of 15 gray wolves. In contrast, genome-wide SNP analysis of 10 K^B/k^y and 10 k^y/k^y North American wolves showed no evidence for recent dog-wolf hybridization (SOM text and fig. S3B).

The dog was domesticated between 15,000 and 40,000 years ago in East Asia from gray wolves (*24*, *25*), and we estimate that K^B is at least 46,886 years old (95% confidence limit: 12,779 to 121,182 years); therefore, we cannot distinguish whether K^B arose before or after domestication. However, if K^B arose in Old World wolves before domestication, our data indicate that it must have been lost from the gene pool and reacquired in North America, perhaps from Native American dogs that accompanied humans across the Bering Strait 12,000 to 14,000 years ago (*26*) (Fig. 3D).

The wolf in the United States faces grave threats, in some cases by eradication, and in others by hybridization, such as in the Great Lakes region (*27*). However, apparent selection for the K^B locus in North American gray wolves shows how genetic diversity—preserved by humans in domestic dogs—may flourish in wild wolf populations. As the available tundra habitat declines because of development and/or global warming, the frequency of the K^B mutation may increase further in northern latitudes. Thus, the introduction of genetic diversity into a natural population from a mutation originally selected in domesticated animals may, ironically, provide a mechanism for that population to adapt to a changing environment. Interspecific hybridization has been widely observed between other domesticated species of animals and plants (*28*–*30*). Our results imply that variants that appear under domestication can be viable in the wild and enrich the genetic legacy of natural populations.

References and Notes

1. P. S. Gipson *et al.*, *Wildl. Soc. Bull.* **30**, 821 (2002).
2. L. D. Mech, L. G. Adams, T. J. Meier, J. W. Burch, B. W. Dale, *The Wolves of Denali* (Univ. of Minnesota Press, Minneapolis, MN, 1998).
3. R. O. Peterson, J. D. Wollington, T. N. Bailey, *Wildl. Monogr.* **88**, 3 (1984).
4. M. Musiani *et al.*, *Mol. Ecol.* **16**, 4149 (2007).
5. P. Jolicoeur, *Evolution* **13**, 283 (1959).
6. M. E. N. Majerus, *Melanism: Evolution in Action* (Oxford Univ. Press, Oxford, 1998).
7. A. L. Ducrest, L. Keller, A. Roulin, *Trends Ecol. Evol.* **23**, 502 (2008).
8. H. Klungland, D. I. Vage, *Ann. N. Y. Acad. Sci.* **994**, 331 (2003).
9. L. Andersson, *Ann. N. Y. Acad. Sci.* **994**, 313 (2003).
10. N. I. Mundy *et al.*, *Science* **303**, 1870 (2004).
11. H. E. Hoekstra, R. J. Hirschmann, R. A. Bundey, P. A. Insel, J. P. Crossland, *Science* **313**, 101 (2006).
12. M. W. Nachman, H. E. Hoekstra, S. L. D'Agostino, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5268 (2003).
13. D. I. Vage *et al.*, *Nat. Genet.* **15**, 311 (1997).
14. J. A. Kerns *et al.*, *Genetics* **176**, 1679 (2007).
15. S. I. Candille *et al.*, *Science* **318**, 1418 (2007).
16. E. E. Bangs, S. Fritts, *Wildl. Soc. Bull.* **24**, 402 (1996).
17. B. M. vonHoldt *et al.*, *Mol. Ecol.* **17**, 252 (2008).
18. K. Lindblad-Toh *et al.*, *Nature* **438**, 803 (2005).
19. P. C. Sabeti *et al.*, *Nature* **419**, 832 (2002).
20. S. M. Woolley, D. Posada, K. A. Crandall, *PLoS ONE* **3**, e1913 (2008).
21. M. S. Roy, E. Geffen, D. Smith, E. A. Ostrander, R. K. Wayne, *Mol. Biol. Evol.* **11**, 553 (1994).
22. H. Tang, D. O. Siegmund, P. Shen, P. J. Oefner, M. W. Feldman, *Genetics* **161**, 447 (2002).
23. E. Randi, V. Lucchini, *Conserv. Genet.* **3**, 29 (2002).
24. C. Vila *et al.*, *Science* **276**, 1687 (1997).
25. P. Savolainen, Y. P. Zhang, J. Luo, J. Lundeberg, T. Leitner, *Science* **298**, 1610 (2002).
26. J. A. Leonard *et al.*, *Science* **298**, 1613 (2002).
27. J. A. Leonard, R. K. Wayne, *Biol. Lett.* **4**, 95 (2008).
28. R. Lecis *et al.*, *Mol. Ecol.* **15**, 119 (2006).
29. N. Halbert, J. Derr, *J. Hered.* **98**, 1 (2007).
30. N. Ellstrand, H. Prentice, J. Hancock, *Annu. Rev. Ecol. Syst.* **30**, 539 (1999).

31. Supported by grants from NIH (G.S.B.), NSF (R.K.W., D.R.S., and D.W.S.), and the Swedish Research Council (J. A.L.). We are grateful to H. Chen and S. Schmutz for advice, to H. Manuel for technical assistance, and to members of the U.S. Department of Agriculture Wildlife Services and private citizens for assistance with sample collection. Sequences generated in this study are deposited in GenBank under accession numbers FJ609634 to FJ609641.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1165448/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
Tables S1 to S5
References

3 September 2008; accepted 15 January 2009
Published online 5 February 2009;
10.1126/science.1165448
Include this information when citing this paper.

Drought Sensitivity of the Amazon Rainforest

Oliver L. Phillips,[1]* Luiz E. O. C. Aragão,[2] Simon L. Lewis,[1] Joshua B. Fisher,[2] Jon Lloyd,[1] Gabriela López-González,[1] Yadvinder Malhi,[2] Abel Monteagudo,[3] Julie Peacock,[1] Carlos A. Quesada,[1,4] Geertje van der Heijden,[1] Samuel Almeida,[5] Iêda Amaral,[4,6] Luzmila Arroyo,[7,8] Gerardo Aymard,[9] Tim R. Baker,[1] Olaf Bánki,[10] Lilian Blanc,[11] Damien Bonal,[12] Paulo Brando,[13,14] Jerome Chave,[15] Átila Cristina Alves de Oliveira,[4] Nallaret Dávila Cardozo,[16] Claudia I. Czimczik,[17] Ted R. Feldpausch,[1] Maria Aparecida Freitas,[5] Emanuel Gloor,[1] Niro Higuchi,[18] Eliana Jiménez,[19] Gareth Lloyd,[20] Patrick Meir,[21] Casimiro Mendoza,[22] Alexandra Morel,[2] David A. Neill,[8,23] Daniel Nepstad,[24,25] Sandra Patiño,[1,11] Maria Cristina Peñuela,[19] Adriana Prieto,[26] Fredy Ramírez,[16] Michael Schwarz,[1,27] Javier Silva,[2] Marcos Silveira,[28] Anne Sota Thomas,[29] Hans ter Steege,[30] Juliana Stropp,[30] Rodolfo Vásquez,[3] Przemyslaw Zelazowski,[2] Esteban Alvarez Dávila,[31] Sandy Andelman,[6] Ana Andrade,[4] Kuo-Jung Chao,[1] Terry Erwin,[32] Anthony Di Fiore,[33] Eurídice Honorio C.,[34] Helen Keeling,[1] Tim J. Killeen,[7] William F. Laurance,[4,35] Antonio Peña Cruz,[3] Nigel C. A. Pitman,[36] Percy Núñez Vargas,[37] Hirma Ramírez-Angulo,[38] Agustín Rudas,[39] Rafael Salamão,[5] Natalino Silva,[40] John Terborgh,[41] Armando Torres-Lezama[38]

Amazon forests are a key but poorly understood component of the global carbon cycle. If, as anticipated, they dry this century, they might accelerate climate change through carbon losses and changed surface energy balances. We used records from multiple long-term monitoring plots across Amazonia to assess forest responses to the intense 2005 drought, a possible analog of future events. Affected forest lost biomass, reversing a large long-term carbon sink, with the greatest impacts observed where the dry season was unusually intense. Relative to pre-2005 conditions, forest subjected to a 100-millimeter increase in water deficit lost 5.3 megagrams of aboveground biomass of carbon per hectare. The drought had a total biomass carbon impact of 1.2 to 1.6 petagrams (1.2×10^{15} to 1.6×10^{15} grams). Amazon forests therefore appear vulnerable to increasing moisture stress, with the potential for large carbon losses to exert feedback on climate change.

Old-growth forests in Amazonia store 120 Pg (1.2×10^{17} g) of carbon in their biomass (*1*), and through photosynthesis and respiration they process 18 Pg C annually (*2*), more than twice the rate of anthropogenic fossil fuel emissions. Relatively small changes in Amazon forest dynamics therefore have the potential to substantially affect the concentration of atmospheric CO_2 and thus the rate of climate change itself. A key parameter in determining the magnitude of this effect is the sensitivity—or resilience—of tropical forests to drought. Increased moisture stress is a dominant feature of some modeled 21st-century climate scenarios for Amazonia, particularly for southern Amazonia (*3*–*5*), and there is some evidence that this has already commenced (*6*). Prolonged tropical droughts can kill trees (*7*–*10*), and some models predict climate-induced Amazon dieback this century (*4*, *11*, *12*). But it has also been suggested that dry conditions may cause Amazon forests to "green up" (*13*, *14*) and that increases in solar radiation during drier periods boost tropical productivity (*15*–*17*). Large-scale on-the-ground assessments of the ecological impacts of tropical droughts are completely lacking, precluding tests of these ideas.

In 2005, large areas of the Amazon Basin experienced one of the most intense droughts of the past 100 years (*18*), providing a unique opportunity to directly evaluate the large-scale sensitivity of tropical forest to water deficits. The 2005 event was driven not by El Niño, as is often the case for Amazonia, but by elevated tropical North Atlantic sea surface temperatures (*18*), which affected the southern two-thirds of Amazonia and especially the southwest through reduced precipitation as well as higher-than-average temperatures (*18*, *19*). Both the anomalous North Atlantic warming and its causal link to Amazon drought are reproduced in some recent modeled scenarios for 21st-century climates (*5*, *12*), and thus the event of 2005 may provide a proxy for future climate conditions. Through a large long-term research network, RAINFOR, we have monitored forest plots across the basin for 25 years. After the drought we conducted an emergency recensus program covering all major Amazon nations, climates, soils, and vegetation types. Here we report the results of this large-scale natural experiment to assess the impact of tropical drought on the ground.

By 2005 the RAINFOR network consisted of 136 permanent plots located in old-growth forest distributed across 44 discrete landscapes ("sites") (*20*). We used tree diameter, wood density, and allometric models to compute biomass at each point in time, as well as rates of biomass gain ("growth") and loss ("mortality") between censuses, correcting for possible sampling effects (*20*). To establish the pre-2005 Amazon baseline, we first determined the long-term biomass changes in our plots. To assess drought impacts, we focused on the 2005 event, evaluating net biomass change, growth, and mortality and the differences in these relative to earlier records, focusing on the 55 plots that were regularly censused both before and after the drought. To

[1]Ecology and Global Change, School of Geography, University of Leeds, Leeds LS2 9JT, UK. [2]Environmental Change Institute, School of Geography and Environment, Oxford University, Oxford OX1 3QY, UK. [3]Jardín Botánico de Missouri, Oxapampa, Pasco, Peru. [4]Instituto Nacional de Pesquisas na Amazônia, Av. Andre Araujo, 1753 CP 478, 69060-011 Manaus AM, Brasil. [5]Museu Paraense Emílio Goeldi, Av. Perimetral, 1901 Terra Firme, CEP: 66077-830 Belém PA, Brasil. [6]Tropical Ecology Assessment and Monitoring Network (TEAM), Conservation International, 2011 Crystal Drive, Suite 500, Arlington, VA 22202, USA. [7]Museo de Historia Natural Noel Kempff Mercado, Casilla 2489, Av. Irala 565, Santa Cruz, Bolivia. [8]Missouri Botanical Garden, Box 299, St. Louis, MO 63166, USA. [9]Programa de Ciencias del Agro y del Mar, Herbario Universitario (PORT), Universidad Nacional Experimental de Los Llanos Occidentales Ezequiel Zamora, Mesa de Cavacas, Portuguesa 3350, Venezuela. [10]Nationaal Herbarium Nederland, W.C. van Unnikgebouw, Heidelberglaan 2, 3584 CS Utrecht, Netherlands. [11]Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), UMR EcoFoG, Campus Agronomique, BP 709, 97387 Kourou Cedex, French Guiana. [12]Institut National de la Recherche Agronomique (INRA), UMR EcoFoG, Campus Agronomique, BP 709, 97387 Kourou Cedex, French Guiana. [13]Instituto de Pesquisa Ambiental da Amazônia, Avenida Nazaré 669, CEP-66035, Belém PA, Brasil. [14]Department of Botany and School of Natural Resources and Environment, University of Florida, P.O. Box 118526, Gainesville, FL 32611, USA. [15]Laboratoire EDB, Université Paul Sabatier, Bâtiment 4R3, 31062 Toulouse, France. [16]Universidad Nacional de la Amazonía Peruana, Iquitos, Loreto, Perú. [17]Department of Earth System Science, University of California, Irvine, CA 92697, USA. [18]Departamento de Silvicultura Tropical, Manejo Florestal, Instituto Nacional de Pesquisas da Amazônia, Av. André Araújo, 2936 Petrópolis, Manaus AM, Brasil. [19]Universidad Nacional de Colombia, Kilómetro 2 Via Tarapacá, Leticia, Amazonas, Colombia. [20]National Australia Bank, UB2211, 800 Bourke Street, Docklands, VIC 3008, Australia. [21]School of Geosciences, University of Edinburgh, Edinburgh EH8 9XP, UK. [22]FOMABO (Manejo Forestal en las Tierras Tropicales de Bolivia), Sacta, Bolivia. [23]Jatun Sacha Foundation, Casilla 17-12-867, Avenida Rio Coca 1734, Quito, Ecuador. [24]Woods Hole Research Center, Falmouth, MA 02540, USA. [25]Gordon and Betty Moore Foundation, P.O. Box 29910, San Francisco, CA 94129, USA. [26]Instituto Alexander von Humboldt, Claustro de San Agustín, Villa de Leyva, Boyacá, Colombia. [27]Bayreuth Center of Ecology and Environmental Research, University of Bayreuth, 95440 Bayreuth, Germany. [28]Depto de Ciências da Natureza, Universidade Federal do Acre, Rio Branco AC 69910-900, Brasil. [29]Faculty of Agriculture and Horticulture, Humboldt University of Berlin, Phillipstrasse 13, 10557 Berlin, Germany. [30]Institute of Environmental Biology, Department of Biology, Faculty of Science, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, Netherlands. [31]Facultad de Ingeniería Forestal, Universidad del Tolima, 546 Ibagué, Colombia. [32]Department of Entomology, National Museum of Natural History, Smithsonian Institution, MRC 187, P.O. Box 37012, Washington, DC 20013, USA. [33]Center for the Study of Human Origins, Department of Anthropology, New York University, New York, NY 10003, USA. [34]Instituto de Investigaciones de la Amazonía Peruana, Av. José A. Quiñones km. 2.5, Apartado Postal 784, Loreto, Perú. [35]Smithsonian Tropical Research Institution, Roosevelt Avenue, Tupper Building 401 Balboa, Ancón, Panamá, República de Panamá. [36]Centro de Investigación y Capacitación del Río de Los Amigos, Madre de Dios, Perú. [37]Universidad Nacional San Antonio Abad del Cusco, Av. de la Cultura 733, Cusco, Apartado Postal N° 921, Perú. [38]INDEFOR, Facultad de Ciencias Forestales, Universidad de Los Andes, Mérida, Venezuela. [39]Instituto de Ciencias Naturales, Universidad Nacional de Colombia, Apartado 7495, Bogotá, Colombia. [40]Serviço Florestal Brasileiro, SCEN Trecho 2, Ed. Sede do IBAMA Bloco H, 70.818-900 Brasília DF, Brasil. [41]Center for Tropical Conservation, Duke University, Box 90381, Durham, NC 27708, USA.

*To whom correspondence should be addressed. E-mail: o.phillips@leeds.ac.uk; Internet: www.rainfor.org

estimate the moisture stress at each location, we compiled meteorological data sets and determined the maximum dry-season intensity for each year in the 2005 measurement interval and for each year in the entire pre-2005 measurement period. Forest sensitivity to drought was then determined by relating the change in biomass dynamics to the change in mean maximum moisture stress. The results presented below are based on the sampling unit of individual plots; in (*20*) we explore the sensitivity of our findings to varying both the spatial scale of the sampling unit and the method of estimating moisture stress.

Before 2005, plots recorded a long-term net increase in aboveground (dry-weight) biomass, weighted by sampling effort, of 0.89 Mg ha^{-1} year^{-1} (bootstrapped 95% confidence intervals: 0.65, 1.12). This increase occurred through a multidecadal period spanning dry and wet episodes, including several El Niño events. The net biomass gain was widespread and is not a sampling artifact (*20*). These results confirm previous measured and modeled indications of a persistent biomass carbon sink—now based on a much larger data set—and are consistent with Amazon forest productivity increasing with time (*21*–*25*).

By contrast, through the 2005 drought period there was no net biomass increase in monitored plots [net rate of change –0.71 (–1.93, +0.30) Mg ha^{-1} year^{-1}; n = 55, interval mean 1.97 years]. Before 2005, 76% of plots (93 of 123) gained biomass, but during the 2005 interval only 51% did so (28 of 55); this difference is highly significant ($P < 0.01$, Mann-Whitney U test). To assess whether biomass changes were drought-related, we developed meteorological and soil data sets to estimate evapotranspirational demand and soil moisture stress (*20*). For plots with longer and more intense moisture deficits than normal, there were clear net losses [–1.62 (–3.16, –0.54) Mg ha^{-1} year^{-1}; n = 38, interval mean 1.96 years]. The distribution through time of all measured biomass dynamics (Fig. 1) reveals that the drought coincided with the first substantial decline in measured biomass in Amazonian plots since measurements started. However, fingerprinting the drought impact is complicated by switching among plots being monitored, the nonequilibrium initial conditions, divergent climatologies and soils, and contrasting conditions in 2005 itself. Within-plot analyses help to control for such effects and confirm the drought's impact: Relative to their extended period of earlier biomass gains, plots monitored through 2005 experienced negative change [difference = –1.50 (–3.01, –0.44) Mg ha^{-1} year^{-1}; n = 43]. Among the 28 plots with longer and more severe water deficits than normal during 2005, the rate of aboveground woody biomass accumulation declined by 2.39 (1.12 to 3.97) Mg ha^{-1} year^{-1}, whereas by contrast the 15 nondroughted plots continued to gain [difference = +0.76 (–0.78, +2.00) Mg ha^{-1} year^{-1}].

The Amazon forest spans a large climatic range, from the almost aseasonal high-precipitation northwest to the strongly seasonal southern fringes with frequent prolonged moisture deficits (*26*, *27*). Distributions of neotropical trees reflect their drought sensitivity (*28*), so we hypothesized that any drought impacts will be experienced by plants as a function of relative departure from their long-term environmental conditions. For each site, we therefore estimated the magnitude of the drought experienced during the 2005 interval relative to local, long-term estimates of water balance. We find that relative drought is indeed strongly

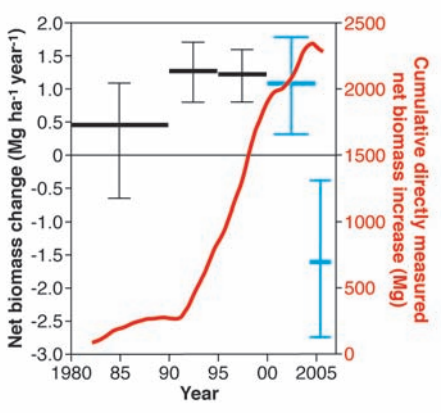


Fig. 1. Interval-by-interval, plot-by-plot net biomass change measured in Amazonia since 1980. The multidecadal carbon sink is evident, strongly reversed in 2005. Long sampling intervals may have obscured earlier fluctuations (see fig. S1). Red line (scale on right) represents the total cumulative biomass increase of Amazon trees ≥10 cm in diameter as actually measured in permanent plots, as a function of the mid-date of each census interval, with a running mean of 50 intervals. Black and blue distributions (scale on left) represent mean and 95% bootstrapped confidence intervals for interval-by-interval biomass change weighted by sampling effort (*20*). Black distributions indicate predefined periods (1980–1989, 1990–1994, 1995–1999) where the chronological span of each bin represents the interval mid-dates that fall within that period. Blue distributions align intervals with the 2005 drought event to reveal its impact, contrasting all 2000–2004 predrought measurements with all droughted plots monitored in 2005.

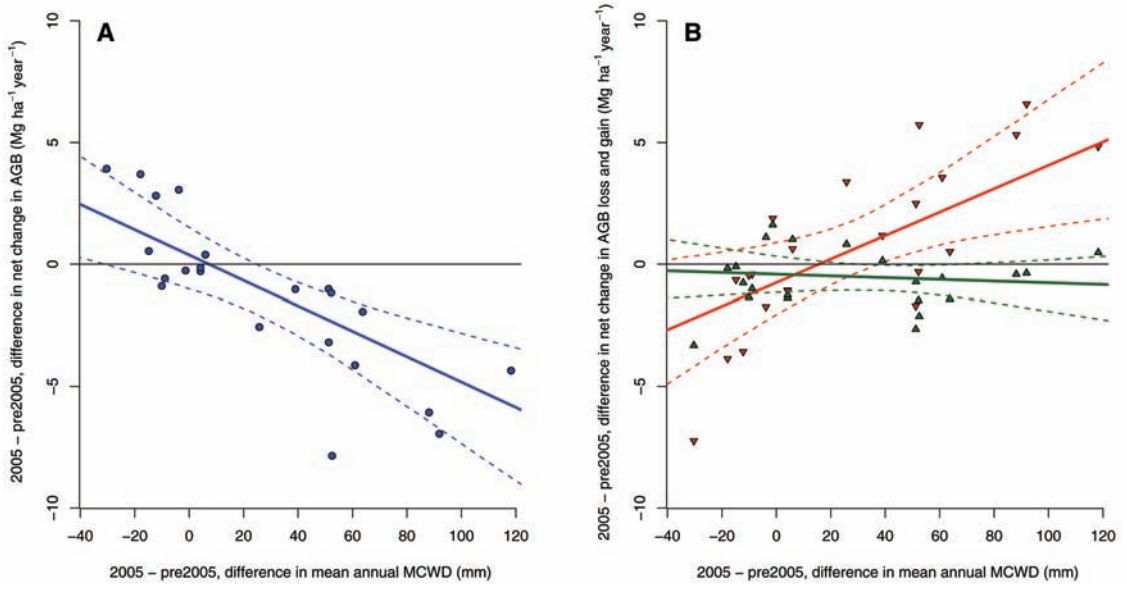


Fig. 2. Biomass dynamics response to the relative intensity of the 2005 drought. Differences in (**A**) plot biomass change (blue) and (**B**) mortality rate (red) and growth rate (green) are shown for trees ≥10 cm in diameter for the drought interval relative to pre-2005 as a linear function of drought relative intensity, weighted by monitoring effort (*20*). Change in drought intensity is measured by change in maximum climatological water deficit (MCWD, accounts only for rainfall). Uncertainty in precipitation is included in the bootstrapped estimates of the relationship of difference in biomass change versus difference in MCWD and confidence intervals (*20*). Plots known to have different 2005 interval MCWD are treated as independent; values are otherwise averaged across contributing plots. Alternative models that account for variation in soil properties, evapotranspiration, and plot definitions give very similar results (*20*). Polynomial or break-point functions do not provide closer fits.

implicated as the driver of the network-wide shift in forest behavior (Fig. 2) but that the absolute intensity of the 2005 dry period was only weakly related to biomass dynamics (fig. S5): Those forests experiencing the most elevated moisture stress relative to their long-term mean tended to lose the most biomass relative to their pre-2005 trend (Fig. 2). These losses were driven by occasionally large mortality increases and by widespread but small declines in growth. Our method may fail to capture growth impacts well because intervals were longer than the period of potential moisture constraint, thereby masking its effects (drought can kill trees but can only temporarily stop growth). Analysis at the site level confirms that the relationship between forest response and droughting is not driven by a few anomalous plots (*20*), and accounting for local soil water-holding capacity, temperature, humidity, and radiation shows this relationship to be robust regardless of how the moisture balance is estimated (*20*). Moreover, just as the earlier net gains were widespread across the basin, the 2005 declines were well distributed spatially (Fig. 3). From Fig. 2, and assuming a proportional impact on smaller trees and lianas (*20*), we estimate that an average forest hectare subject to a 100-mm increase in maximum water deficit lost 5.3 Mg of aboveground biomass carbon over the average 1.97-year drought census interval relative to pre-2005 conditions (bootstrapped confidence intervals 3.0, 8.1).

We also recorded the identity of trees that died. Fast-growing, light-wooded trees may be especially vulnerable to drought by cavitation or carbon starvation (*7*, *29*–*31*), and consistent with this, trees dying during the 2005 period had lower wood densities than those dying before. In 25 drier-than-average plots with dead trees identified, trees recorded as dead in 2006 were 5% lighter than in previous censuses [mean wood density of dead trees fell from 0.60 to 0.57 g cm^{-3} ($P = 0.02$) (*20*)]. Apparently, Amazon drought kills selectively and therefore may also alter species composition, pointing to potential consequences of future drought events on the biodiversity in the Amazon region.

Relative to the predrought sink, we estimate a total impact of –1.21 Pg C (–2.01, –0.57) by simply scaling the per-plot impact by the total droughted area (~3.3×10^8 ha) and assuming that nonmeasured components of biomass were equally affected. Scaling the per-site impact yields slightly greater values (*20*). Alternatively, we can scale the observed relationship between relative biomass change in plots and droughting (Fig. 2) by the moisture deficits across Amazonia estimated from remotely sensed rainfall data (*19*, *20*). This suggests an even greater impact on the biomass carbon balance of the droughted area: –1.60 Pg C (–2.63, –0.83). Site-based scaling-up indicates similar values (*20*). Although better understanding of soils is needed to determine the local effects of meteorological drought, the magnitude and consistency of these estimates demonstrate Amazonia's vulnerability to drought and the potential for changes in tropical climates to have large carbon cycle impacts. Our on-the-ground data reveal that, despite apparent "greening up" during dry periods (*13*, *14*), Amazon drought accelerates mortality over large areas (Fig. 2B) (*20*).

The exceptional growth in atmospheric CO_2 concentrations in 2005, the third greatest in the global record (*32*), may have been partially caused by the Amazon drought effects documented here. However, our findings do not translate simply into instantaneous flux estimates because carbon fluxes from necromass will lag the actual tree death events. Drought can suppress respiration (*17*, *33*), so the system as a whole might even contribute a temporary net sink even though the live biomass was in negative mass balance. Nonetheless, our results constrain the aggregate impacts of drought because trees are by far the largest and longest-lived of the aboveground carbon stores. Tropical droughts may intensify and become more frequent this century as a result of anthropogenic climate change (*1*, *3*–*5*, *11*). In addition to directly affecting Amazonian peoples and biodiversity, such events appear capable of strongly altering the regional carbon balance and thereby accelerating climate change.

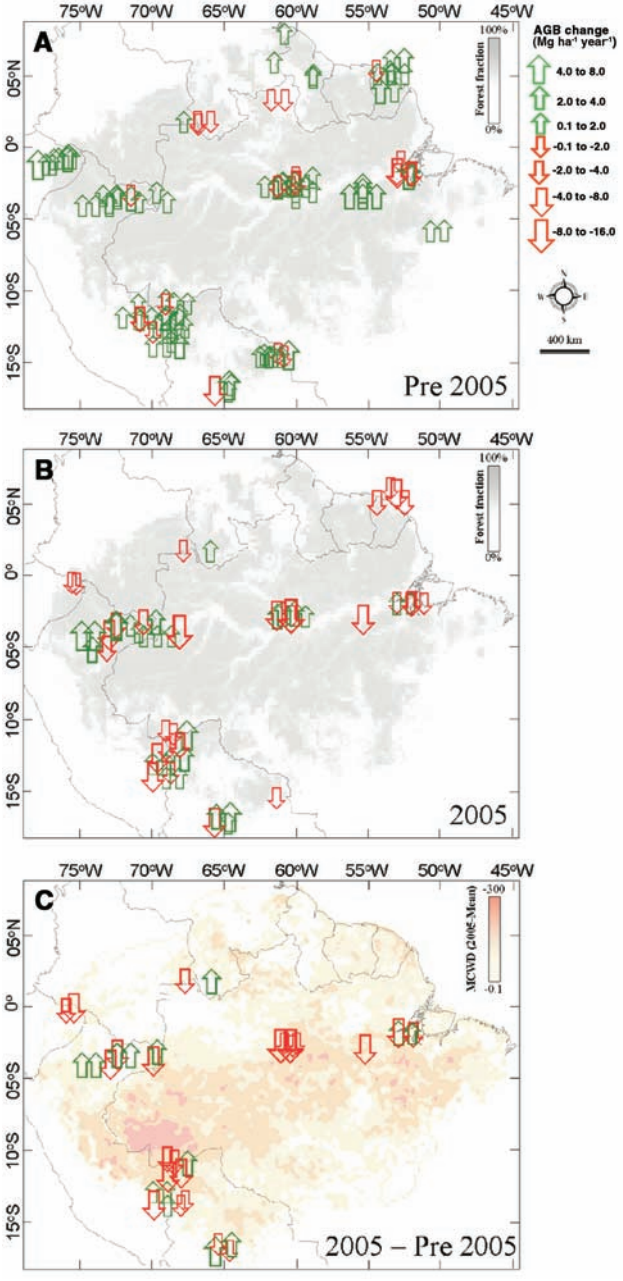


Fig. 3. Aboveground biomass change in the Amazon Basin and contiguous lowland moist forests. The 2005 drought reversed a multidecadal biomass carbon sink across Amazonia. Symbols represent magnitude and direction of measured change and approximate location of each plot. (**A**) Annual aboveground biomass change before 2005. (**B**) Annual aboveground biomass change during the 2005 interval. (**C**) Difference in rates of change in aboveground biomass, 2005 versus pre-2005, for those plots monitored throughout. Grayscale shading in (A) and (B) represents proportion of area covered by forests. Colored shading in (C) indicates the intensity of the 2005 drought relative to the 1998–2004 mean as measured from space using radar-derived rainfall data [Tropical Rainfall Measuring Mission (TRMM)].

References and Notes

1. Y. Malhi *et al.*, *Science* **319**, 169 (2008); published online 27 November 2007 (10.1126/science.1146961).
2. Y. Malhi, J. Grace, *Trends Ecol. Evol.* **15**, 332 (2000).
3. J. H. Christensen *et al.*, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon *et al.*, Eds. (Cambridge Univ. Press, Cambridge, 2007).

4. L. F. Salazar, C. A. Nobre, M. D. Oyama, *Geophys. Res. Lett.* **34**, L09708 (2007).
5. P. M. Cox *et al*., *Nature* **453**, 212 (2008).
6. W. Li, R. Fu, R. I. Negrón Juárez, K. Fernandes, *Philos. Trans. R. Soc. London Ser. B* **363**, 1767 (2008).
7. R. Condit, S. P. Hubbell, R. B. Foster, *Ecol. Monogr.* **65**, 419 (1995).
8. G. B. Williamson *et al*., *Conserv. Biol.* **14**, 1538 (2000).
9. D. C. Nepstad, I. M. Tohver, D. Ray, P. Moutinho, G. Cardinot, *Ecology* **88**, 2259 (2007).
10. P. M. Brando *et al*., *Philos. Trans. R. Soc. London Ser. B* **363**, 1839 (2008).
11. P. M. Cox, R. A. Betts, C. D. Jones, S. A. Spall, I. J. Totterdell, *Nature* **408**, 184 (2000).
12. P. P. Harris, C. Huntingford, P. M. Cox, *Philos. Trans. R. Soc. London Ser. B* **363**, 1753 (2008).
13. A. R. Huete *et al*., *Geophys. Res. Lett.* **13**, L06405 (2006).
14. S. R. Saleska, K. Didan, A. R. Huete, H. R. da Rocha, *Science* **318**, 612 (2007).
15. R. Condit *et al*., *J. Trop. Ecol.* **20**, 51 (2004).
16. E. A. Graham, S. Mulkey, K. Kitajima, K. N. Phillips, S. J. Wright, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 572 (2003).
17. D. Bonal *et al*., *Global Change Biol.* **14**, 1917 (2008).
18. J. A. Marengo *et al*., *J. Clim.* **21**, 495 (2008).
19. L. E. O. C. Aragão *et al*., *Geophys. Res. Lett.* **34**, L07701 (2007).
20. See supporting material on *Science* Online.
21. O. L. Phillips *et al*., *Science* **282**, 439 (1998).
22. R. R. Nemani *et al*., *Science* **300**, 1560 (2003).
23. T. R. Baker *et al*., *Philos. Trans. R. Soc. London Ser. B* **359**, 353 (2004).
24. S. L. Lewis *et al*., *Philos. Trans. R. Soc. London Ser. B* **359**, 421 (2004).
25. K. Ichii, H. Hashimoto, R. Nemani, M. White, *Global Planet. Change* **48**, 274 (2005).
26. C. von Randow *et al*., *Theor. Appl. Climatol.* **78**, 5 (2004).
27. N. Hasler, R. Avissar, *J. Hydrometeorol.* **8**, 380 (2007).
28. B. M. Engelbrecht *et al*., *Nature* **447**, 80 (2007).
29. M. Tyree, J. Sperry, *Annu. Rev. Plant Phys. Plant Mol. Biol.* **40**, 19 (1989).
30. U. Hacke, J. Sperry, W. Pockman, S. Davis, K. McCuloh, *Oecologia* **126**, 457 (2001).
31. N. McDowell, W. Pockman, C. Allen, D. Breshears, N. Cobb, *New Phytol.* **178**, 719 (2008).
32. National Oceanic and Atmospheric Administration, Trends in Atmospheric Carbon Dioxide, Global (www.esrl.noaa.gov/gmd/ccgg/trends, 2008).
33. P. Meir, D. B. Metcalfe, A. C. Costa, R. A. Fisher, *Philos. Trans. R. Soc. London Ser. B* **363**, 1849 (2008).
34. This paper is a product of the RAINFOR network. Funding for the 2006 recensuses came primarily from a Natural Environment Research Council (NERC) Urgency Grant. The results summarized here also depend on contributions from numerous field assistants and rural communities in Brazil, Bolivia, Colombia, Ecuador, French Guiana, Guyana, Peru, and Venezuela, many previously acknowledged. I. Huamantupa, N. Jaramillo, N. Saavedram (Peru), V. Peña (Ecuador), J.-C. Arias, and D. Navarrete (Colombia) assisted with recent censuses. We thank CNPQ (Brazil), MCT (Brazil), Ministerio del Medio Ambiente, Vivienda y Desarrollo Territorial (Colombia), Ministerio de Ambiente (Ecuador), the Forestry Commission (Guyana), INRENA (Perú), and Ministerio del Ambiente para el Poder Popular (Venezuela) for research permissions, and the Large Scale Biosphere-Atmosphere Experiment in Amazonia (LBA) for logistical support. This paper was supported by the Leverhulme Trust (O.L.P.), NERC grants NE/B503384/1 and NE/D01025X/1 (O.L.P.), NER/A/S/2003/00608/2 (Y.M.), WOTRO (W84-581) (H.t.S., O.B.), the Royal Society (S.L.L., Y.M.), University of Leeds (T.R.B., K.-J.C., G.L.-G.), and a Gordon and Betty Moore Foundation grant to RAINFOR. Some of the data used in these analyses were collected with support from the TEAM Network of Conservation International, funded by the Gordon and Betty Moore Foundation. We thank D. Galbraith, L. Anderson, C. Michelaki, J. Ratnam, and M. Wagner for discussions; three anonymous reviewers for comments on the manuscript; M. García-Hernandez for help in managing co-author contributions; and A. Manson and D. Appleyard for help in creating the figures.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5919/1344/DC1
Materials and Methods
SOM Text
Figs. S1 to S8
Tables S1 to S7
References

31 July 2008; accepted 24 December 2008
10.1126/science.1164033

Species Response to Environmental Change: Impacts of Food Web Interactions and Evolution

Jason P. Harmon,[1]* Nancy A. Moran,[2] Anthony R. Ives[1]

How environmental change affects species abundances depends on both the food web within which species interact and their potential to evolve. Using field experiments, we investigated both ecological and evolutionary responses of pea aphids (*Acyrthosiphon pisum*), a common agricultural pest, to increased frequency of episodic heat shocks. One predator species ameliorated the decrease in aphid population growth with increasing heat shocks, whereas a second predator did not, with this contrast caused by behavioral differences between predators. We also compared aphid strains with stably inherited differences in heat tolerance caused by bacterial endosymbionts and showed the potential for rapid evolution for heat-shock tolerance. Our results illustrate how ecological and evolutionary complexities should be incorporated into predictions of the consequences of environmental change for species' populations.

Species throughout the world face many anthropogenic environmental disturbances (*1*). Some disturbances, such as land-use change, occur progressively and predictably. Others take place as increases in the frequency or magnitude of environmental shocks, such as the anticipated increase in tropical storm severity (*2*). Regardless of the mode of disturbance, changes in species abundance will depend on the multigenerational response of their survival and reproduction within ecosystems. Although the response of species' populations depends on the direct effects of environmental disturbances on species physiology, behavior, and life history (*3*, *4*), three additional complexities may play major roles in the long-term change in species' populations (*5*).

First, the change in a species' population growth rate in response to an environmental disturbance depends on how the species interacts ecologically with other species in the ecosystem (*6*). For example, if a competitively dominant species is sensitive to a disturbance, then a competitively subordinate species may benefit indirectly from the disturbance through competitive release (*7*). Although the role of food web interactions is well-known in theoretical work (*8*) and a growing number of empirical studies document these effects (*9*–*11*), most of this work has not considered how the strength of these interactions might change because of density-dependent effects during the environmental change.

A second complexity is the possibility that species may evolve tolerance to the environmental change (*12*). Empirical studies have now documented a growing list of species that have undergone evolutionary responses to environmental changes (*13*, *14*). If genetic variation exists, then environmental disturbances with large impacts on population growth rates may drive rapid evolution of tolerance.

The third complexity is that ecological and evolutionary complexities might interact (*15*). If ecological interactions modify the response of population growth rates to environmental changes, then they might also modify the selective regime for tolerance and, hence, evolution. In turn, evolution may change population growth rates and interactions among species, thereby increasing the complexities of predicting population changes.

Here, we investigate these three complexities for predicting population changes of pea aphids in response to increasing frequency of episodic heat shocks. To show that ecological interactions can modify population responses to environmental disturbances, we subjected field-caged populations of pea aphids and predators to an experimentally increased frequency of heat shocks (*16*). Our goal was to contrast the effects of two similar ladybeetle predators, investigating how species-specific differences in aphid density–dependent attack rates affect the change in aphid population growth rates when subjected to environmental change. To investigate the potential for evolution, we constructed aphid strains that differed in the presence of stably inherited endosymbionts that affect heat-shock tolerance. We

[1]Department of Zoology, University of Wisconsin, Madison, WI 53706, USA. [2]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA.

*To whom correspondence should be addressed. E-mail: jharmon@wisc.edu

deployed these strains in field experiments run at the same time scale as the ecological experiments to measure the rate of evolution of heat-shock tolerance. Finally, we derived a model parameterized exclusively from field data that illustrates how these ecological and evolutionary processes may interact. We selected heat shocks as an environmental disturbance because global climate models predict that short exposures to high temperature will occur with increasing frequency and intensity (*17*).

The pea aphid, *Acyrthosiphon pisum* (Harris) (Homoptera: Aphididae), is an ecologically and evolutionarily well-studied organism. Heat shocks affect pea aphid population growth rates by reducing fecundity or even sterilizing females, with less severe effects on survival and development times (*16*). These effects are similar to those documented for many species; even short periods of high temperatures can denature proteins and cause numerous physiological and developmental problems (*18*). Tolerance to heat shocks in pea aphid strains may be conferred by certain secondary (facultative) bacterial symbionts (*19*, *20*) and is also strongly affected by a common mutation affecting heat-shock genes in the obligate bacterial symbiont *Buchnera* (*21*). Because endosymbionts are invariably transmitted during parthenogenetic reproduction, they are analogous to inherited traits in monoclonal aphid lines (*22*). Secondary symbionts conferring heat tolerance have higher prevalence after periods of summer heat (*19*) and are present in 100% of pea aphids in hot desert sites (*16*), consistent with selection for heat-shock tolerance. The secondary endosymbiont we used occurs naturally at low frequency in populations at our study site; in 2008, 2 out of 57 assayed aphids contained the endosymbiont conferring heat-shock tolerance (*16*). The *Buchnera* allele conferring heat sensitivity occurs variably in field populations; it occurred in 21% of individuals in one sample from our study area (*21*) and, in others, ranged from 66% in 1999 to 0% in 2008. The existence of natural strains varying in tolerance (and our knowledge of the bases of this variation) makes pea aphids a good model system for studying the consequences of environmental changes.

Pea aphids sometimes attain very high population densities and destroy crops of their legume hosts (*23*). However, in south-central Wisconsin, USA, pea aphids in alfalfa rarely reach densities high enough to cause economic crop damage due to a suite of natural enemies, especially two predatory ladybird beetles: *Coccinella septempunctata* L. and *Harmonia axyridis* (Pallas) (Coleoptera: Coccinellidae) (*24*). Although both predators occur in alfalfa, they have distinctly different behavioral responses to pea aphid abundance (*16*). As a consequence, *C. septempunctata* occurs when few pea aphids are in alfalfa fields and is only slightly more abundant when more aphids are present. In contrast, *H. axyridis* is absent from fields with low aphid abundance but becomes more common quickly as aphid abundance increases (Fig. 1). These differences in predator responses to aphid densities lead to contrasting predictions about how predation will modify pea aphid population growth rates when subjected to heat shocks. Because *C. septempunctata* populations do not rapidly decline with decreasing pea aphid abundance, *C. septempunctata* should continue to exert predation pressure in the presence of the heat shocks that can directly reduce aphid abundance. Conversely, predation pressure from *H. axyridis* rapidly decreases with declining aphid densities, and therefore, its effect on aphid population growth should diminish when aphids face heat-shock disturbances. This would indirectly ameliorate the impact of heat shocks on pea aphids, because the shocked aphid population would suffer less predation.

To test these predictions, we conducted a field experiment using 2- × 2- × 2-m mesh cages and a pea aphid strain sensitive to heat shocks, having a heat-sensitive genotype in the primary symbiont and lacking secondary endosymbionts that provide protection (*16*). We used a 2 × 2 factorial design: with or without supplemented heat shocks and with or without predators. Heat shocks were experimentally imposed by covering cages with clear plastic sheeting for 4 hours at midday three times per week. This increased temperatures by ~5°C, exceeding the threshold at which pea aphid fecundity is affected but remaining within the temperature range naturally observed at the study site. Hence, these manipulations represent increasing frequency of heat shocks rather than increasing magnitude of temperature beyond natural variation.

We performed the experiment twice, once with *C. septempunctata* and once with *H. axyridis* as the predator. In both experiments, heat shocks

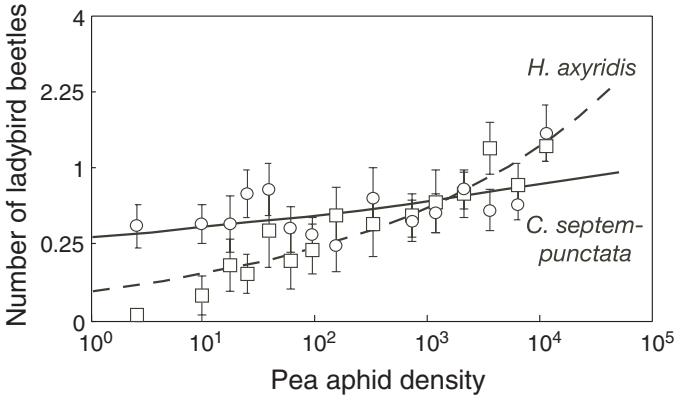


Fig. 1. Relation between pea aphid density in alfalfa fields and the relative densities of adult *C. septempunctata* (circles, solid line) and *H. axyridis* (squares, dashed line) for 2003–2007. For display, we grouped data into 15 aphid density bins having equal numbers of samples. *C. septempunctata* and *H. axyridis* differed strongly in their responses to changing pea aphid density for data from the first and second cycles ($t_{647} = 6.01$, $P < 10^{-6}$) (*16*). Error bars indicate ± 1 SE.

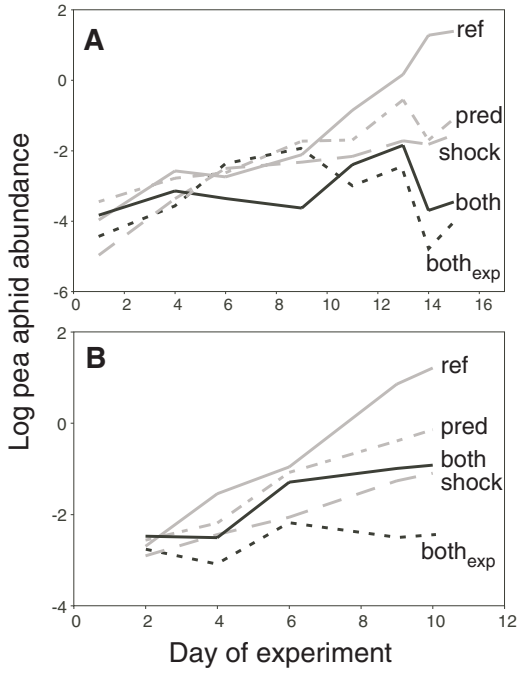


Fig. 2. Interactive effects of heat shocks and predation by (**A**) *C. septempunctata* and (**B**) *H. axyridis* on pea aphid populations in 20 field cages, as revealed by coefficients estimated in a generalized linear mixed model (GLMM) (*16*). The different lines indicate no heat shocks and no predators (solid gray lines, labeled "ref"); heat shocks and no predators (gray lines with long dashes, "shock"); no heat shocks and predators (gray lines with short dashes, "pred"); and heat shocks and predators (solid black lines, "both"). The black dashed line ("$both_{exp}$") gives the expected pea aphid densities if there were no interactions between heat shock and predators. The graphs present the expected values for each treatment day, calculated from the GLMM (*16*). Although for *C. septempunctata* the interaction of heat shock*predators*day was significant ($\chi^2_8 = 76.8$, $P < 10^{-6}$), it was significant only because of the negative interactions on days 6 and 9; confining the statistical analysis to the last 4 days gave no interaction ($\chi^2_4 = 2.4$, $P > 0.5$). For *H. axyridis*, there was a statistically significant positive interaction ($\chi^2_5 = 41.8$, $P < 10^{-6}$).

caused statistically significant reductions in pea aphid abundances, as did the presence of predators. When *C. septempunctata* was the predator, heat shocks caused the same proportional reduction in aphid population growth rates when the predator was absent (Fig. 2A, "ref" versus "shock") or present ("pred" versus "both"); by the end of the experiment, the pea aphid densities that were expected if the effects of heat shock and predation were additive matched the observed densities ("$both_{exp}$" versus "both"). In contrast, when *H. axyridis* was the predator, the effect of heat shock was ameliorated in the presence of predation (Fig. 2B,"$both_{exp}$" versus "both"); at low aphid densities, predation by *H. axyridis* diminished. If the data set for *C. septempunctata* is shortened to the same length as that of the *H. axyridis* data, this contrast is even stronger with a statistically significant interaction for *C. septempunctata* in the opposite direction from *H. axyridis* (*16*). This pattern of predation by *H. axyridis* is consistent with the field observations (Fig. 1); both patterns are probably driven by a suite of correlated behaviors that encourages *H. axyridis* to focus foraging at high aphid densities (*16*). These results show that the impact of heat-shock disturbance on pea aphids depends not just on the presence of predators, but also on the identity of the predator that is present.

To investigate the potential for evolution for heat-shock tolerance, we used four clones that differed in endosymbionts to maximize differences in heat sensitivity. Pea aphids have two genetically based color morphs, green and red, and we took advantage of this color polymorphism to measure the population growth rate of two different clones within the same field cages. We used two different pairs of green and red clones such that in pair A, the green clone was heat-shock tolerant (contained heat-resistant primary endosymbiont plus a protective secondary endosymbiont) and the red clone was susceptible to heat shocks; in pair B, the red clone was tolerant and the green clone was susceptible. These two pairs allowed us to control for potential ecological differences that have been attributed to pea aphid color morph (*25*). We conducted an experiment subjecting either pair A or B to either ambient or increased heat-shock conditions (*16*). For both pairs, the heat-shock sensitive clone had slightly higher population growth rates than did the heat-shock tolerant clone in the absence of experimental heat shocks, but both heat-shock sensitive clones also had greatly reduced population growth rates in the presence of heat shocks. These population growth rates translate into strong selection against heat-sensitive clones in the presence of heat shocks (Table 1).

To illustrate the possible combined effects of predation and evolution on the long-term response of pea aphid populations to increased exposure to heat shocks, we produced a simplified mathematical model that uses only information derived from our studies. Although not designed to make quantitative predictions, our model nonetheless addresses qualitative expectations about interactions between ecological and evolutionary processes. In the model, there are sensitive and tolerant aphid clones, but initially the pea aphid population is almost entirely composed of heat-sensitive clones. Sensitive and tolerant clones have different population growth rates under ambient and shocked regimes corresponding to the average growth rates of sensitive and tolerant clones under each environmental regime in the evolution experiment (Table 1). We included predation in the model by assuming that predation pressure follows the same function of aphid density as we observed in field surveys of ladybeetle abundance, with *C. septempunctata* showing a smaller reduction in abundance than *H. axyridis* when pea aphid densities are low (Fig. 1). We then assumed that the environmental regime changes from our experimental conditions without heat shocks to those with heat shocks. This led to a direct change in the pea aphid population growth rates and an indirect change in predation pressure.

In the model parameterized for *C. septempunctata* (Fig. 3A), a rapid decrease in the population abundance of pea aphids is followed by a substantial recovery that coincides with the evolutionary increase in tolerance (Fig. 3B). In the absence of evolution, however, the pea aphid population would have declined to extinction because of the combined deleterious effects of the heat shock and *C. septempunctata* (Fig. 3A). In contrast, in the model parameterized for *H. axyridis*, the initial decrease in pea aphid density is more modest, and evolution leads to a recovery of the population to near the level before the change in environmental regimes. The rate of evolution was identical, regardless of the predator modeled (Fig. 3B) because, in the model, the rate of evolution depends on the relative fitness of tolerant versus sensitive phenotypes, not on their absolute fitness. The only way that predation could change the rate of evolution is if predators selectively attacked heat-tolerant or -sensitive

Table 1. Per capita population growth rates and selection coefficients for heat-sensitive and heat-tolerant pea aphid clones in a field experiment in which the presence of heat shocks was manipulated (*16*). Selection coefficients are placed under the aphid clone selected against for a given pair of aphid clones in a given treatment.

Parameter	Sensitive	Tolerant
	Pair A	
Population growth rate, r (no shock)	0.243 ± 0.009	0.214 ± 0.007
Selection coefficient*	–	0.25
Population growth rate, r (shock)	0.155† ± 0.027	0.234† ± 0.023
Selection coefficient	0.55	–
	Pair B	
Population growth rate, r (no shock)	0.269 ± 0.008	0.247 ± 0.010
Selection coefficient	–	0.20
Population growth rate, r (shock)	0.129† ± 0.033	0.208† ± 0.031
Selection coefficient	0.55	–

*Selection coefficients are calculated as $1 - \exp(-|r_s - r_t|T)$, where r_s and r_t are the population growth rates of heat-sensitive and heat-tolerant clones, and assuming a generation time of $T = 10$ days. †For both pair A and pair B (different combinations of color morphs), the decrease in r of the heat-sensitive clone due to heat shocking was greater than the decrease in r of the heat-tolerant clone; likelihood ratio test, pair A: $\chi^2_1 = 12.1$, $P < 0.001$; pair B, $\chi^2_1 = 6.96$, $P < 0.01$.

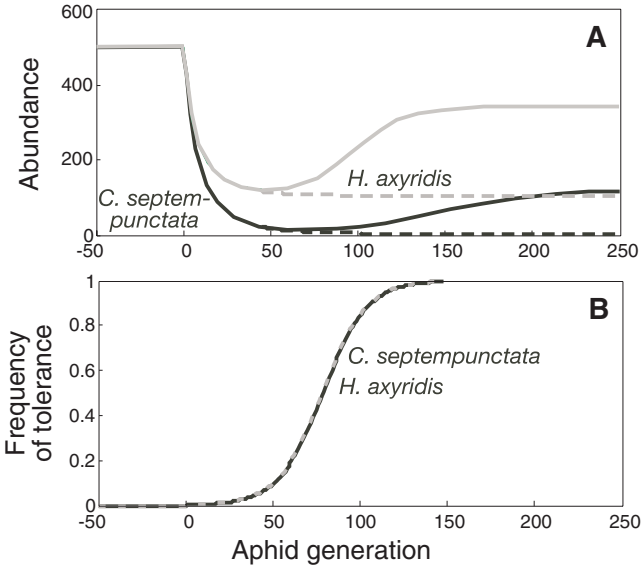


Fig. 3. Model incorporating results from field studies (Fig. 1 and Table 1) on pea aphid responses to heat shock environmental regimes in the presence of different ladybeetle species. (**A**) Pea aphid abundance with *H. axyridis* (gray lines) and *C. septempunctata* (black lines), with (solid lines) and without (dashed lines) evolution. (**B**) The frequency of tolerant phenotype in the population (results for both predators coincide). At generation 0, the environmental regime was changed to include heat shocks, and the tolerant phenotype was introduced at a frequency of 0.002.

aphids when they are mixed within a population; under laboratory conditions, they show no such selectivity (*16*).

Our results highlight three lessons about the consequences of ecological and evolutionary complexities for how environmental change affects species abundances. First, changes in population abundances depend not only on the interactions with other species in a food web, but also on the strengths of these interactions and how the strengths change during environmental disturbances. *C. septempunctata* and *H. axyridis* had different effects on pea aphid abundances because their attack rates showed different relations to aphid abundance. This complexity is a challenge for studies on the effects of environmental change, because the role of species interactions might depend on the species-specific ecologies that affect these interactions (*26*). Whereas studies have considered direct effects of climate change on species interactions—for example, by increasing transmission or attack rates from pathogens and predators (*27*, *28*) or by causing phenological mismatches between plants and pollinators (*5*, *29*)—in our study, the interaction strengths are affected indirectly through changes in species densities.

Second, ecological and evolutionary processes operate on the same time scales. Our field experiments on ecological species interactions and evolution of heat-shock tolerance were conducted at the same time scales (2 to 3 aphid generations), and they showed strong ecological or evolutionary effects. Our experiments add to the growing number of studies documenting rapid evolution and the artificial distinction between ecological and evolutionary time scales (*30*).

Third, ecological and evolutionary processes that modify how species abundances respond to environmental change may not interact (Fig. 3). For pea aphids, the identity of the predator species affects the absolute aphid population growth rate in response to increasing frequency of heat shocks, but it is unlikely to affect the relative growth rate of sensitive and tolerant aphid strains. Thus, even though species interactions themselves may have evolutionary consequences for traits that affect the interactions (*31*), they may have few consequences for traits that affect species tolerances to a different selective pressure. This separation of ecological and evolutionary complexities may simplify predictions of the impacts of environmental changes.

Our study illustrates the ecological and evolutionary complexities of predicting the responses of species to environmental changes. Changes in species' abundances may depend on the specific characteristics of the species with which they interact, and evolution can occur so rapidly that it cannot be ignored, even in the short term. Nonetheless, it is possible to address both ecological and evolutionary complexities simultaneously, and it is necessary to understand both to predict how environmental changes will affect species.

References and Notes

1. Millennium Ecosystem Assessment, *Ecosystems and Human Wellbeing: General Synthesis* (Island Press, Washington, DC, 2005).
2. P. J. Webster, G. J. Holland, J. A. Curry, H.-R. Chang, *Science* **309**, 1844 (2005).
3. A. Clarke, *Trends Ecol. Evol.* **18**, 573 (2003).
4. W. F. Morris *et al.*, *Ecology* **89**, 19 (2008).
5. C. Parmesan, *Annu. Rev. Ecol. Evol. Syst.* **37**, 637 (2006).
6. J. M. Tylianakis, R. K. Didham, J. Bascompte, D. A. Wardle, *Ecol. Lett.* **11**, 1351 (2008).
7. T. M. Frost, S. R. Carpenter, A. R. Ives, T. K. Kratz, in *Linking Species and Ecosystems*, C. G. Jones, J. H. Lawton, Eds. (Chapman & Hall, New York, 1995), pp. 224–239.
8. A. R. Ives, *Ecology* **76**, 926 (1995).
9. E. Post, C. Pedersen, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12353 (2008).
10. L. Jiang, P. J. Morin, *J. Anim. Ecol.* **73**, 569 (2004).
11. K. B. Suttle, M. A. Thomsen, M. E. Power, *Science* **315**, 640 (2007).
12. P. Gienapp, C. Teplitsky, J. S. Alho, J. A. Mills, J. Merila, *Mol. Ecol.* **17**, 167 (2008).
13. A. S. Jump, J. Penuelas, *Ecol. Lett.* **8**, 1010 (2005).
14. A. A. Hoffmann, Y. Willi, *Nat. Rev. Genet.* **9**, 421 (2008).
15. C. de Mazancourt, E. Johnson, T. G. Barraclough, *Ecol. Lett.* **11**, 380 (2008).
16. Supporting material is available on *Science* Online.
17. N. S. Diffenbaugh, J. S. Pal, R. J. Trapp, F. Giorgi, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15774 (2005).
18. M. E. Feder, G. E. Hofmann, *Annu. Rev. Physiol.* **61**, 243 (1999).
19. C. B. Montllor, A. Maxmen, A. H. Purcell, *Ecol. Entomol.* **27**, 189 (2002).
20. J. A. Russell, N. A. Moran, *Proc. R. Soc. London Ser. B.* **273**, 603 (2006).
21. H. E. Dunbar, A. C. C. Wilson, N. R. Ferguson, N. A. Moran, *PLoS Biol.* **5**, e96 (2007).
22. N. A. Moran, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8627 (2007).
23. J. J. Davis, in *USDA Bull. No. 276* (Washington, DC, 1915), pp. 5–9.
24. B. J. Cardinale, C. T. Harvey, K. Gross, A. R. Ives, *Ecol. Lett.* **6**, 857 (2003).
25. J. E. Losey, A. R. Ives, J. Harmon, F. Ballantyne, C. Brown, *Nature* **388**, 269 (1997).
26. M. D. Bertness, P. J. Ewanchuk, *Oecologia* **132**, 392 (2002).
27. J. A. Pounds *et al.*, *Nature* **439**, 161 (2006).
28. J. M. Durant, D. O. Hjermann, G. Ottersen, N. C. Stenseth, *Clim. Res.* **33**, 271 (2007).
29. J. Memmott, P. G. Craze, N. M. Waser, M. V. Price, *Ecol. Lett.* **10**, 710 (2007).
30. J. N. Thompson, *Trends Ecol. Evol.* **13**, 329 (1998).
31. A. A. Agrawal, J. A. Lau, P. A. Hamback, *Q. Rev. Biol.* **81**, 349 (2006).

32. We thank G. Burke for screening aphids for symbionts; M. Meisner, D. Rowlands, and K. Smith for assistance with experiments; D. Frye and the staff of the University of Wisconsin Arlington Research Station for establishment and maintenance of alfalfa fields; and K. C. Abbott, S. R. Carpenter, M. A. Duffy, R. T. Gilman, C. Gratton, and W. E Snyder for insights and help with the manuscript. Supported by US-NSF grants 0313737 to N.A.M. and A.R.I.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5919/1347/DC1
SOM Text
Figs. S1 and S2
Tables S1 to S4
References

20 October 2008; accepted 22 January 2009
10.1126/science.1167396

Sensing Chromosome Bi-Orientation by Spatial Separation of Aurora B Kinase from Kinetochore Substrates

Dan Liu,[1] Gerben Vader,[2]* Martijn J. M. Vromans,[2] Michael A. Lampson,[1]†‡ Susanne M. A. Lens[2]†

Successful cell division requires that chromosomes attach to opposite poles of the mitotic spindle (bi-orientation). Aurora B kinase regulates chromosome-spindle attachments by phosphorylating kinetochore substrates that bind microtubules. Centromere tension stabilizes bi-oriented attachments, but how physical forces are translated into signaling at individual centromeres is unknown. Using fluorescence resonance energy transfer–based biosensors to measure localized phosphorylation dynamics in living cells, we found that phosphorylation of an Aurora B substrate at the kinetochore depended on its distance from the kinase at the inner centromere. Furthermore, repositioning Aurora B closer to the kinetochore prevented stabilization of bi-oriented attachments and activated the spindle checkpoint. Thus, centromere tension can be sensed by increased spatial separation of Aurora B from kinetochore substrates, which reduces phosphorylation and stabilizes kinetochore microtubules.

Accurate chromosome segregation during cell division is essential to maintain genome integrity. Before segregation, kinetochores of sister chromatids attach to microtubules from opposite spindle poles (bi-orientation). This configuration is achieved through a trial-and-error process in which correct attachments exert tension across the centromere, which stabilizes kinetochore-microtubule interactions. Incorrect attachments, for example, if both sister chromatids attach to a single spindle pole, exert less tension and are destabilized, providing a new opportunity to bi-orient (*1*, *2*). How tension is coupled to kinetochore-microtubule stability is not known.

The mitotic kinase Aurora B (Ipl1 in budding yeast) localizes to the inner centromere, between sister kinetochores, and destabilizes

microtubule attachments by phosphorylating kinetochore substrates, including Dam1 and the Ndc80 complex (*3*–*10*). An appealing model is that Aurora B substrates are selectively phosphorylated at incorrect attachments. To test this model, we first examined phosphorylation of centromere protein A (CENP-A) Ser-7, a known kinetochore substrate (*11*). We used an assay in which Aurora B inhibition leads to a high frequency of syntelic attachment errors, with sister chromatids connected to a single spindle pole (*6*, *12*) (fig. S1A). We compared phospho-CENP-A staining at correct and incorrect attachments 10 min after removing the reversible Aurora B kinase inhibitor ZM447439 (*13*), which reactivates Aurora B. Phospho-CENP-A staining was strongest at incorrect attachments, identified as unaligned kinetochores (Fig. 1, A and B), which indicates that phosphorylation of an Aurora B substrate at the kinetochore depends on the microtubule attachment state.

We considered two models to explain how selective phosphorylation at incorrect attachments might be achieved. First, tension could directly regulate kinase activity by inducing a conformational change either in the kinase itself or in its associated regulatory proteins (*14*). Second, kinase activity could be constant, but tension pulls kinetochores away from Aurora B at the inner centromere (*4*, *15*). This spatial separation might lead to decreased phosphorylation of Aurora B substrates at the kinetochore. To distinguish between these models, we measured phosphorylation of an Aurora B substrate localized to either the centromere or the kinetochore, using biosensors that report quantitative changes in phosphorylation by Aurora B in living cells through changes in fluorescence resonance energy transfer (FRET) (*16*). Sensors were targeted either to the centromere using the centromere-targeting domain of CENP-B (*17*) or to the kinetochore using Mis12 (*18*) (fig. S1A), and their localization was verified (Fig. 1, C to F). Both sensors respond to Aurora B inhibition, consistent with previous findings (*16*) (fig. S1, B and C).

We first examined steady-state phosphorylation levels in cells treated with either nocodazole to depolymerize microtubules, the kinesin-5 inhibitor monastrol to create monopolar spindles and syntelic attachments (*19*), or the proteasome inhibitor MG132 to arrest cells at metaphase with bi-oriented attachments. Measurements of interkinetochore distances (*20*) confirmed that tension was low in cells treated with nocodazole (0.6 ± 0.1 μm,

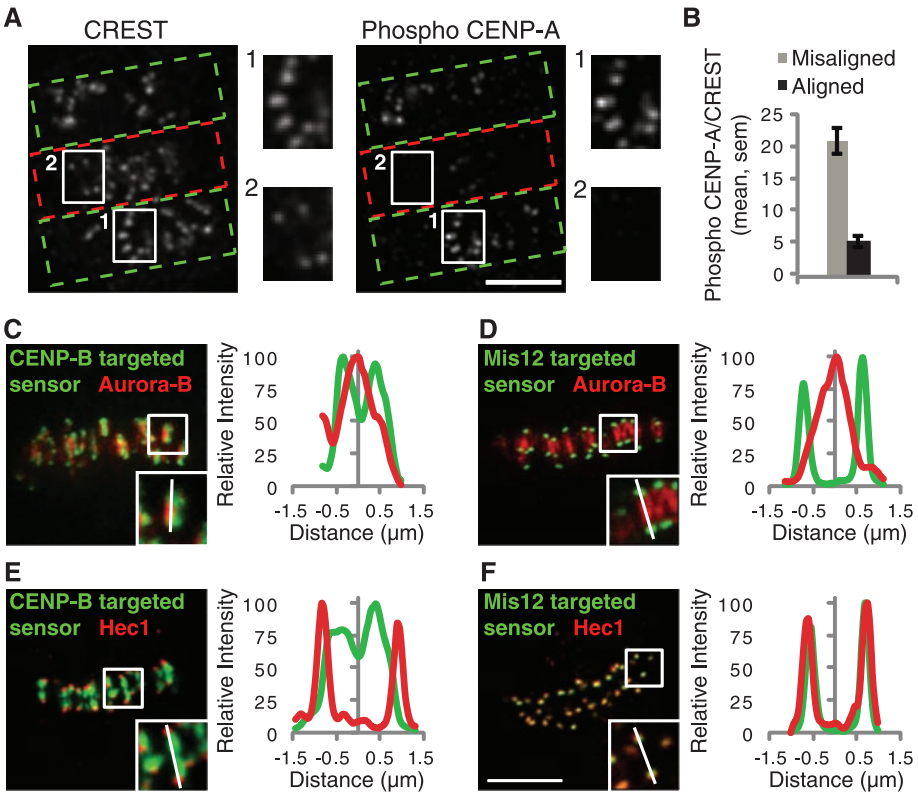


Fig. 1. Phosphorylation of a kinetochore Aurora B substrate depends on the microtubule attachment state. (**A** and **B**) HeLa cells with both incorrect [(A), green boxes] and bi-oriented [(A), red boxes] attachments were fixed and stained for kinetochores (CREST) and phospho-CENP-A. Insets show stronger phospho-CENP-A staining on unaligned (1) versus aligned (2) kinetochores. The phospho-CENP-A/CREST ratio was calculated at individual aligned (N = 146) and unaligned (N = 89) kinetochores from multiple cells (B). (**C** to **F**) HeLa cells expressing either the CENP-B–targeted [(C) and (E)] or Mis12-targeted [(D) and (F)] sensor were fixed and stained for either Aurora B [(C) and (D)] or Hec1 [(E) and (F)] as markers for the inner centromere and outer kinetochore, respectively. YFP emission (green) shows sensor localization relative to Aurora B or Hec1 (red). Insets show individual centromere pairs used for line scans. Scale bars, 5 μm. Error bars, mean ± SEM.

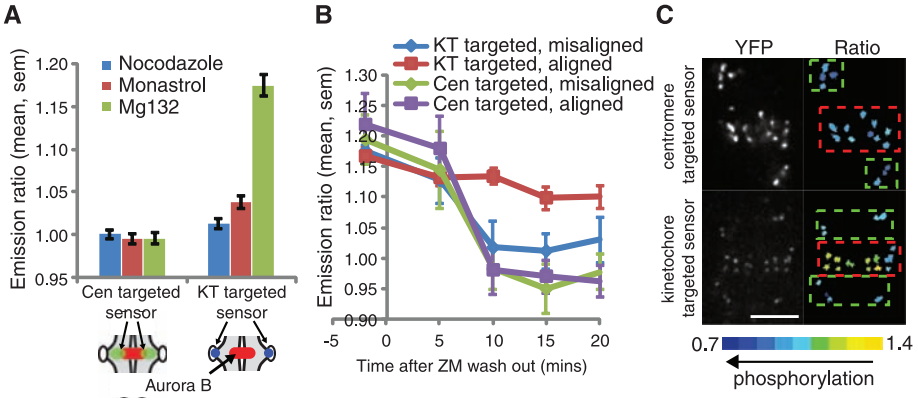


Fig. 2. Phosphorylation of Aurora B substrates depends on spatial separation of kinase from substrate. HeLa cells expressing either the kinetochore (KT)–targeted or centromere (Cen)–targeted sensors were treated as indicated and imaged live. (**A**) Cells were treated with nocodazole, monastrol, or MG132, and the YFP/TFP emission ratio was calculated. An increase in emission ratio indicates dephosphorylation. Each bar represents an average over multiple cells, ≥30 kinetochores analyzed per cell. The *y* axis is scaled to include the range of the data. Schematic shows sensor localization relative to Aurora B, as in fig. S1A. (**B** and **C**) Cells were treated as in Fig. 1A, then imaged live before and after ZM447439 washout. The emission ratio was calculated at aligned and unaligned centromeres/kinetochores at the indicated time points (B). Each point represents ≥14 kinetochores/centromeres analyzed over multiple cells. Images (C) show sensor localization (YFP) and a color-coded representation of the emission ratio 10 min after ZM447439 washout. Boxes indicate aligned (red) and unaligned (green) centromeres/kinetochores. Scale bar, 5 μm. Error bars, mean ± SEM.

[1]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Department of Medical Oncology, University Medical Center, 3584 CG Utrecht, Netherlands.

*Present address: Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA.
†These authors contributed equally to this work.
‡To whom correspondence should be addressed. E-mail: lampson@ sas.upenn.edu

mean ± SD) or monastrol (0.7 ± 0.1 μm) and high in cells treated with MG132 (1.4 ± 0.3 μm) (fig. S1, D and E). The centromere-targeted sensor was phosphorylated under all conditions examined, independent of tension, whereas the kinetochore-targeted sensor was phosphorylated when tension was low and dephosphorylated when tension was high (Fig. 2A). Thus, phosphorylation depends on spatial separation of Aurora B from its substrate, rather than on direct regulation of Aurora B kinase activity.

To determine whether phosphorylation dynamics occur on relevant time scales for regulation of kinetochore-microtubule attachments, we followed kinase activity during the correction of attachment errors (*6*). Cells were treated as in Fig. 1A and imaged live after Aurora B

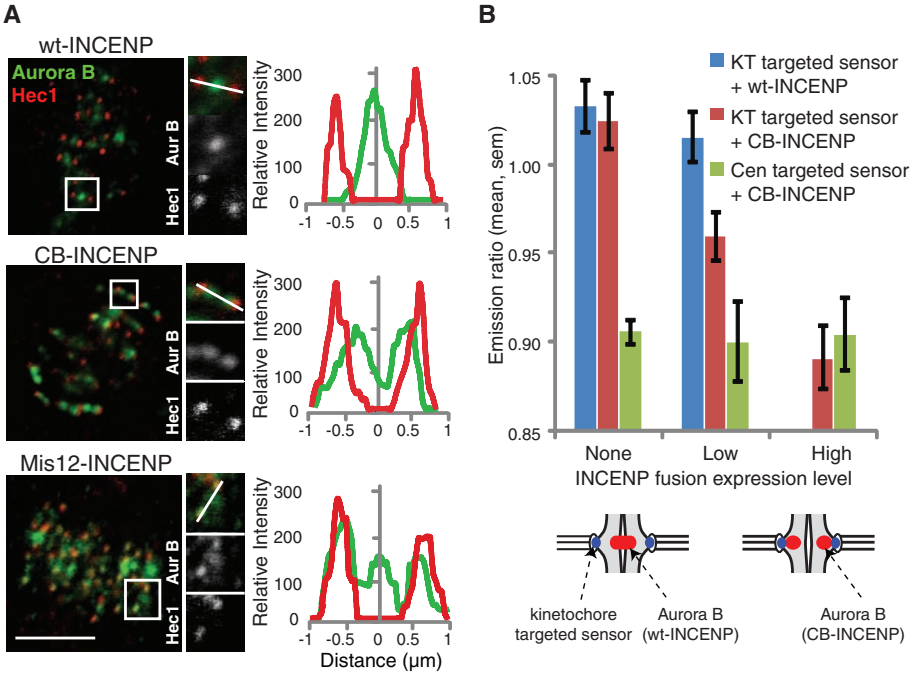


Fig. 3. Positioning Aurora B closer to the kinetochore leads to increased phosphorylation of kinetochore substrates. (**A**) U2OS cells expressing the indicated vesicular stomatitis virus epitope (VSV)–tagged INCENP fusion proteins were fixed and stained for Aurora B (green) and Hec1 (red). Scale bars, 5 μm. (**B**) HeLa cells were transfected with a phosphorylation sensor together with mCherry-tagged wt-INCENP or CB-INCENP as indicated, imaged live, and the YFP/TFP emission ratio was calculated. Cells were grouped based on levels of the INCENP fusion proteins at centromeres, calculated based on mCherry fluorescence intensity (wt-INCENP was not present at as high levels as CB-INCENP). Each bar represents an average over multiple cells, ≥30 kinetochores analyzed per cell. Error bars, mean ± SEM.

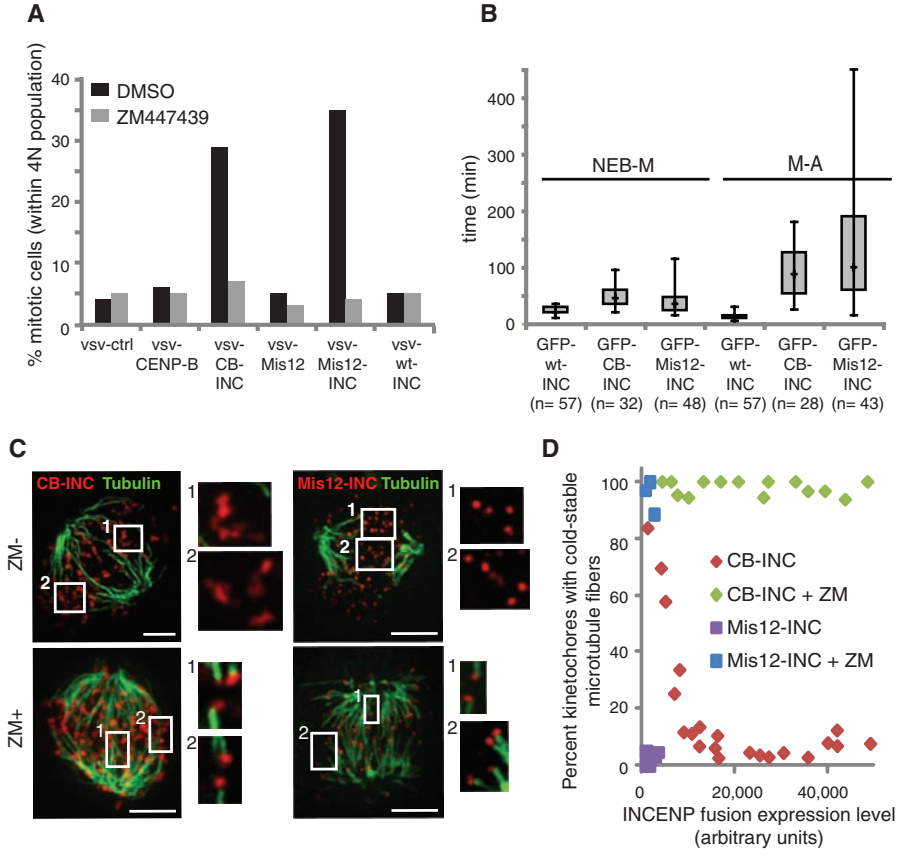


Fig. 4. Positioning Aurora B closer to the kinetochore activates the spindle checkpoint and destabilizes kinetochore-microtubule attachments. (**A**) U2OS cells expressing the indicated proteins were released from a G1/S block with or without ZM447439 for 14 hours. The mitotic index was determined by propidium iodide/MPM2 monoclonal antibody labeling and fluorescence-activated cell sorting analysis. One representative experiment out of three is shown. (**B**) Box-and-whisker plots (median, interquartile range, full range) showing time from nuclear envelope breakdown to metaphase (NEB-M) and from metaphase to anaphase (M-A) of U2OS cells expressing the indicated INCENP (INC) fusion proteins. (**C** and **D**) HeLa cells expressing VSV-tagged CB-INCENP or Mis12-INCENP were treated with or without ZM447439, then analyzed for cold-stable microtubules (green) and VSV immunofluorescence (red). Brightness of VSV-Mis12-INCENP staining is enhanced relative to VSV-CB-INCENP so that it is visible. The percent of kinetochores with cold-stable microtubules was determined and plotted versus expression level of the INCENP fusion proteins (D). Each data point represents >80 kinetochores from one cell.

reactivation, and we measured phosphorylation dynamics at aligned and unaligned kinetochores and centromeres. The centromere-targeted sensor was rapidly phosphorylated at both aligned and unaligned centromeres within 10 min of removing the Aurora B inhibitor (Fig. 2B). The kinetochore-targeted sensor was also rapidly phosphorylated at unaligned kinetochores, but it remained dephosphorylated at aligned kinetochores (Fig. 2B). Color-coded images of the yellow fluorescent protein/teal fluorescent protein (YFP/TFP) emission ratio showed that the kinetochore-targeted sensor was selectively phosphorylated at unaligned kinetochores 10 min after inhibitor washout, whereas the centromere-targeted sensor was phosphorylated at all centromeres (Fig. 2C and fig. S1F). Thus, Aurora B is active at both correct and incorrect attachments, and differential phosphorylation of kinetochore substrates may depend on distance from the kinase.

The spatial separation model predicts that positioning Aurora B closer to the kinetochore should lead to increased phosphorylation of kinetochore substrates in metaphase. Aurora B localizes to the inner centromere through interactions with the inner centromere protein (INCENP) (*21*). We manipulated the position of Aurora B by creating INCENP fusion proteins in which the inner centromere-targeting domain of INCENP is replaced either with the centromere-targeting domain of CENP-B (CB-INCENP) (*17*, *22*) or with full-length Mis12 (Mis12-INCENP) (fig. S2A). Endogenous Aurora B localized closer to or at the kinetochore in cells expressing CB-INCENP or Mis12-INCENP, respectively (Fig. 3A). The extent of Aurora B redistribution increased with the expression level of CB-INCENP (fig. S2C), which suggests that CB-INCENP competes with endogenous INCENP for Aurora B binding. To determine the relationship between kinase position and phosphorylation in cells expressing CB-INCENP, we measured phosphorylation of the kinetochore-targeted sensor at chromosomes aligned on the metaphase plate. Phosphorylation increased as the expression level of CB-INCENP increased (Fig. 3B), whereas the average interkinetochore distances (1.4 ± 0.3 μm, mean ± SD) remained constant (fig. S2, D and E), which indicates that the increased phosphorylation was not due to reduced tension. Thus, phosphorylation of a kinetochore substrate depends on the relative position of the kinase.

Because phosphorylation of endogenous Aurora B substrates at the kinetochore reduces microtubule binding (*7*–*10*), the spatial separation model predicts that repositioning the kinase closer to the kinetochore should impair the stabilization of bi-oriented attachments and activate the spindle checkpoint (*2*). Indeed, expression of either CB-INCENP or Mis12-INCENP induced a dramatic mitotic arrest, with a mitotic index >25% versus ~5% in control cells (Fig. 4A and fig. S3, A and B). The mitotic arrest was dependent on Aurora B kinase activity (Fig. 4A and fig. S3A) and on the spindle checkpoint (fig. S3B). Live imaging revealed that the timing of chromosome alignment increased from 20 min (median) in cells expressing wild-type INCENP (wt-INCENP) to 45 or 35 min for CB-INCENP or Mis12-INCENP, respectively. Moreover, the median time spent in metaphase increased substantially from 10 min for wt-INCENP to 87.5 min or 100 min for CB-INCENP or Mis12-INCENP, respectively (Fig. 4B; fig. S3, C and D; and movies S1 to S3). Paired sister chromatids frequently exited and reentered the metaphase plate during the metaphase arrest (fig. S3E and movie S4), followed by aberrant anaphases with lagging chromosomes (fig. S3, D and E, and movies S1 to S4). In addition, most metaphase cells contained at least one kinetochore positive for Mad1 (fig. S4). Thus, microtubules can sufficiently attach to allow congression, but stabilization of bi-oriented attachments is impaired when Aurora B is positioned close to the kinetochore.

To determine directly whether repositioning Aurora B prevents the stabilization of kinetochore microtubules, we examined cold-stable microtubules. Kinetochore fibers remain intact at 4°C, while most other microtubules in the cell depolymerize (*23*). The number of kinetochores with cold-stable attachments decreased as CB-INCENP expression increased, and at high CB-INCENP levels, where the kinetochore-targeted sensor was highly phosphorylated (Fig. 3B), very few kinetochores had cold-stable microtubule fibers (Fig. 4, C and D). Kinetochores also lacked cold-stable microtubule fibers in cells expressing detectable Mis12-INCENP (Fig. 4, C and D), whereas fibers were stable in cells expressing wt-INCENP (fig. S3F). Aurora B inhibition stabilized kinetochore microtubules in cells expressing either CB-INCENP or Mis12-INCENP (Fig. 4, C and D), which indicates that the destabilization depends on increased kinase activity at the kinetochore. Thus, repositioning Aurora B closer to the kinetochore leads to both increased phosphorylation of kinetochore substrates (Fig. 3B) and destabilization of kinetochore microtubules (Fig. 4).

How bi-oriented attachments are selectively stabilized by tension across the centromere is a long-standing question. Our findings support a mechanism linking tension to phosphorylation of Aurora B substrates at the kinetochore and regulation of kinetochore-microtubule stability. In the absence of tension, kinetochore substrates are phosphorylated because they are in close proximity to Aurora B at the inner centromere. Phosphorylation reduces the affinity of these substrates for microtubules (*7*–*10*), so attachments are destabilized. Tension is exerted when spindle microtubules pull bi-oriented sister kinetochores in opposite directions, away from the inner centromere, so that kinetochore substrates are dephosphorylated, which increases affinity for microtubules and stabilizes attachments. Many biological systems that respond to mechanical forces do so through protein conformational changes, such as those induced in mechanosensitive ion channels (*24*). We present another strategy by which physical forces can be converted into biochemical changes in the cell through modulation of the spatial separation of a kinase from its substrate.

References and Notes

1. T. U. Tanaka, *Chromosoma* **117**, 521 (2008).
2. R. B. Nicklas, *Science* **275**, 632 (1997).
3. S. Biggins, A. W. Murray, *Genes Dev.* **15**, 3118 (2001).
4. T. U. Tanaka *et al*., *Cell* **108**, 317 (2002).
5. S. Hauf *et al*., *J. Cell Biol.* **161**, 281 (2003).
6. M. A. Lampson, K. Renduchitala, A. Khodjakov, T. M. Kapoor, *Nat. Cell Biol.* **6**, 232 (2004).
7. I. M. Cheeseman, J. S. Chappie, E. M. Wilson-Kubalek, A. Desai, *Cell* **127**, 983 (2006).
8. J. G. DeLuca *et al*., *Cell* **127**, 969 (2006).
9. C. Ciferri *et al*., *Cell* **133**, 427 (2008).
10. D. R. Gestaut *et al*., *Nat. Cell Biol.* **10**, 407 (2008).
11. S. G. Zeitlin, R. D. Shelby, K. F. Sullivan, *J. Cell Biol.* **155**, 1147 (2001).
12. Materials and methods are available as supporting material on *Science* Online.
13. C. Ditchfield *et al*., *J. Cell Biol.* **161**, 267 (2003).
14. S. Ruchaud, M. Carmena, W. C. Earnshaw, *Nat. Rev. Mol. Cell Biol.* **8**, 798 (2007).
15. P. D. Andrews *et al*., *Dev. Cell* **6**, 253 (2004).
16. B. G. Fuller *et al*., *Nature* **453**, 1132 (2008).
17. A. F. Pluta, N. Saitoh, I. Goldberg, W. C. Earnshaw, *J. Cell Biol.* **116**, 1081 (1992).
18. I. M. Cheeseman, A. Desai, *Nat. Rev. Mol. Cell Biol.* **9**, 33 (2008).
19. T. M. Kapoor, T. U. Mayer, M. L. Coughlin, T. J. Mitchison, *J. Cell Biol.* **150**, 975 (2000).
20. J. C. Waters, R. H. Chen, A. W. Murray, E. D. Salmon, *J. Cell Biol.* **141**, 1181 (1998).
21. R. R. Adams *et al*., *Curr. Biol.* **10**, 1075 (2000).
22. D. M. Eckley, A. M. Ainsztein, A. M. Mackay, I. G. Goldberg, W. C. Earnshaw, *J. Cell Biol.* **136**, 1169 (1997).
23. C. L. Rieder, *Chromosoma* **84**, 145 (1981).
24. A. W. Orr, B. P. Helmke, B. R. Blackman, M. A. Schwartz, *Dev. Cell* **10**, 11 (2006).

25. We thank T. M. Kapoor (Rockefeller University) and R. H. Medema (University Medical Center, Utrecht) for their support during the initiation of these studies. We also thank B. E. Black, G. Kops, and R. H. Medema for critical reading of the manuscript and helpful comments and K. V. Le and T. Chiang for assistance with preparing reagents. This work was supported by grants from the American Cancer Society (IRG-78-002-30, M.A.L.), the National Institutes of Health (GM083988, M.A.L.), the Searle Scholars Program (M.A.L.), the Penn Genome Frontiers Institute (M.A.L.), and the Netherlands Organization for Scientific Research (Vidi 917.66.332, S.M.A.L.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1167000/DC1
Materials and Methods
Figs. S1 to S4
Movies S1 to S4
References

9 October 2008; accepted 6 January 2009
Published online 15 January 2009;
10.1126/science.1167000
Include this information when citing this paper.

Geometric Cue for Protein Localization in a Bacterium

Kumaran S. Ramamurthi,[1] Sigolene Lecuyer,[2] Howard A. Stone,[2] Richard Losick[1]*

Proteins in bacteria often deploy to particular places within the cell, but the cues for localization are frequently mysterious. We found that the peripheral membrane protein SpoVM (VM) recognizes a geometric cue when localizing to a particular membrane during sporulation in *Bacillus subtilis*. Sporulation involves an inner cell maturing into a spore and an outer cell nurturing the developing spore. VM is produced in the outer cell, where it embeds in the membrane that surrounds the inner cell but not in the cytoplasmic membrane of the outer cell. We found that VM localized by discriminating between the positive curvature of the membrane surrounding the inner cell and the negative curvature of the cytoplasmic membrane. Membrane curvature could be a general cue for protein localization in bacteria.

Proteins often localize to particular positions within bacteria, sometimes in a dynamic manner. A striking but mysterious example of subcellular localization occurs during spore formation in *Bacillus subtilis* when the protein SpoVM (VM) localizes to a particular patch of membrane (*1*). How VM discriminates between different membrane surfaces in the same cell is unknown.

During sporulation, the cell divides asymmetrically to create mother-cell and forespore compartments. Next, the mother cell engulfs the forespore, enveloping it with inner and outer membranes (Fig. 1A). After engulfment, a protein coat is deposited around the outer forespore membrane (*2*). Coat assembly depends on VM, a 26-residue peptide that is produced in the mother cell (*3*). VM is an amphipathic α helix (*4*) that inserts into the membrane with its long axis parallel to the membrane and its hydrophobic face buried in the lipid bilayer (*5*). During engulfment, VM localizes to the membrane that tracks around the forespore, eventually decorating the entire surface of the forespore, as visualized by its fusion to green fluorescent protein (VM-GFP) (Fig. 1C) (*1*). Proline 9 (P9) (Fig. 1B) is critical for this localization (*1*), because substitution of P9 with alanine (VM^{P9A}-GFP) resulted in localization to both the cytoplasmic and outer forespore membranes (Fig. 1D).

After engulfment, the outer forespore membrane becomes topologically isolated from the cytoplasmic membrane. We wondered if VM would adhere to the outer forespore membrane after isolation. We engineered cells to produce VM-GFP in response to an inducer and triggered synthesis of the fusion protein after engulfment. To monitor topological isolation, we stained the membranes with a membrane-permeating dye, which stains all membranes, and a nonpermeating dye, which can only access the engulfment membrane before membrane fusion (*6*). VM-GFP was localized almost exclusively to the outer forespore membrane even when the forespore was topologically isolated (Fig. 1, G to I). As a control, VM^{P9A}-GFP synthesized after engulfment localized promiscuously (Fig. 1, J to L). Thus, VM-GFP partitions between both membranes, but the wild-type fusion protein is retained in the outer forespore membrane.

What does VM recognize? We wondered whether the localization cue for VM was geometric, because the outer forespore membrane is the only membrane in the mother cell with a positive (convex) curvature. We examined the localization of VM-GFP in cells of a triple mutant [SpoIID/SpoIIM/SpoIIP (D/M/P)] arrested with a straight polar septum (*7*). VM-GFP behaved indiscriminately in such mutant cells, localizing to the cytoplasmic and engulfing membranes (Fig. 2B).

As a further test of the idea that VM recognizes curvature, we took advantage of the fact that at low frequency (~1%) the D/M/P triple mutant develops a fissure in the septum, allowing the membrane to bulge into the mother cell (*8*). VM-GFP should localize to this convex membrane protrusion. Indeed, VM-GFP localized with high selectivity to such bulges (Fig. 2C).

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. [2]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: losick@mcb.harvard.edu

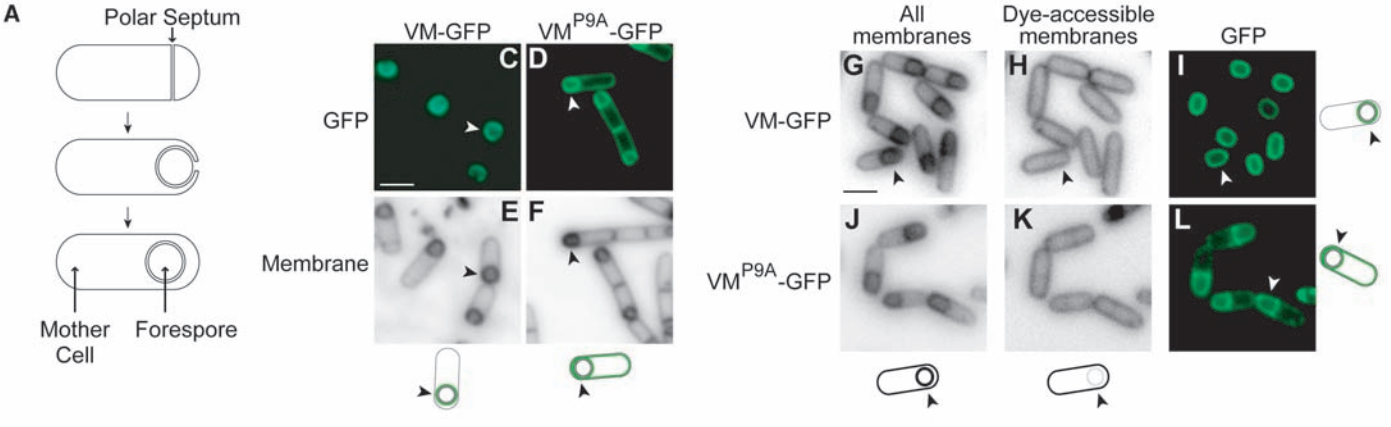


Fig. 1. VM-GFP localizes selectively to the surface of the forespore (*11*). (**A**) Stages of sporulation. (Top) Division creates a mother cell (left) and a forespore (right). (Middle) The mother cell engulfs the forespore. (Bottom) The forespore is pinched off as a protoplast. (**B**) α-helical model of VM. Shown on the left is the hydrophobic membrane-embedded face and on the right a 180° rotation of the helix along its long axis with positively charged residues labeled green. A, alanine; F, phenylalanine; I, isoleucine; K, lysine; L, leucine; M, methionine; P, proline; R, arginine; T, threonine; V, valine; Y tyronine. N refers to the amino terminus. Red square indicates proline 9; gray circles indicate hydrophobic residues on the positively charged face. (**C**) VM-GFP localizes to the surface of the forespore, whereas VM^{P9A}-GFP localizes to all membranes (**D**). (**E** and **F**) Membrane-stained cells in (C) and (D), respectively. Arrowheads identify the cell depicted in the illustrations. (**G** to **L**) Localization of VM-GFP or VM^{P9A}-GFP produced after topological isolation. The membrane surrounding the forespore was stained by a membrane-permeating dye [(G) and (J)] but not by a nonpermeating dye [(H) and (K)]. VM-GFP (I), but not VM^{P9A}-GFP (L), localized selectively to the surface of the forespore. Scale bars, 2 μm.

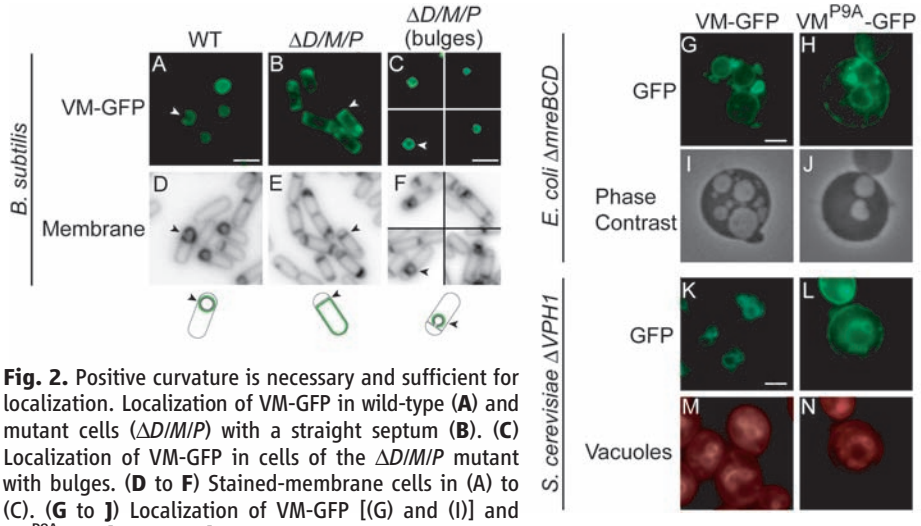


Fig. 2. Positive curvature is necessary and sufficient for localization. Localization of VM-GFP in wild-type (**A**) and mutant cells (Δ*D/M/P*) with a straight septum (**B**). (**C**) Localization of VM-GFP in cells of the Δ*D/M/P* mutant with bulges. (**D** to **F**) Stained-membrane cells in (A) to (C). (**G** to **J**) Localization of VM-GFP [(G) and (I)] and VM^{P9A}-GFP [(H) and (J)] in mutant *E. coli* cells (Δ*mreBCD*) that produce internal vesicles. [(I) and (J)] Vesicles were visualized by phase contrast microscopy. (**K** to **N**) Localization of VM-GFP [(K) and (M)] and VM^{P9A}-GFP [(L) and (N)] in mutant yeast cells (Δ*VPH1*) that produced fragmented vacuoles. [(I) and (J)] Vacuoles were stained with FM4-64. Scale bars, 2 μm.

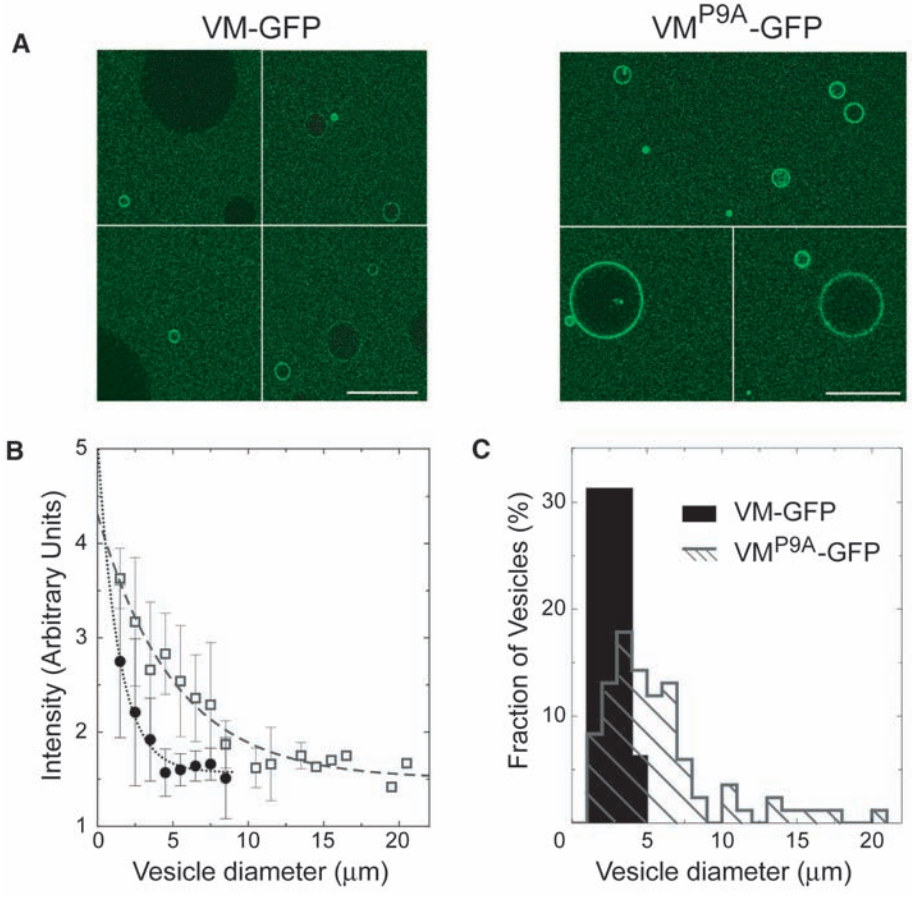


Fig. 3. VM-GFP detects curvature in vitro. (**A**) Confocal fluorescence micrographs of purified VM-GFP (left) or VM^{P9A}-GFP (right) incubated with phospholipid vesicles. Scale bars, 20 μm. (**B**) Membrane fluorescence intensity as a function of vesicle diameter. Vesicles were incubated with similar concentrations of VM-GFP (solid circles) or VM^{P9A}-GFP (open squares). The data points are mean ± SD measured on *n* vesicles (*n* = 3 to 18) and were fit with an exponential decay (*11*). (**C**) Size distribution of fluorescent vesicles whose intensity is at least 2.5-fold greater than the background.

To test the curvature hypothesis, we examined the localization of VM-GFP in a heterologous host that also exhibits convex membranes, an *Escherichia coli* mutant lacking the cytoskeletal protein MreB (*9*) that forms internal vesicles similar in size to the forespore. In mutant cells, VM-GFP localized almost exclusively to the surface of the vesicles (Fig. 2G). As a control, VM^{P9A}-GFP not only localized to the surface of vesicles but also to the cytoplasmic membrane (Fig. 2H). We also examined the localization of VM-GFP in a mutant of *Saccharomyces cerevisiae* that produces fragmented vacuoles whose sizes are again similar to that of the forespore (*10*). In these cells, VM-GFP localized primarily to the surface of vacuoles, whereas VM^{P9A}-GFP localized to the concave periphery of the cells as well (Fig. 2, K and L). VM-GFP and VM^{P9A}-GFP were stable and was produced at similar levels in *B. subtilis*, *E. coli*, or *S. cerevisiae* (fig. S1). Taken together, VM appears to respond to a geometric cue rather than a *B. subtilis*–specific feature of the membrane.

If so, then VM should bind preferentially to any phospholipid bilayer with a curvature similar to that of the outer forespore membrane. We asked whether purified VM-GFP would adsorb to the surface of phosphoplipid vesicles and whether it would favor vesicles similar in size to the forespore. We prepared by means of electroformation a heterogeneously sized population of unilamellar phospholipid vesicles, ranging in diameter from approximately 1 to 30 μm, with a distribution that peaked at 8 μm (*11*). We incubated the vesicles with purified VM-GFP harboring a C-terminal polyhistidine tag (VM-GFP-His$_6$) and examined the distribution of fluorescence with confocal laser microscopy. VM-GFP preferentially adsorbed to the smallest observable vesicles in the population (Fig. 3A). In contrast, VM^{P9A}-GFP displayed a diminished preference for smaller vesicles and adsorbed to the surfaces of even large vesicles (Fig. 3A). The fluorescent signal resulted from surface-localized VM-GFP-His$_6$, and a label introduced into the lipids resulted in uniform and equally fluorescent vesicles (*11*).

The concentration of surface-adsorbed VM-GFP decreased as the diameter of the vesicles increased (Fig. 3B). Thus, smaller vesicles harbored more VM-GFP per surface area than did larger vesicles. When vesicles were incubated with VM^{P9A}-GFP, however, the decrease of fluorescence intensity with increasing vesicle size was much less pronounced, suggesting that VM^{P9A}-GFP was recruited more readily to larger vesicles (Fig. 3B). Next, we measured the number of vesicles whose fluorescence was at least 2.5 times higher than background fluorescence (Fig. 3C). Strong adsorption of VM-GFP occurred for vesicles less than 5 μm in diameter and peaked near the smallest vesicles in the population at around 2.5 μm in diameter (by comparison, a forespore is approximately 1 μm in diameter). In our analysis, vesicles less than 1 μm in diameter were grossly

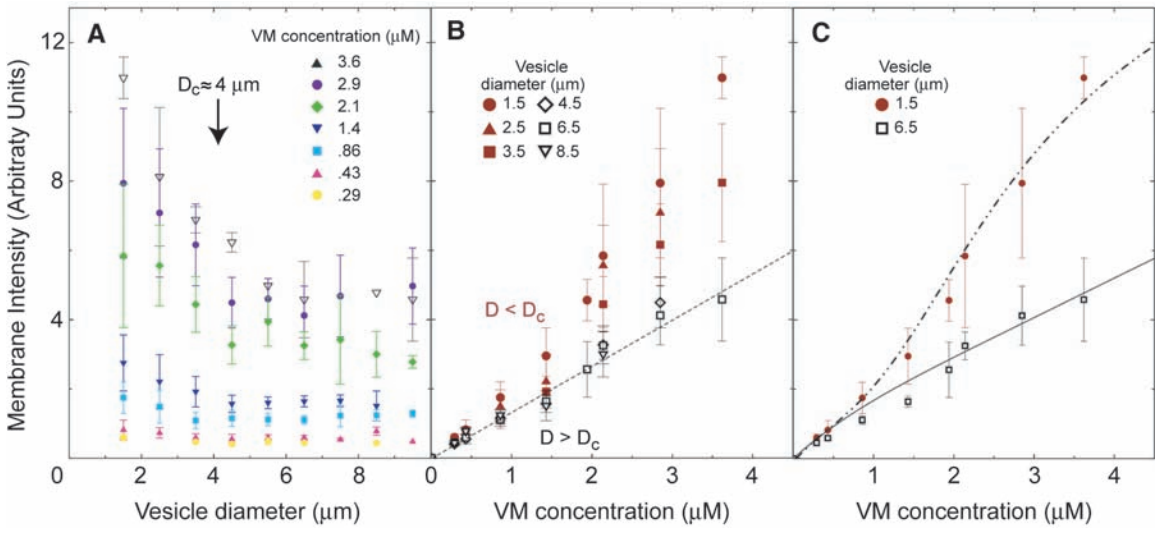


Fig. 4. Preferential adsorption of VM-GFP onto smaller vesicles is concentration-dependent. (**A**) Membrane fluorescence intensity as a function of vesicle diameter for increasing bulk concentrations of VM-GFP. At higher concentrations, vesicles with a diameter below D_c ~ 4 μm were preferentially labeled. (**B**) Membrane fluorescence intensity as a function of VM-GFP concentration for different vesicle diameters (indicated diameters are ±0.5 μm). Above D_c, all data points were similar and well approximated by a linear fit (dotted line), whereas below D_c the isotherms were steeper and deviated from linear. (**C**) Typical behaviors for small (1.5 ± 0.5 μm) and large (6.5 ± 0.5 μm) vesicles and theoretical curves obtained using a model for cooperativity (*11*). Data points are mean ± SD measured on *n* vesicles (*n* = 3 to 26), resulting from three independent experiments.

underrepresented. Thus, we probably overestimated the peak size of vesicles to which VM-GFP was recruited, and VM-GFP may preferentially adsorb onto even smaller vesicles (Figs. 3B and 4A). In contrast to VM-GFP, VM^{P9A}-GFP adsorbed onto vesicles up to 20 μm in diameter, resulting in a broad distribution with a peak at about 5 μm. Substitution of P9 does not simply increase the affinity of VM for membranes. Thus, the adsorption of VM-GFP to membranes is sensitive to curvature and dependent on P9.

How can VM, which is 40 Å in length or less (*12*), be sensitive to the curvature of micrometer-sized spheres? A precedent is adenosine 5′-diphosphate–ribosylation factor guanosine triphosphatase–activating protein 1 (ArfGAP1), a Golgi-associated protein that preferentially associates with yeast vesicles via a stretch of α-helix (*13*, *14*) and is sensitive to curvature. But the vesicles recognized by ArfGAP1 are far smaller (~50 nm) and more highly curved than those recognized by VM. For a 40 Å rod lying flat on the surface of a 1-μm-diameter sphere, the maximum distance between one end of the rod and the surface of the sphere is less than 0.2 Å. It therefore seems improbable that the partitioning of individual VM molecules between differently sized spheres could be influenced by such a small degree of curvature. Several molecules of VM may thus be required to display a collective sensitivity for slightly curved surfaces. We tested whether the preferential adsorption of VM-GFP to smaller vesicles was dependent on the concentration of VM-GFP by measuring membrane fluorescence intensity as a function of vesicle size for different concentrations of VM-GFP. The preferential adsorption of VM-GFP onto smaller vesicles increased with increasing concentrations of protein (Fig. 4A). Moreover, our data suggest a critical value of the vesicle diameter (D_c) of about 4 μm. For all concentrations, above D_c the amount of VM-GFP adsorbed per unit area of membrane did not substantially vary. However, below D_c and at higher concentrations, the adsorbed amount strongly increased with a decreasing vesicle diameter.

We then constructed an "adsorption isotherm," which is a signature of the adsorption mechanism. For a given vesicle diameter, we plotted membrane fluorescence intensity as a function of VM-GFP bulk concentration (Fig. 4B). When the vesicle diameter was larger than D_c, all curves were similar and approximately linear, indicating that the adsorption mechanism was the same and not dependent on membrane curvature. However, below D_c the curves deviated from linear and became progressively steeper as the vesicle diameter decreased, which suggests an adsorption mechanism that involves cooperative interactions (clustering) among VM-GFP. Preliminary theoretical analysis points to clusters consisting of just a few VM molecules (fig. S2) (*11*). VM-GFP molecules do not appear to interact with each other directly (fig. S3). An alternative possibility, however, and one that we favor, is that the insertion of a VM molecule into the membrane indirectly recruits other VM molecules to its vicinity analogous to the clustering of certain phospholipids (*15*) and membrane proteins (*16*, *17*): The energetic cost due to bilayer deformation induced by the insertion of VM could be minimized by the clustering of VM molecules, resulting in an apparent cooperativity that does not involve contact between protein molecules.

Might geometric cues represent a strategy by which proteins localize to particular patches of membrane in bacteria? Convex surfaces, such as that of the forespore, are not typical in bacteria. Nonetheless, some bacteria produce organelles with positively curved membranes, such as photosynthetic vesicles and magnetosomes (*18*, *19*). Perhaps amphipathic α-helices are used to identify the membranes of these organelles for protein localization. A major challenge in bacterial cell biology is to identify cues that recruit proteins to the poles of cells. Conceivably, membrane curvature, in this case extreme negative curvature, is used to identify the inside surface at the pole.

References and Notes

1. C. van Ooij, R. Losick, *J. Bacteriol.* **185**, 1391 (2003).
2. A. O. Henriques, C. P. Moran Jr., *Annu. Rev. Microbiol.* **61**, 555 (2007).
3. P. A. Levin *et al.*, *Mol. Microbiol.* **9**, 761 (1993).
4. R. S. Prajapati, T. Ogura, S. M. Cutting, *Biochim. Biophys. Acta* **1475**, 353 (2000).
5. K. S. Ramamurthi, K. R. Clapham, R. Losick, *Mol. Microbiol.* **62**, 1547 (2006).
6. D. Z. Rudner, Q. Pan, R. M. Losick, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8701 (2002).
7. P. Eichenberger, P. Fawcett, R. Losick, *Mol. Microbiol.* **42**, 1147 (2001).
8. B. Blaylock, X. Jiang, A. Rubio, C. P. Moran Jr., K. Pogliano, *Genes Dev.* **18**, 2916 (2004).
9. F. O. Bendezu, P. A. de Boer, *J. Bacteriol.* **190**, 1792 (2008).
10. E. S. Seeley, M. Kato, N. Margolis, W. Wickner, G. Eitzen, *Mol. Biol. Cell* **13**, 782 (2002).
11. Materials and methods are available as supporting material on *Science* Online.
12. D. J. Barlow, J. M. Thornton, *J. Mol. Biol.* **201**, 601 (1988).
13. R. Parthasarathy, J. T. Groves, *Soft Matter* **3**, 24 (2007).
14. J. Bigay, J. F. Casella, G. Drin, B. Mesmin, B. Antonny, *EMBO J.* **24**, 2244 (2005).
15. K. C. Huang, R. Mukhopadhyay, N. S. Wingreen, *PLoS Comput. Biol.* **2**, e151 (2006).
16. M. Goulian, R. Bruinsma, P. Pincus, *Europhys. Lett.* **22**, 145 (1993).
17. K. S. Kim, J. Neu, G. Oster, *Biophys. J.* **75**, 2274 (1998).

18. C. Mackenzie *et al.*, *Annu. Rev. Microbiol.* **61**, 283 (2007).
19. A. Komeili, Z. Li, D. K. Newman, G. J. Jensen, *Science* **311**, 242 (2006).
20. We thank R. Gaudet, D. Rudner, C. Schmidt, L. Shapiro, R. Schekman, D. Weitz, and N. Wingreen for discussions and S. Lacefield, B. Scheid, and P. de Boer for advice. This work was supported by the Harvard Nanoscale Science and Engineering Center, a Fulbright grant to S.L., and NIH grant GM18568 to R.L.

Supporting Online Material
www.sciencemag.org/cgi/content/full/323/5919/1354/DC1
Materials and Methods
Figs. S1 to S3
Table S1
References

2 December 2008; accepted 13 January 2009
10.1126/science.1169218

A Kinase-START Gene Confers Temperature-Dependent Resistance to Wheat Stripe Rust

Daolin Fu,[1]*† Cristobal Uauy,[1]*‡ Assaf Distelfeld,[1,2]* Ann Blechl,[3] Lynn Epstein,[4] Xianming Chen,[5] Hanan Sela,[2] Tzion Fahima,[2] Jorge Dubcovsky[1]§

Stripe rust is a devastating fungal disease that afflicts wheat in many regions of the world. New races of *Puccinia striiformis*, the pathogen responsible for this disease, have overcome most of the known race-specific resistance genes. We report the map-based cloning of the gene *Yr36* (*WKS1*), which confers resistance to a broad spectrum of stripe rust races at relatively high temperatures (25° to 35°C). This gene includes a kinase and a putative START lipid-binding domain. Five independent mutations and transgenic complementation confirmed that both domains are necessary to confer resistance. *Yr36* is present in wild wheat but is absent in modern pasta and bread wheat varieties, and therefore it can now be used to improve resistance to stripe rust in a broad set of varieties.

Bread wheat (*Triticum aestivum* L.) provides ~20% of the calories consumed by humankind. The increasing world demand for cereals requires improved strategies to reduce yield losses due to pathogens. Wheat stripe rust, caused by the fungus *Puccinia striiformis* f. sp. *tritici* (PST, table S1), affects millions of hectares of wheat, and virulent races that have appeared within the past decade are causing large yield losses (*1*–*3*). Historically, resistant varieties have provided an economical and "environmentally friendly" method to control stripe rust. Numerous race-specific resistance genes have been deployed by breeders, but each has had limited durability, presumably because of rapid pathogen evolution. In contrast, partial resistance genes (i.e., "slow-rusting") offer a broader spectrum of resistance than race-specific genes; they are generally more effective at adult plant stages and usually confer more durable resistance (*1*). Unfortunately, our understanding of partial resistance to cereal rusts is limited because none of these genes has yet been cloned.

We report here the positional cloning of the high-temperature stripe rust resistance gene *Yr36*. This gene was first discovered in wild emmer wheat (*T. turgidum* ssp. *dicoccoides* accession FA15-3, henceforth DIC) (*4*). Analysis of *Yr36* isogenic lines in different genetic backgrounds confirmed that this gene confers partial resistance to PST under field conditions and is associated with significant yield increases when the pathogen is present. In controlled environments, plants with *Yr36* are resistant at relatively high temperatures (25° to 35°C) but susceptible at lower temperatures (e.g., 15°C) (*4*). *Yr36* resistance, originally discovered in adult plants, has some effectiveness in seedlings at high temperatures (fig. S1). Other high-temperature partial resistance genes have provided durable resistance to stripe rust and are used frequently in wheat breeding programs (*5*–*8*).

To clone *Yr36*, we crossed the susceptible durum wheat variety Langdon (LDN, Fig. 1A) with the resistant isogenic recombinant substitution line RSL65 (Fig. 1B), which carries *Yr36* in a LDN genetic background. We screened a population of 4500 F_2 plants using *Yr36* flanking markers *Xucw71* and *Xbarc136* (*4*) and identified 121 lines with recombination events between these two markers. On the basis of genes from the rice colinear region (*9*), nine polymerase chain reaction (PCR) markers were developed to construct a high-density map of *Yr36* (Fig. 1, C and D, and table S2). With the use of replicated field trials and controlled environment inoculations (tables S3 and S4 and figs. S2 and S3), *Yr36* was mapped to a 0.14-cM

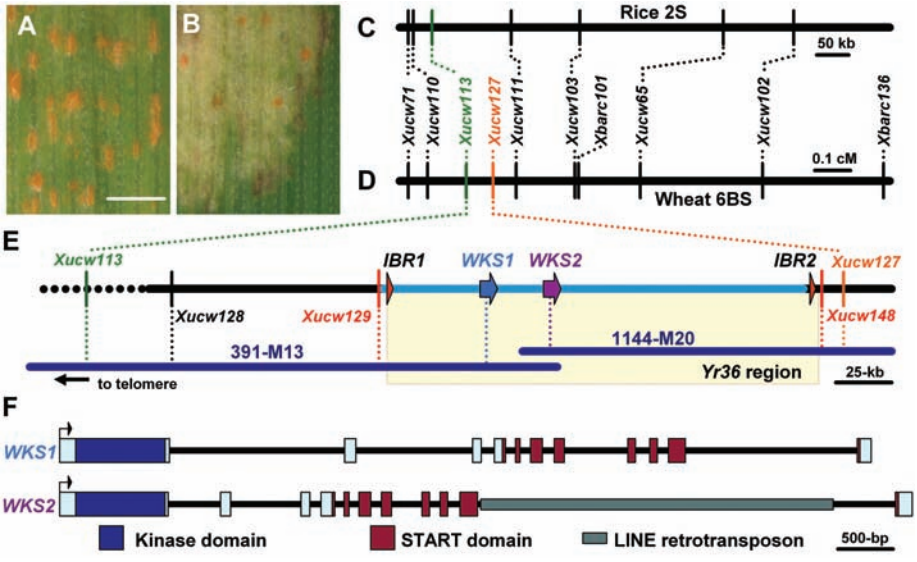


Fig. 1. Map-based cloning of *Yr36*. (**A** and **B**) Phenotype of susceptible parent Langdon with PST sporulation (A) and partially resistant parent RSL65 (B). Scale bar, 1 mm. (**C** and **D**) Genetic maps of colinear regions of rice chromosome 2 (C) and wheat chromosome 6B (D). (**E**) Physical map of the *Yr36* region. Genes are represented by colored arrows and the deleted region in Langdon by a light blue line. (**F**) Structure of the *WKS* genes. Exons are represented by rectangles, and the kinase and START domains are shown in different colors.

[1]Department of Plant Sciences, University of California, Davis, CA 95616, USA. [2]Department of Evolutionary and Environmental Biology, University of Haifa, Haifa 31905, Israel. [3]USDA-ARS, Western Regional Research Center, Albany, CA 94710, USA. [4]Department of Plant Pathology, University of California, Davis, CA 95616, USA. [5]USDA-ARS and Department of Plant Pathology, Washington State University, Pullman, WA 99164, USA.

*These authors contributed equally to this work.
†Present address: Department of Agronomy, Shandong Agricultural University, Tai'an, Shandong 271018, China.
‡Present address: John Innes Centre, Colney, Norwich NR4 7UH, UK.
§To whom correspondence should be addressed. E-mail: jdubcovsky@ucdavis.edu

interval delimited by markers *Xucw113* and *Xucw111* (Fig. 1D).

Screening the RSL65 bacterial artificial chromosome (BAC) library (*10*) with the distal marker *Xucw113* yielded six BACs (fig. S4). BAC ends were used to rescreen the library and extend the contig by chromosome walking. BAC-end marker *Xucw127* (table S2 and fig. S4) was mapped proximal to *Yr36*, thereby completing the physical map (Fig. 1E). BAC clones 391M13 and 1144M20 were sequenced and a contiguous 314-kb sequence including the flanking markers was annotated and deposited in GenBank (EU835198, fig. S5). New markers were developed from the sequence (table S2), and *Yr36* resistance (eight PST races, table S5) was mapped between *Xucw129* and *Xucw148* (0.02 cM).

This region has two pairs of duplicated genes (fig. S5). The first pair includes two short putative genes (*IBR1* and *IBR2*) with an "in between RING finger" domain (IBR, pfam01485). The two other duplicated genes, which we designated *WHEAT KINASE-START 1* and *2* (*WKS1* and *WKS2*, Fig. 1F), encode 86% identical proteins that have a predicted kinase domain followed by a predicted steroidogenic acute regulatory protein-related lipid transfer domain (START, pfam01852). *WKS1*, *WKS2*, and *IBR1* are deleted in the susceptible parent (Fig. 1E). The *WKS* genes were prioritized for functional characterization because their domains have been associated with plant responses to pathogens in other species (*11*–*13*).

Primers specific for *WKS1* and *WKS2* kinase and START domains (table S6) were used to screen a population of 1536 ethyl methanesulfonate (EMS)–mutagenized M_2 lines from the common wheat breeding line UC1041+*Yr36* (*14*). Of the 117 mutants found in the TILLING screen (*15*), we selected for functional characterization six mutants with changes in conserved amino acids in WKS1 (figs. S6 and S7A) and three with premature stop codons in *WKS2* (table S7).

Of the six *WKS1* mutants, five showed susceptible reactions similar to the susceptible UC1041 control line (Fig. 2, A to F, and figs. S8 and S9). In contrast, none of the *WKS2* truncation mutants was susceptible (fig. S8), which suggested that *WKS1* is *Yr36*. Both the kinase (fig. S8) and START domains (fig. S9) were necessary for the resistance response. Laser point scanning confocal microscopy showed that the T6-312 mutant had an unrestricted network of fungal growth, whereas the control line with a functional *WKS1* gene had a resistance response inside the leaf with reduced fungal growth delimited by autofluorescing plant cells (Fig. 2, G to J).

To confirm the identity between *WKS1* and *Yr36*, we transformed the susceptible wheat variety Bobwhite with a 12.2-kb genomic frag-

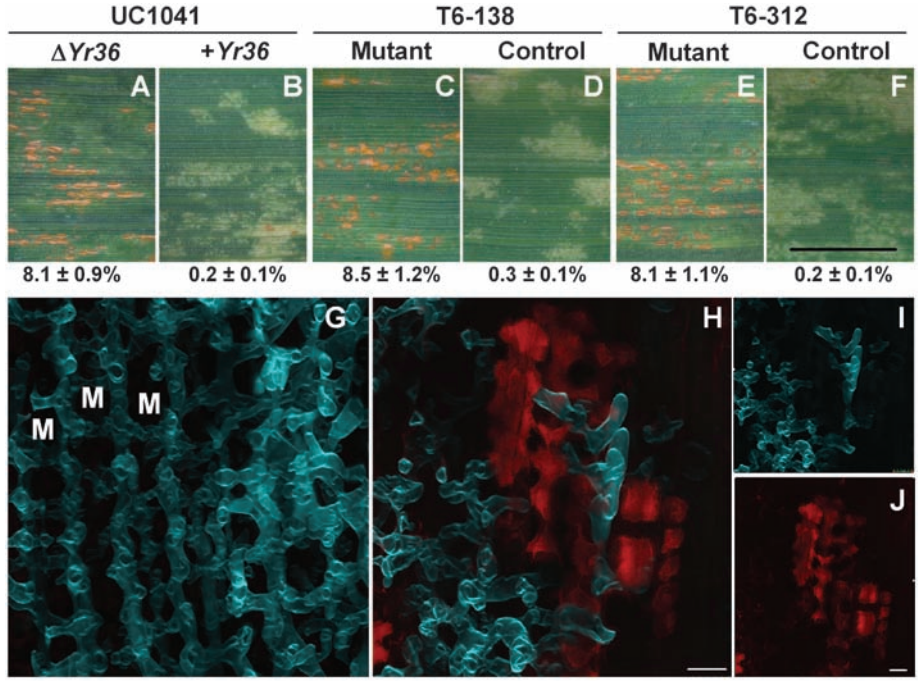


Fig. 2. Functional validation of *Yr36* by mutational analysis. (**A** to **F**) Leaf surfaces 11 days after PST inoculation. Scale bar, 5 mm. Numbers below leaves are average percent leaf area with pustules ± SEM (N = 8, fig. S2). An analysis of variance (ANOVA) of the log-transformed data showed significant differences ($P < 0.01$) between mutant and control lines. (A) UC1041 without *Yr36*. (B) UC1041+*Yr36* isogenic line used for mutagenesis. (C and E) Lines T6-138 and T6-312 with homozygous mutations in the *WKS1* kinase domain. (D and F) Sister lines without the mutations. These and additional mutant lines are described in table S7 and figs. S6 to S9. (**G** and **H**) A dual-channel, confocal microscopic z-series inside a wheat leaf 13 days after PST inoculation. Scale bar, 20 μm. The fungus stained with Uvitex 2B (Polysciences Inc., Warrington, PA; false-color blue) and autofluorescing wheat leaf cells (false-color red) are visible. (G) The susceptible T6-312 mutant has an extensive mycelial network in which each (invisible) plant mesophyll cell (selected cells shown as M) is encircled by a hypha. (H) The T6-312 control line has a poorly developed fungal network surrounded by autofluorescent mesophyll cells that presumably were involved in the resistance response. (**I** and **J**) Separate channels of (H). Scale bar, 20 μm.

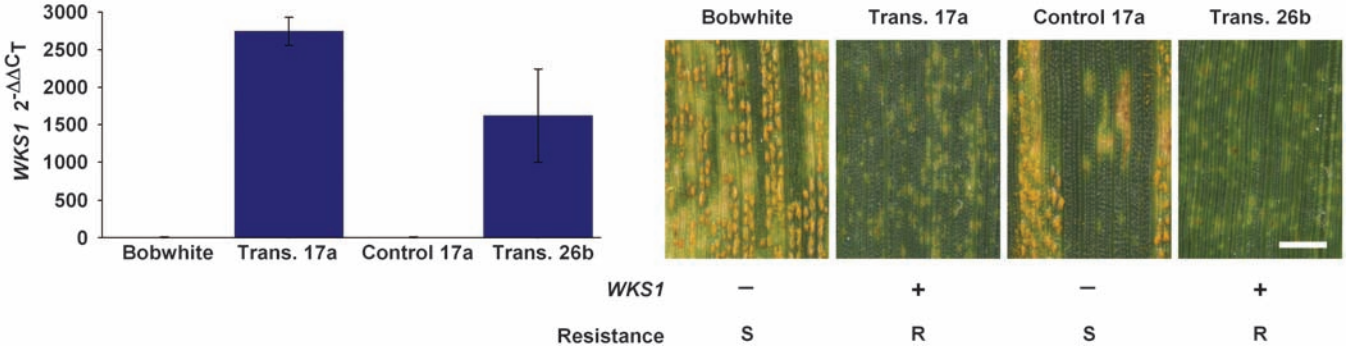


Fig. 3. *WKS1* transcript levels and resistance phenotype in transgenic wheat plants. Left panel: Average *WKS1* transcript levels (±SEM) in independent transgenic events 17a (five plants) and 26b (seven plants) were determined by quantitative reverse transcription (RT) PCR. Negative controls are the untransformed variety Bobwhite and the average of three T_1 sister lines of 17a without the transgene. Right panel: Leaf phenotypes (S, susceptible; R, resistant). Scale bar, 2 mm. Southern blots and transcription profiles of individual T_1 plants are shown in fig. S10.

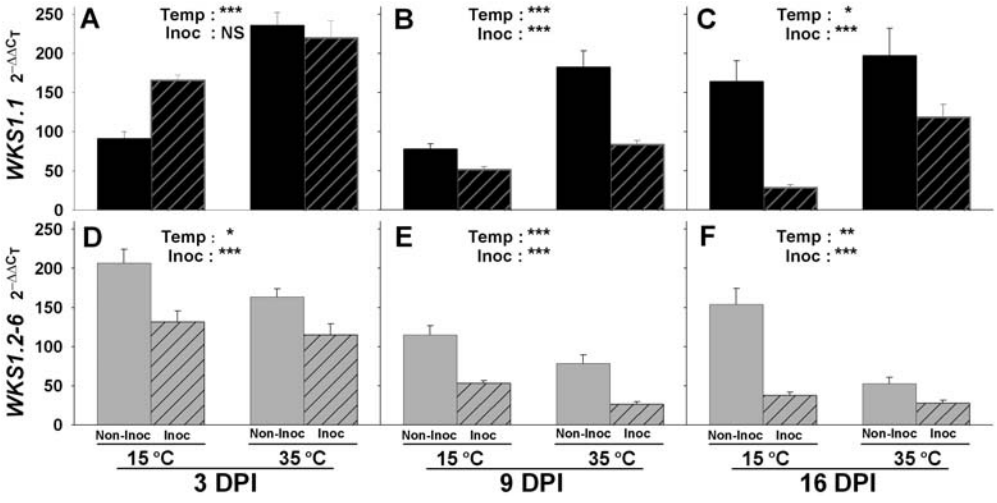


Fig. 4. (**A** to **F**) Effect of temperature and PST inoculation on transcript levels of *WKS1* transcript variants WKS1.1 and WKS1.2-6 in RSL65. Quantitative RT-PCR transcripts of WKS1.1 are indicated in black and those of WKS1.2-6 in gray. PST-inoculated plants are indicated by stripes and non-inoculated controls by solid colors. No significant interactions between temperature and inoculation were detected in the individual two-way ANOVAs, except for (A) (significant differences between inoculation classes only at low temperature). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; NS, not significant. Each data point is an average based on six replicates (±SEM). Overall ANOVAs are presented in table S8; comparisons between WKS1.1 and WKS1.2-6 transcript levels are shown in fig. S12.

ment that includes the complete *WKS1* coding and flanking regions (*14*). Only two of the nine independent T_1 transgenic lines had complete *WKS1* transcripts, and they were both resistant to stripe rust (Fig. 3 and fig. S10), thereby demonstrating that *WKS1* is *Yr36*.

The cloning and sequencing of 56 full-length *WKS1* cDNAs revealed six alternative transcript variants (WKS1.1 to 1.6, fig. S11). WKS1.1 encodes a complete WKS1 protein, whereas the other five (WKS1.2 to WKS1.6, henceforth WKS1.2-6) lack exon 11 and encode proteins with truncated START domains. Some of the missing amino acids are well conserved across the plant kingdom (fig. S7B). Quantitative PCR showed that even the lowest transcript levels of WKS1.1 and WKS1.2-6 are only one-third of those of *ACTIN*, indicating relatively high transcript levels. Overall, high temperature up-regulates WKS1.1 (Fig. 4, A to C, and fig. S12) and down-regulates WKS1.2-6 (Fig. 4, D to F, and fig. S12) ($P < 0.0001$, table S8).

PST inoculation consistently down-regulated WKS1.2-6 across temperature and time, but the effect on WKS1.1 transcript levels varied with sampling times (Fig. 4, A to C). Comparisons between WKS1.1 and WKS1.2-6 transcript levels in PST-inoculated plants (fig. S12, A to C) showed no significant differences at low temperature (susceptible response, $P > 0.55$) and significantly higher values of WKS1.1 relative to WKS1.2-6 at high temperature (resistant response, $P < 0.01$) for all 3 days.

The relative increase in transcript levels of the variant with the complete START domain (WKS1.1) at high temperature parallels the observed high-temperature resistance conferred by *Yr36*. START domain proteins in humans are known to play important roles in lipid trafficking, metabolism, and sensing; and their binding with sterols and ceramides results in protein conformational changes [reviewed in (*16*)]. If the putative WKS1 START domain has the ability to bind lipids from PST (or redirected by PST) at high temperature and change its conformation, this may cause the kinase domain to initiate a signaling cascade leading to the observed programmed cell death (Fig. 2 and fig. S8). The WKS1 serine-threonine kinase domain (pfam00069) was confirmed to have kinase activity (fig. S13).

The combination of the kinase and START domains in WKS1 apparently is the result of a novel domain shuffling, because these two domains are not found together in other organisms (*14*). The most similar protein in *Arabidopsis* to the putative WKS1 START domain is EDR2, a protein that negatively regulates plant defense to the powdery mildew pathogen *Golovinomyces cichoracearum* (*12*–*14*). EDR2 has PH (pfam00169) and DUF1336 (pfam07059) domains, which are absent in WKS1. The WKS1 kinase has high similarity to several *Arabidopsis* WAK-like kinases (fig. S6), but WKS1 lacks the additional domains characteristic of WAK-like kinases (*17*). The WKS1 kinase belongs to the non-RD kinases, which are frequently involved in the early steps of the innate immune response (*11*).

The appearance of this novel gene architecture preceded the origin of the Triticeae, because *WKS1* and *WKS2* were detected in several species from this tribe (table S9 and fig. S14). However, the presence of these two genes was rare among Triticeae species and varied across accessions within those species where they were detected. This suggests that *WKS1* and *WKS2* were lost repeatedly in several grass lineages, including the diploid donors of the A and D genomes of polyploid wheat (table S9). Among 131 wild and cultivated tetraploid wheat accessions, *WKS1* was detected only in wild wheat (24% of accessions), which suggests that *WKS1* was not incorporated into the initial domesticated forms. In hexaploid wheat, *WKS1* was present only in five accessions where the DIC segment was incorporated recently (table S10).

Introgression of *WKS1* in transgenic Bobwhite wheat and in susceptible varieties by backcrossing improved their resistance to stripe rust (*4*). This indicates either that WKS1 is sufficient to improve resistance, or that WKS1 can trigger intermediate genes still present in these varieties that initiate the hypersensitive response. Because *WKS1* is absent from almost all modern commercial varieties of pasta and bread wheat (table S10), the introgression of *Yr36* could have a broad impact in improving resistance to this pathogen. *Yr36* resistance has remained effective against the numerous stripe rust races present in California (2004–2008 field tests) and to all races tested so far in controlled environments (table S5). Moreover, *Yr36* has improved resistance in a variety carrying the partial resistance gene *Yr18* (*4*), which suggests that pyramiding appropriate combinations of partial resistance genes may provide adequate resistance against this pathogen. The discovery of different proteins and resistance mechanisms for the partial resistance genes *Yr36* and *Yr18/Lr34* (*18*) suggests that this type of resistance may involve a heterogeneous group of genes and mechanisms.

References and Notes

1. R. P. Singh, H. M. William, J. Huerta-Espino, G. Rosewarne, paper presented at the 4th International Crop Science Congress, Brisbane, Australia, 26 September to 1 October 2004 (www.cropscience.org.au/icsc2004/symposia/3/7/141_singhrp.htm).
2. X. M. Chen, *Aust. J. Agric. Res.* **58**, 648 (2007).
3. A. M. Wan, X. M. Chen, Z. H. He, *Aust. J. Agric. Res.* **58**, 605 (2007).
4. C. Uauy *et al.*, *Theor. Appl. Genet.* **112**, 97 (2005).
5. A. Qayoum, R. F. Line, *Phytopathology* **75**, 1121 (1985).
6. H. S. Shang, *Sci. Agri. Sin.* **31**, 46 (1998).
7. R. F. Line, *Annu. Rev. Phytopathol.* **40**, 75 (2002).
8. Q. Ma, H. S. Shang, *J. Plant Pathol.* **86**, 19 (2004).
9. A. Distelfeld *et al.*, *Funct. Integr. Genomics* **4**, 59 (2004).
10. A. Cenci *et al.*, *Theor. Appl. Genet.* **107**, 931 (2003).
11. C. Dardick, P. Ronald, *PLoS Pathog.* **2**, e2 (2006).
12. D. Tang, J. Ade, C. A. Frye, R. W. Innes, *Plant J.* **44**, 245 (2005).
13. S. Vorwerk *et al.*, *BMC Plant Biol.* **7**, 35 (2007).
14. See supporting material on *Science* Online.
15. C. M. McCallum, L. Comai, E. A. Greene, S. Henikoff, *Nat. Biotechnol.* **18**, 455 (2000).
16. F. Alpy, C. Tomasetto, *J. Cell Sci.* **118**, 2791 (2005).
17. J. A. Verica, Z. H. He, *Plant Physiol.* **129**, 455 (2002).

18. S. G. Krattinger *et al.*, *Science* **323**, 1360 (2009); published online 19 February 2009 (10.1126/science.1166453).
19. Supported by USDA-CSREES grants 2005-00975 and 2006-55606-16629 and by U.S. and Israel BARD Fund grant US-4024-07. We thank C. Li, F. Paraiso, L. Penman, K. L. Richardson, S. Bassein, M.R. Paddy, Q. Lam, and M. Jindal for technical assistance and R. Thilmony and M. Whalen for critical reading of the manuscript. Gene sequences have been deposited in GenBank with accession numbers EU835198-EU835200, FJ154103-FJ154118, and FJ155069-FJ155070.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1166289/DC1
Materials and Methods
SOM Text
Figs. S1 to S14
Tables S1 to S10
References

23 September 2008; accepted 19 December 2008
Published online 19 February 2009;
10.1126/science.1166289
Include this information when citing this paper.

A Putative ABC Transporter Confers Durable Resistance to Multiple Fungal Pathogens in Wheat

Simon G. Krattinger,[1]* Evans S. Lagudah,[2]*† Wolfgang Spielmeyer,[2] Ravi P. Singh,[3] Julio Huerta-Espino,[4] Helen McFadden,[2]‡ Eligio Bossolini,[1]§ Liselotte L. Selter,[1] Beat Keller[1]†

Agricultural crops benefit from resistance to pathogens that endures over years and generations of both pest and crop. Durable disease resistance, which may be partial or complete, can be controlled by several genes. Some of the most devastating fungal pathogens in wheat are leaf rust, stripe rust, and powdery mildew. The wheat gene *Lr34* has supported resistance to these pathogens for more than 50 years. *Lr34* is now shared by wheat cultivars around the world. Here, we show that the LR34 protein resembles adenosine triphosphate–binding cassette transporters of the pleiotropic drug resistance subfamily. Alleles of *Lr34* conferring resistance or susceptibility differ by three genetic polymorphisms. The *Lr34* gene, which functions in the adult plant, stimulates senescence-like processes in the flag leaf tips and edges.

Improved control of fungal rust diseases in cereals through breeding varieties with durable rust resistance is critical for world food security. International attention has been recently drawn to the continuing major threat of fungal rust diseases of cereals, highlighting the need for effective and durable sources of rust resistance. The most profitable and environmentally friendly strategy for farmers to control wheat rusts in both the developing and the developed world is to grow genetically resistant wheat varieties.

The wheat gene *Lr34* is associated with resistance to two rust diseases of wheat, leaf rust (caused by *Puccinia triticina*) (Fig. 1, A and B), and stripe rust (*P. striiformis*) (*1*–*3*), as well as powdery mildew (*Blumeria graminis*) (*4*). *Lr34* provides an important source of partial resistance that is expressed in adult plants during the critical grain-filling stage and is most effective in the uppermost leaf, the so-called flag leaf. When deployed with other adult plant resistance genes, near-immunity can be achieved (*5*). Flag leaves of many wheat cultivars containing *Lr34* develop a necrotic leaf tip, a morphological marker described as leaf tip necrosis (Fig. 1C) (*6*, *7*). The gene was first documented in Canada by Dyck although *Lr34*-containing germplasm has been a part of wheat improvement since the early 20th century. Wheat cultivars containing *Lr34* occupy more than 26 million ha in various developing countries alone and contribute substantially to yield savings in epidemic years (*8*).

The *Lr34* gene has remained durable, and no evolution of increased virulence toward *Lr34* has been observed for more than 50 years. This is in contrast to many other rust resistance genes, the so-called gene-for-gene class, that provide resistance to some but not all strains of a rust species (*9*–*12*). Despite the importance of adult plant resistance genes (*13*), no such gene has been cloned to date. Understanding the molecular nature of this class of resistance has important implications for long-term control of rust diseases. Previous studies have localized the codominant gene *Lr34* on the short arm of chromosome 7D between the two markers gwm1220 and SWM10 (*14*, *15*). We further reduced the target interval in a map-based cloning approach based on three high-resolution populations (*16*) (table S1). High-resolution mapping revealed a 0.15-cM target interval for *Lr34* flanked by *XSWSNP3/XcsLVA1* and *XcsLVE17* (Fig. 1D). The 363-kb physical interval containing both flanking markers was fully sequenced in the *Lr34*-containing hexaploid wheat cultivar Chinese Spring (FJ436983). Sequence analysis revealed the presence of a gene-rich island containing eight open reading frames (Fig. 1E) predicted to encode proteins with homologies to a hexose carrier, an ATP-binding cassette (ABC) transporter, two cytochromes P450, two lectin receptor kinases, a cysteine proteinase, and a glycosyl transferase. The latter two genes were interrupted by repetitive elements and were excluded as candidates for *Lr34*. Molecular markers derived from the coding sequences resembling one of the two lectin receptor kinases (SWDEL3), the ABC transporter (SWDEL2/csLVD2), and the hexose carrier (SWDEL1) were cosegregating with *Lr34*.

To determine whether one of these cosegregating genes corresponds to *Lr34*, we examined for sequence differences in their coding regions from the three pairs of +/−*Lr34* parental lines of the mapping populations. Consistent sequence polymorphism between the alleles of all parental pairs was found only in the putative ABC transporter gene. Second, we sequenced locus-specific DNA fragments covering parts of the six candidate genes on two γ-irradiation (m19 and m21) and six sodium azide–induced *Lr34* mutants (2B, 2F, 2G, 3E, 4C, and 4E) that were selected for loss-of-function of the *Lr34* resistance. Each mutant showed sequence alterations in the putative ABC transporter gene (table S2), leading to either splice site mutations resulting in strongly reduced splicing efficiency or mis-splicing (fig. S1), amino acid exchanges, frame shifts, or premature stop codons (Fig. 1F). To test for the presence of additional mutations in the other genes cosegregating with *Lr34*, we sequenced DNA fragments covering 12 to 15 kb of the other five candidate genes and intergenic regions on the six mutants 2B, 3E, 4C, 4E, m19, and m21 without finding any sequence polymorphism. Hence, we can exclude the possibility that the eight independent mutations found in the putative ABC transporter gene are due to a generally very high mutation frequency in these lines, and we conclude that this gene is responsible for conferring the durable *Lr34* disease resistance.

Lr34 cosegregated with partial resistance to adult plant stripe rust (*Yr18*), powdery mildew (*Pm38*), as well as leaf tip necrosis (*Ltn1*). The mutants were more susceptible to leaf rust, stripe rust, and powdery mildew, and they did not show leaf tip necrosis. These observations, based on eight independent mutations within a single putative ABC transporter gene, strongly suggest that the same gene controls resistance based on

[1]Institute of Plant Biology, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland. [2]Commonwealth Scientific and Industrial Research Organisation (CSIRO) Plant Industry, GPO Box 1600, Canberra, ACT, 2601, Australia. [3]International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600 Mexico DF, Mexico. [4]Campo Experimental Valle de Mexico INIFAP, Apdo. Postal 10, 56230 Chapingo, Edo de Mexico, Mexico.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: evans.lagudah@csiro.au (E.S.L.), bkeller@botinst.uzh.ch (B.K.)
‡Present address: Science and Research Division, Department of Innovation, Industry, Science and Research, Industry House Level 6, 10 Binara Street, Canberra, ACT, 2601, Australia.
§Present address: Institute of Plant Sciences, University of Bern, 3013 Bern, Switzerland.

Lr34, *Yr18*, and *Pm38* as well as leaf tip necrosis. Depending on the genetic background, *Lr34* was also shown to confer resistance against stem rust (*17*), and the tested mutants were more susceptible to stem rust than the wild type. Further evidence that the putative ABC transporter gene confers slow-rusting resistance came from the study of the early infection process of leaf rust in seedlings of mutants. It was shown earlier that *Lr34* conferred resistance at the seedling stage to leaf rust at low temperatures (*18*). Analysis of infection processes in seedlings grown at 4° to 8°C revealed differences in resistance response to leaf rust (Fig. 1G). From the fourth week after infection, the colonized area in mutants m19 and m21 was larger relative to the wild type. External symptoms of sporulation were evident in the mutants by the fifth week, whereas the presence of the active *Lr34* gene delayed visible symptoms until after the sixth week after infection.

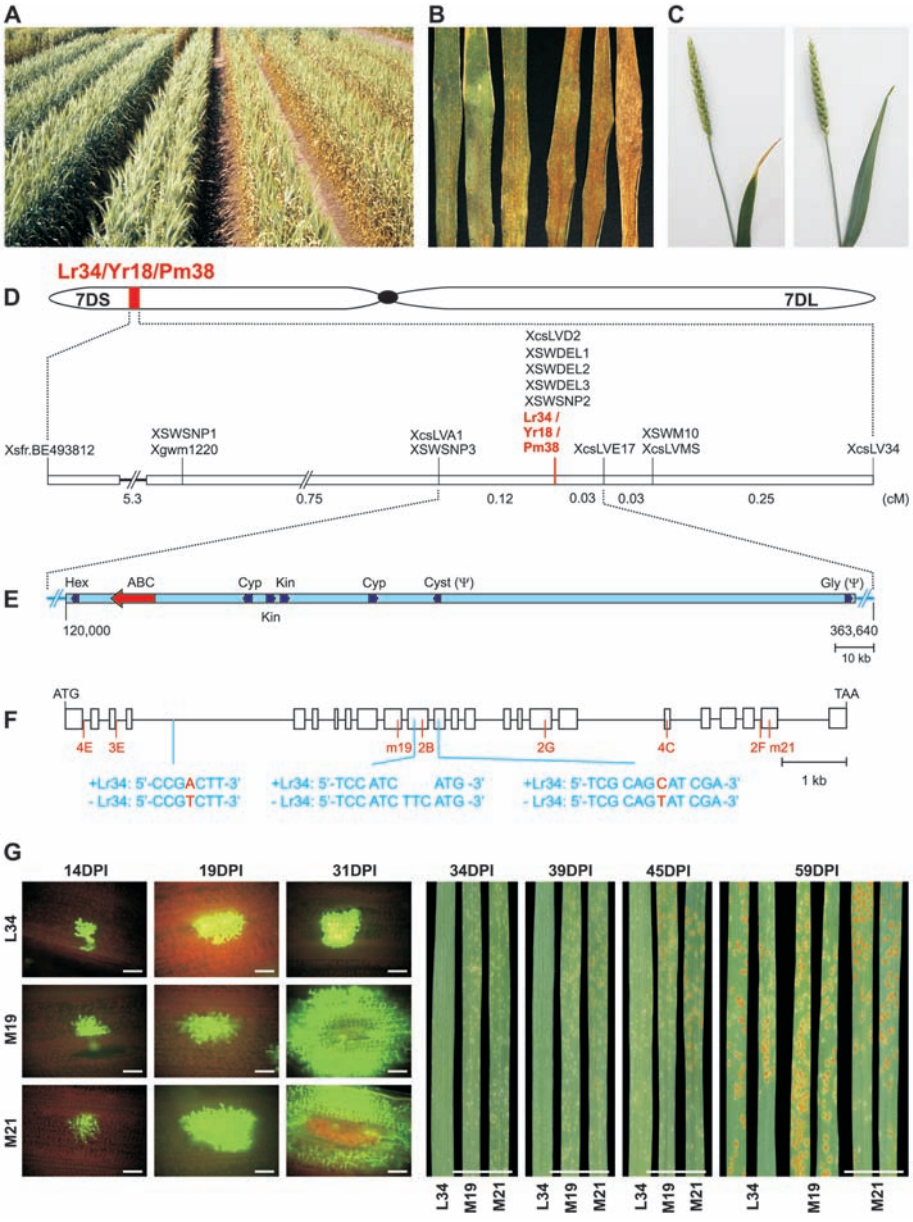


Fig. 1. *Lr34* phenotypes and mapping. (**A**) *Lr34* confers a partial slow-rusting resistance. The images show rows of the resistant selection Jupateco R (left) and of the susceptible near-isogenic line Jupateco S (right) infected with leaf rust in Mexico. (**B**) Progression of leaf rust infection on three successive aging flag leaves of Jupateco R (left) and Jupateco S (right). (**C**) *Lr34* is associated with leaf tip necrosis (*Ltn1*) that can be observed on the resistant near-isogenic line Arina *Lr34* (left) but not on the *–Lr34* Swiss winter wheat cultivar Arina (right). (**D**) Consensus genetic map of *Lr34* based on three high-resolution mapping populations defined a 0.15-cM target interval for *Lr34*. (**E**) The corresponding physical target interval sequenced on the *+Lr34* cultivar Chinese Spring contains eight open reading frames (arrows). Blue lines indicate repetitive regions without genes and numbers refer to the respective positions within the 363-kb interval. Gly, glycosyl transferase; Cyst, cysteine proteinase; Cyp, cytochrome P450; Kin, lectin receptor kinase; ABC, ABC transporter; Hex, hexose carrier; Ψ, pseudogene. (**F**) Gene structure of *Lr34*. Open boxes indicate exons; introns are shown as adjoining lines. Red marks indicate the mutation sites of the eight mutants 2B, 2F, 2G, 3E, 4C, 4E, m19, and m21. The three sequence polymorphisms between susceptible and resistant alleles are indicated in blue. (**G**) Low temperature–induced seedling resistance of *Lr34*. (Left) Microscopic visualization of wheat germ agglutinin stained fungal colonization of mesophyll cells between 14 and 31 days post inoculation (DPI) of Lalbahadur *Lr34*(L34) and derived mutant m19 and m21 seedlings (scale bar, 100 μm). (Right) Progression of leaf rust sporulation on seedling leaf surfaces between 34 and 59 DPI (scale bar, 10 mm).

The nucleotide sequence of *Lr34* spans 11,805 base pairs (bp). Sequencing of the full-length cDNA revealed that *Lr34* consist of 24 exons (Fig. 1F). The predicted 1401–amino acid protein (fig. S2A) belongs to the pleiotropic drug resistance subfamily of ABC transporters. Pleiotropic drug resistance transporters share a common basic structure containing two cytosolic nucleotide binding domains and two hydrophobic transmembrane domains. Fifteen pleiotropic drug resistance–like genes have been identified in the genome of *Arabidopsis*, and 23 members were described in rice. The closest LR34 homolog in rice is OsPDR23, showing 86% amino acid identity. In *Arabidopsis*, the closest homologs are 56% identical to LR34 at the amino acid level (fig. S3). Pleiotropic drug resistance transporters are known to confer resistance to various drugs, but little is known about their substrate specificity (*19*). In *Arabidopsis*, it has previously been reported that PEN3/PDR8 contributes to resistance toward non–host pathogens (*20*). The current model suggests that PEN3 may be involved in translocating toxic compounds derived from glucosinolates into the apoplast (*21*).

We next determined the sequence differences between the *Lr34* alleles in cultivars with or without *Lr34*-based resistance. Comparison of genomic sequences of the putative pleiotropic drug resistance transporter in the *+Lr34* cultivar Chinese Spring and the *–Lr34* French winter wheat cultivar Renan (FJ436985) revealed that the gene was present in both wheat varieties. Only three polymorphisms distinguished the alleles of Chinese Spring and Renan (Fig. 1F). One single-nucleotide polymorphism was located in the large intron 4. The other two sequence differences were located in exons. A deletion of 3 bp (ttc) found in exon 11 in Chinese Spring resulted in the deletion of a phenylalanine residue, whereas a second single-nucleotide polymorphism in exon 12 converted a tyrosine to a histidine in the resistant cultivar. Both sequence differences located in exons affect the first transmembrane domain connecting the two nucleotide binding domains and may alter the structure and substrate specificity of the transporter (fig. S2B). Sequence comparison of 2 kb of the putative *Lr34* promoter regions did not reveal any differences between the two cultivars. Three breeding lineages of *Lr34* in wheat germplasm have been identified:

(i) Far-East germplasm; (ii) spring wheat lines from North and South America that were traced back to *Lr34* cultivar sources developed in Italy; and (iii) winter wheat material in Europe. The same resistance haplotype was found in these three breeding lineages, suggesting a single origin of *Lr34* (Table 1).

Lr34 is more effective during adult growth stage than in seedlings under field conditions. Because leaf tip necrosis develops spontaneously and independent of infection, expression of *Lr34* does not depend on the presence of pathogens. To determine whether the difference in resistance between seedlings and adult plants in lines containing *Lr34* is related to the expression level of *Lr34*, we performed reverse transcription–polymerase chain reaction (RT-PCR) using uninfected leaf material of the near-isogenic lines Thatcher and Thatcher *Lr34*. As expected, the putative pleiotropic drug resistance transporter was expressed at very low levels in 14-day-old seedlings grown at 20°C, whereas the expression level was clearly higher in flag leaves of adult plants before (53-day-old plants) and after the development (63-day-old plants) of leaf tip necrosis (Fig. 2). There was no visible difference in expression between resistant and susceptible plants, which is in agreement with the absence of sequence polymorphisms in the putative promoter region. The unspliced *Lr34* gene product also accumulated in adult plants after 63 days. It has been reported that splicing may be very inefficient in some genes that are expressed at low levels (*22*).

The level of *Lr34*-mediated resistance for leaf rust infection correlated with the development of leaf tip necrosis (fig. S4). In addition, flag leaf tips of *Lr34*-containing cultivars are more resistant than their respective leaf bases. Rubiales and Niks reported that *Lr34* is associated with reduced intercellular hyphal growth but not with a hypersensitive response or papilla formation (*18*). Analysis of microarray studies revealed that genes with similar annotations were up-regulated both in uninfected flag leaves of *Lr34*-containing near-isogenic lines (*23*) and senescing wheat flag leaves (*24*). We therefore tested the hypothesis whether leaf senescence, a highly controlled process (*25*) starting from the leaf tips, may contribute to *Lr34* resistance. The barley cDNA HvS40 is known to be up-regulated during leaf senescence (*26*). Northern blot analysis revealed that HvS40 was highly expressed in uninfected flag leaf tips of Thatcher *Lr34* but not of Thatcher and the six azide-

Table 1. Diagnostic value of the three sequence differences between *+Lr34* and *–Lr34* alleles. The three different origins of *Lr34* are indicated. Some lines without *Lr34* were included for the American and European material. SNP, single-nucleotide polymorphism.

Genotype	Origin	+/– Lr34	A/T SNP intron 4	3-bp deletion exon 11	C/T SNP exon 12
		Far-East wheat germplasm			
Chinese Spring	China	+	A	Deletion	C
RL6058	China	+	A	Deletion	C
Fukuho	Japan	+	A	Deletion	C
		Spring wheat from North and South America			
Mentana	Italy	+	A	Deletion	C
Frontana	Brazil	+	A	Deletion	C
Fronteira	Brazil	–	T	No deletion	T
Ardito	Italy	+	A	Deletion	C
Jupateco R	CIMMYT	+	A	Deletion	C
Jupateco S	CIMMYT	–	T	No deletion	T
Glenlea	Canada	+	A	Deletion	C
Thatcher	Canada	–	T	No deletion	T
Anza	USA	+	A	Deletion	C
Chris	USA	+	A	Deletion	C
Condor	Australia	+	A	Deletion	C
Penjamo 62	CIMMYT	+	A	Deletion	C
Inia 66	CIMMYT	–	T	No deletion	T
Lalbahadur Lr34	CIMMYT	+	A	Deletion	C
Lalbahadur	India	–	T	No deletion	T
		European winter wheat cultivars			
Forno	Switzerland	+	A	Deletion	C
Arina	Switzerland	–	T	No deletion	T
Pegaso	Italy	+	A	Deletion	C
Bezostaja	Russia	+	A	Deletion	C
Kavkaz	Russia	+	A	Deletion	C
Roazon	France	–	T	No deletion	T
Capelle Desprez	France	–	T	No deletion	T
Maris Huntsman	UK	–	T	No deletion	T
Renan	France	–	T	No deletion	T

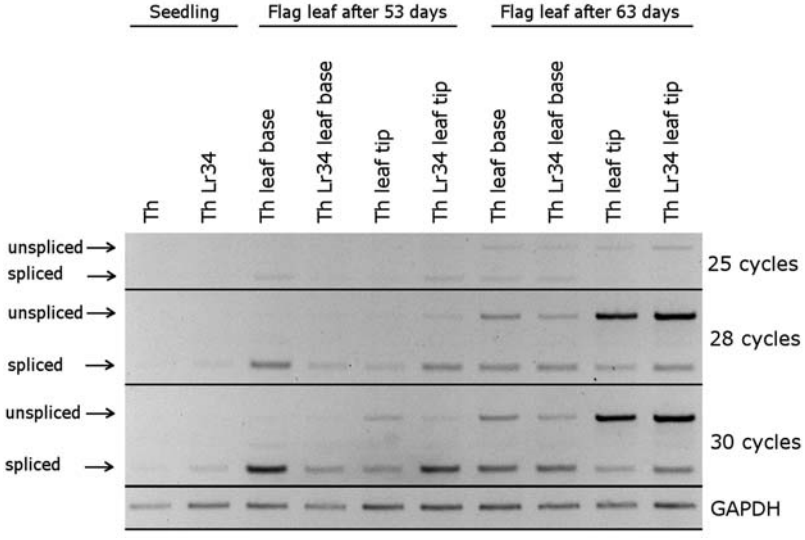


Fig. 2. Expression analysis of *Lr34*. RT-PCR was performed on uninfected leaf material with a primer pair amplifying the first three exons of the gene. Leaves of the near-isogenic lines Thatcher and Thatcher *Lr34* were harvested at the seedling stage after 14 days and of adult flag leaves before (53-day-old plants) and after development (63-day-old plants) of leaf tip necrosis. Expression levels of leaf base and leaf tip were separately determined from adult leaves, because leaf tips of Thatcher *Lr34* are much more resistant than the respective leaf base. The correctly spliced and the unspliced *Lr34* gene product are indicated by arrows. Contamination with genomic DNA can be excluded because the GAPDH control did not amplify an unspliced fragment. Different panels represent independent gel runs and distances between spliced and unspliced band vary. Th, Thatcher; Th *Lr34*, Thatcher *Lr34*; GAPDH, glyceraldehyde 3-phosphate dehydrogenase.

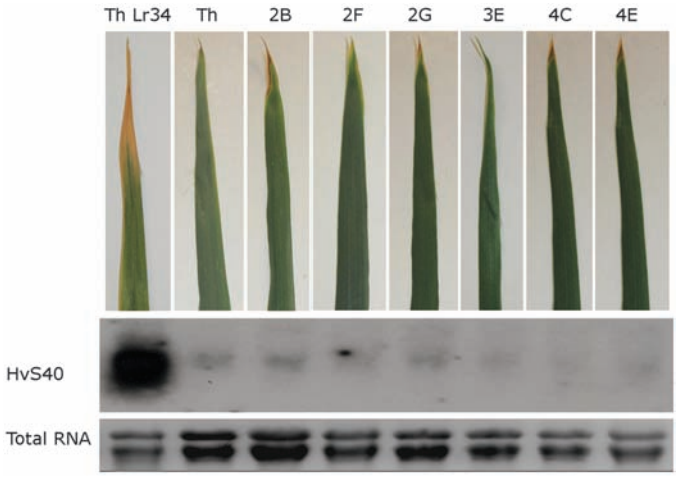


Fig. 3. *Lr34* stimulates senescence-like processes in flag leaves. Northern blot using the probe HvS40 on 63-day-old flag leaf tips (after development of leaf tip necrosis) of the near-isogenic lines Thatcher and Thatcher *Lr34* and the azide-induced *Lr34* mutants 2B, 2F, 2G, 3E, 4C, and 4E. Th, Thatcher; Th *Lr34*, Thatcher *Lr34*.

induced mutants in 63-day-old plants (Fig. 3). Further evidence of *Lr34*'s involvement in leaf senescence came from the analysis of chlorophyll degradation products (*27*). Nonfluorescent chlorophyll catabolites, which are hallmarks of leaf senescence, were detected in the flag leaf tip of Thatcher *Lr34* but not in Thatcher (fig. S5). Thus, it is possible that *Lr34* resistance is the result of senescence-like processes. Alternatively, LR34 may play a more direct role in resistance by exporting metabolites that affect fungal growth, similar to the proposed role for PEN3. *Arabidopsis* PEN3 is a pleiotropic drug-resistance protein and was shown to be involved in non–host resistance to barley powdery mildew. Thus, LR34 and PEN3 belong to the same protein family, raising the possibility of similar defense mechanisms in non–host resistance and durable resistance to an adapted pathogen.

The observation that multiple pathogen resistance in wheat, which comprises *Lr34*, *Yr18*, *Pm38*, as well as the phenotypic marker *Ltn1*, is controlled by the same gene demonstrates the existence in plants of single genetic factors that act durably against several diseases.

References and Notes

1. P. L. Dyck, D. J. Samborski, R. G. Anderson, *Can. J. Genet. Cytol.* **8**, 665 (1966).
2. R. A. McIntosh, *Plant Pathol.* **41**, 523 (1992).
3. R. P. Singh, *Phytopathology* **82**, 835 (1992).
4. W. Spielmeyer, R. A. McIntosh, J. Kolmer, E. S. Lagudah, *Theor. Appl. Genet.* **111**, 731 (2005).
5. R. P. Singh, R. Trethowan, in *Breeding Major Food Staples*, M. Kang, P. M. Priyadarshan, Eds. (Blackwell, Ames, IA, 2007), pp. 109–140.
6. P. L. Dyck, *Crop Sci.* **31**, 309 (1991).
7. R. P. Singh, *Crop Sci.* **32**, 874 (1992).
8. C. N. Marasas, M. Smale, R. P. Singh, *Agric. Econ.* **29**, 253 (2003).
9. B. Keller, S. Bieri, E. Bossolini, N. Yahiaoui, in *Genomic-Assisted Crop Improvement*, R. K. Varshney, R. Tuberosa, Eds. (Springer, Dordrecht, Netherlands, 2007), pp. 103–127.
10. S. Cloutier *et al.*, *Plant Mol. Biol.* **65**, 93 (2007).
11. C. A. Webb *et al.*, *Genetics* **162**, 381 (2002).
12. J. G. Ellis, P. N. Dodds, G. J. Lawrence, *Annu. Rev. Phytopathol.* **45**, 289 (2007).
13. S. N. Panter, D. A. Jones, *Adv. Bot. Res.* **38**, 251 (2002).
14. W. Spielmeyer *et al.*, *Theor. Appl. Genet.* **116**, 481 (2008).
15. E. Bossolini, S. G. Krattinger, B. Keller, *Theor. Appl. Genet.* **113**, 1049 (2006).
16. Materials and methods are available as supporting material on *Science* Online.
17. P. L. Dyck, *Genome* **29**, 467 (1987).
18. D. Rubiales, R. E. Niks, *Plant Dis.* **79**, 1208 (1995).
19. B. Rogers *et al.*, *J. Mol. Microbiol. Biotechnol.* **3**, 207 (2001).
20. M. Stein *et al.*, *Plant Cell* **18**, 731 (2006).
21. U. Lipka, R. Fuchs, V. Lipka, *Curr. Opin. Plant Biol.* **11**, 404 (2008).
22. B. R. Graveley, *Nature* **453**, 1197 (2008).
23. S. H. Hulbert, J. Bai, J. P. Fellers, M. G. Pacheco, R. L. Bowden, *Phytopathology* **97**, 1083 (2007).
24. P. L. Gregersen, P. B. Holm, *Plant Biotechnol. J.* **5**, 192 (2007).
25. P. O. Lim, H. J. Kim, H. G. Nam, *Annu. Rev. Plant Biol.* **58**, 115 (2007).
26. K. Krupinska *et al.*, *Plant Physiol.* **130**, 1172 (2002).
27. S. Hortensteiner, *Annu. Rev. Plant Biol.* **57**, 55 (2006).

28. This work was supported by the Swiss National Science Foundation (B.K.), the Indo-Swiss Collaboration in Biotechnology (B.K.), and the Grains Research and Development Corporation, Australia (E.S.L and R.P.S.). We thank B. Senger for support in field experiments; P. Streckeisen for advice on infection; L. Viccars and S. Chandramohan for technical assistance; M. Ayliffe, M. Talbot, and R. White for assistance with histology; T. Wicker for bioinformatic support; and S. Hortensteiner for nonfluorescent chlorophyll catabolite analysis. GenBank accession numbers: FJ436983 and FJ436985. Patents: CSIRO, Grains Research and Development Corporation, University of Zurich, Australian Provisional Patent Application 2008904364.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1166453/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S3
References

26 September 2008; accepted 9 December 2008
Published online 19 February 2009;
10.1126/science.1166453
Include this information when citing this paper.

MOLECULAR IMAGING

NO SWISS ARMY APPROACH

Imaging of living subjects is of growing importance in biomedical research, particularly in the field of molecular imaging, which represents the cutting edge of these technologies. Utilizing novel technologies and new imaging agents, molecular imaging is allowing scientists to noninvasively visualize specific molecular targets and dynamic events in living animals. Major advances have recently been made in the ability to detect multiple signals simultaneously, enabling visualization of the complex orchestra of molecular communications like never before. Multimodal imaging is gaining momentum, with fusion technologies that combine the power of individual structural and molecular approaches. Ultimately, the hope is that molecular imaging will yield critical information that accelerates the development of diagnostics and therapeutics. **By Ursula Calef**

A big challenge for in vivo molecular imaging is the multiplexing of many signals simultaneously. Sanjiv Gambhir, director of the molecular imaging program at **Stanford University,** says that their latest work on Raman spectroscopy with Raman nanoparticles has helped to solve this problem. His group has made possible, for the first time, the use of Raman optical techniques for small animal imaging by modifying Raman cell microscopy instruments. Gambhir says, "This is, so far, one of the most multiplexable, if you will, strategies we've ever come up with." While he says there is no "Swiss army knife of imaging—where one tool does it all—if you need multiplexing, Raman might be the right way to go."

Raman imaging can be used to study the interplay of many cell populations, e.g., cancer cells and their neovasculature, pre- and posttreatment. Raman spectroscopy measures the inelastic scattering of light "where one out of every 10 million or so photons actually changes its wavelength when it bounces off the target area." Gambhir and colleagues use nanoparticles that enhance this weak effect. The Raman particles "cause some light to shift to one wavelength, some light to shift to a different wavelength, creating unique spectra," explains Gambhir. These spectra are very sharp and tight, and the composition of the particles can be modified to produce unique spectral signatures that are easily distinguished, thereby allowing high multiplexing even on the same cell.

Gambhir uses two kinds of Raman particles: SWNT nanoparticles, or single-walled carbon nanotubes, with intense intrinsic Raman peaks, and SERS-biotags, or surface enhanced Raman scattering active nanoparticles, with a gold core, commercialized by **Oxonica** as Nanoplex Biotags. The molecules with which these particles are functionalized may be varied, so that they can hone in on the specific molecular events being interrogated. Gambhir says that, compared to Raman, the use of different colors of fluorescent probes is more "limited in animals to colors that are red-shifted or in the infrared, because other colors don't do well penetrating through tissue. Even quantum dots," continues Gambhir, "with their many colors, have broad emission peaks. So, if you have four in the same place, it's very hard to separate each of their concentrations."

Raman imaging accomplishes straightforward multiplexing while maintaining exquisitely high sensitivity, and is also semiquantitative. Further, since natural Raman scattering is so low, it results in very little background. For these reasons, Gambhir expects Raman imaging, eventually including tomography, to be a huge future growth area. A main limitation of the approach is that "Raman particles are bigger and not all of them will reach the target site," says Gambhir. While

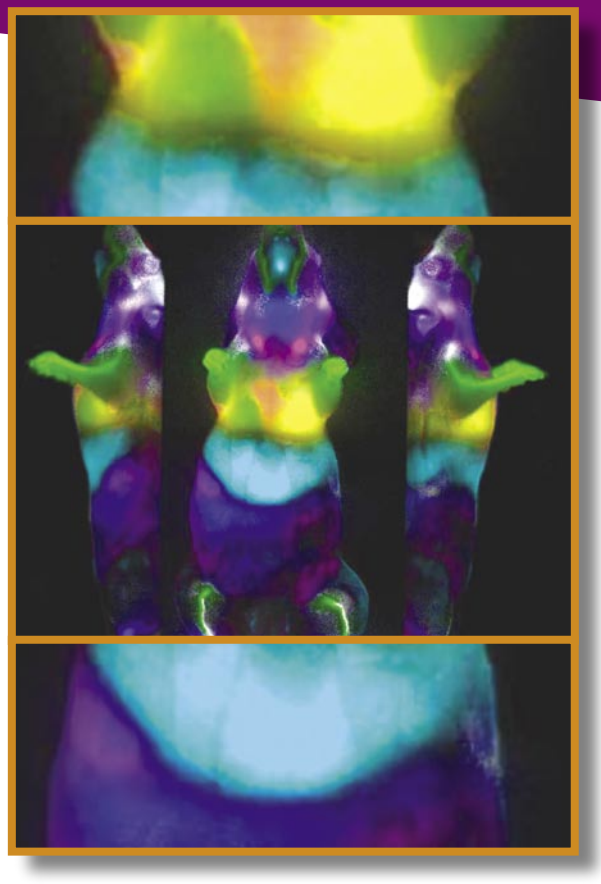


IMAGE COURTESY OF JAMES MANSFIELD, CRI

> "There is no Swiss army knife of imaging, where one tool does it all."

Look for these **Upcoming** Articles

Genomics 1 — April 10

Molecular Diagnostics — May 8

Nucleic Acid Purification and Manipulation— May 15

Inclusion of companies in this article does not indicate endorsement by either AAAS or Science, *nor is it meant to imply that their products or services are superior to those of other companies.*

fluorescent probes might be very small (around 2 nm), "a Raman particle might be tens of nanometers big."

Multiplexing the Colors of the Rainbow

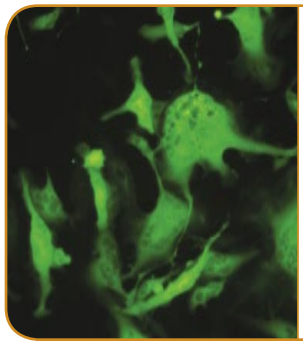
Great strides have similarly been made in multiplexing fluorescent signals. Fluorescence-based imaging is especially useful for in vivo molecular imaging and can follow molecular targets on either a microscopic or macroscopic scale. In 2008, Roger Tsien was awarded the Nobel Prize in Chemistry in part for his work on developing the first of a whole color palette of fluorescent proteins. Additionally, there has been much work on activatable reagents that fluoresce when they react with specific targets.

Tissue autofluorescence has previously limited the sensitivity and utility of fluorescence methods, says James Mansfield, director of multispectral imaging systems at **Cambridge Research & Instrumentation (CRi).** However, recent advances in spectral imaging with CRi's Maestro 2 system have enabled multiplexing of up to five markers in vivo. This technology breaks down light into its spectral components and reads the intensity, "kind of like spectral filtering," explains Mansfield. Using a multispectral camera and sophisticated mathematics, signals are "unmixed," thereby greatly improving contrast and quantitation. Mansfield believes that the Maestro systems are 50–300 times more sensitive than monochrome fluorescence systems, reducing detection limits and facilitating multicolor imaging.

Other multispectral imaging systems include **Caliper Life Sciences'** IVIS Spectrum, which allows the multiplexing of several markers. To select the readable wavelength, IVIS uses large format fixed filters, while the Maestro from CRi uses a tunable filter, explains Stephen Oldfield, senior director of imaging marketing at Caliper Life Sciences. This difference allows for a more variable wavelength range for the Maestro 2 which is designed for fluorescence applications, and the reading of very low light levels by the IVIS Spectrum which provides enhanced sensitivity for bioluminescence. The IVIS can additionally "generate a 3D representation of a fluorescent or a bioluminescent signal using various tomographic techniques," explains Oldfield. Such noninvasive, 3D imaging allows accurate quantitation and localization. Finally, Caliper's IVIS Kinetic includes a real-time, fast imaging capability for studying millisecond events, e.g., perfusion rates of metastatic tumors, says Oldfield.

Photoacoustic Molecular Imaging

Gambhir and colleagues at Stanford University have also recently, for the first time, married photoacoustics to molecular imaging by developing nanoparticle imaging agents. This technique is ideally suited for high multiplexing, high sensitivity small animal imaging. In photoacoustics imaging, Gambhir explains, "Light goes in and interacts with the photoacoustic particle which absorbs the light, heats up slightly, and produces sound. The sound produced, unlike the light, has no trouble penetrating through tissue." This approach therefore offers higher spatial resolution and allows deeper tissue imaging in 3D compared with most optical techniques. The main limitations are that the light going in only penetrates a shallow depth, and sound does not work well near bones or in the lungs. Still, clinical translation of photoacoustic molecular imaging has huge growth potential, e.g., in breast cancer diagnosis, and is currently being commercialized by **Endra** and **Seno Medical.**



Tissue autofluorescence has previously limited the sensitivity and utility of fluorescence methods.

Multimodal Power

While multiplexing lets scientists peek at intricate molecular interactions, advances in multimodal imaging promise to show them precisely where molecular signals are coming from inside animals. Researchers are seeking to combine the strengths of complementary approaches, especially focusing on techniques that combine detailed imaging of anatomy with sensitive imaging of molecular signals. Simon Cherry, director of the Center for Molecular and Genomic Imaging at **University of California, Davis,** and his colleagues have developed a preclinical positron emission tomography/magnetic resonance imaging (PET/MRI) hybrid that for the first time allows simultaneous acquisition of in vivo images. The PET scanner is built into the MRI magnet. PET/computed tomography (CT) and single photon emission computed tomography (SPECT)/CT are well-established approaches that provide sequential scans for molecular and structural information, respectively.

Cherry explains that MRI has major advantages over CT, providing better soft-tissue contrast and allowing you to get "lots of types of information in addition to structure." Importantly, Cherry has put a lot of emphasis on "simultaneous imaging." The challenge in building these integrated systems is to not degrade the performance of either. Fortunately, this is being overcome via better isolation of PET detectors/electronics and the development of newer photodetectors that are less sensitive to magnetic fields. Cherry suggests one good application of PET/MRI would be to study drug delivery into the brain. "You would label the drug so you can see it with PET and coinject gadolinium contrast agent to measure permeability by MRI" to simultaneously assess blood-brain barrier permeability and brain drug concentration, e.g., before and after a treatment to improve drug delivery. Clinical translation is also possible, and **Siemens** has now built the first human PET/MRI prototype systems that are undergoing testing at several sites in the United States and Europe.

One commercially available multimodal research system is **Carestream Molecular Imaging's** Kodak Multispectral FX. It offers four modalities: luminescence, multispectral fluorescence allowing multiplexing of up to four markers, radioisotopic (nuclear), and radiographic (X-ray). This is the first multispectral system that enables "multimodal coregistration of molecular imaging (optical or nuclear) with high resolution X-ray anatomical imaging," explains William McLaughlin, director of research and development at Carestream Molecular Imaging. "All four modes are captured at essentially the same focal plane, so the images can be very precisely overlaid, allowing researchers to more quickly and much more accurately identify the molecular targets of interest." Further multimodal combinations are promised in the future. continued >

In Living Action

Chris Vega, product marketing manager of confocal microscopy at **Leica Microsystems,** asserts that in vivo imaging is exciting because "you can look at the organism living and whole, and watch the naturally occurring processes happening in real time, getting us closer to observing the dynamic processes of life." To accomplish this, Leica Microsystems offers the Leica FCM1000 Confocal Microscope (previously the Cellvizio), which for the first time allows real-time acquisition of data at 12 frames per second. This focal laser device comes with various fibered microprobes to use externally, endoscopically via minor incision, or stereotactically. Its minimal invasiveness permits long-term experiments. Vega explains that with the Leica FCM1000 you can do functional imaging of neurons, cell migratory studies, and "anything where you need to actually look deeper into the animal; at a micro level, this cyber probe allows you to see micro structures in living action."

Super Zoom

Some questions are best addressed with a system that allows you to "essentially start out with a world view and zoom in down to the level of the house, or maybe even the people in the house," describes Vega. He is referring to Leica Microsystems' TCS Large Scale Imaging (LSI) Confocal with super zoom capability. The Leica TCS LSI is the first tool of its kind with "a true confocal scanning system on a macroscope," says Vega. The macroscope provides the large field of view (up to 60 mm), while the optical and confocal zooms, plus additional microscope objectives, allow a range of magnifications from the whole animal down to the cell, all at high resolution. Advanced time lapse software enables longitudinal studies. These features, explains Vega, make LSI ideal for developmental biology applications. "You can watch an organism, such as a zebrafish, go through stages of development," he says.

Molecular Toolbox for a Clearer View

In vivo images are only as good as the imaging agents and methodologies you are using, many of which have recently been added to the toolbox. CRi's Dynamic Contrast Enhancement (DyCE) imaging methodology utilizes time series analysis of images of the injection of a near-infrared (NIR) dye, which allows for the first time the generation of all-optical anatomic maps by charting the path of the dye over time. Mansfield elucidates: "Each organ that has a different time signature can be put into a different color and we use those time signatures to pull out the anatomic information." These anatomical surface maps of major internal organs can be readily coregistered with molecular data.

Other new products include **LI-COR Biosciences'** NIR dyes, such as IRDye 800CW. This dye "excites and emits at the sweet spot for optical imaging" where tissue background interference is lowest, explains D. Michael Olive, vice president of science and technology at LI-COR, and produces a high signal-to-noise ratio with deeper penetration. Carestream Molecular Imaging has two novel small molecule dyes that also excite in the NIR range with very large Stokes shift. McLaughlin says that "excitation and emission peaks are about three times farther apart than typical dyes," significantly improving signal over background. Additionally, Carestream offers fluorochrome-embedded nanospheres that are bright, organic, nontoxic, consistent in size, and have a large number of attachment groups. Finally, the realization of the increased power of fusion technologies is leading to the development of flexible, "multimodal imaging agents" for which different imaging modalities can be used to detect signal, notes McLaughlin.

When Money Matters

Remarkable advances in molecular imaging systems come at a premium. Cash-strapped scientists need cost-efficient alternatives. At under $75,000, "LI-COR's Pearl Imaging System is the most affordable system out there," declares Olive. The Pearl, used for NIR imaging, is a "laser-based system which gives deep tissue penetration and very high signal to background. The resolution of fine structures at the macroscopic level is really remarkable," says Olive. This fluorescence reflectance system is suited for imaging events at the surface, obtaining 2D images, and doing relative quantitation experiments. While various companies do offer lower cost, manually operated base platforms for optical imaging, the Pearl "is truly a one button operation," emphasizes Olive.

Toward the Clinic

Most advances in optical techniques described here are limited to small animal imaging, mainly because light only penetrates 1–2 cm. Also, to use them in humans, imaging agents would need to be clinically validated. Still, remarks CRi's Mansfield, "Unlike bioluminescence, fluorescence imaging is potentially 'translatable' for any antibody label or any activatable reagent. Optical imaging can be used for anything where the surface can be imaged, using any available optics, including endoscopes." McLaughlin notes, "Many are working toward new technologies that may make optical imaging in large animals and humans achievable at a greater depth." Ultimately, the hope is that molecular imaging technologies allow earlier disease detection, separation of patient subgroups for treatment, and more rapid evaluation of therapeutic response.

Ursula Calef is a medical writer and the founder of Calef Biomedical Consulting based in Laguna Beach, California.

DOI: 10.1126/science.opms.p0900032

Featured Participants

Caliper Life Sciences
www.caliperls.com

Cambridge Research Instrumentation (CRi)
www.cri-inc.com

Carestream Molecular Imaging
www.carestreamhealth.com/molecular-imaging.html

Endra, Inc.
www.enlightbio.com/content/endra/

Leica Microsystems
www.leica-microsystems.com

LI-COR Biosciences
www.licor.com

Oxonica
www.oxonica.com

Seno Medical
www.senomedical.com

Siemens
www.medical.siemens.com

Stanford University
www.stanford.edu

University of California, Davis
www.ucdavis.edu

Vibration Isolation Workstation

The 2800 Series LLHP is an ultra-low-frequency vibration-isolation workstation for equipment that weighs in the 100-pound range, with a 200-pound maximum. The system features proprietary trifilar pendulum mounts and Active-Air suspension to provide horizontal axis vibration isolation that outperforms other mechanical systems, and a high level of vertical axis isolation. The 2800 Series tabletop is 30 inches square and 2 inches thick. Its ergonomic design incorporates automatic leveling and low natural frequencies (1.1 Hz along the horizontal axis and 1.4 Hz along the vertical axis). The workstation can achieve vertical isolation efficiency of 96% and horizontal isolation efficiency of 97% (at 10 Hz and above). Suitable for supporting instruments such as atomic force microscopes and analytical balances, it can be configured for Class 100 clean-room compatibility.

Kinetic Systems Inc.
For information 617-522-8700
www.kineticsystems.com

Living Cell Measurement

The AxioVision Physiology software module enables image capture with a camera and the measurement of physiological parameters in living cells both during microscopic observation (online) and afterwards (offline). The software documents both the images and all the major steps of the experiment and makes them available together with the measurement data. The software is mainly for physiologists, cell biologists, neurobiologists, and electrophysiologists who determine calcium concentrations or pH values through ratiometric calculation of fluorescence images after the addition of indicators such as Fura-2 or Indo-1 (emission or excitation ratio imaging).The AxioVision physiology module makes it just as easy and reliable to measure the change in fluorescence intensity over time—particularly of fluorescent proteins—and to analyze the interaction of proteins using the fluorescence recovery energy transfer method.

Carl Zeiss
For information +49 (0) 3641 64-2770
www.zeiss.de

Slide Loader System

The PL-200 is an innovative, high-speed automated slide loading system designed to be compatible with a wide range of microscopes and suitable for a variety of high-throughput microscopy applications. The PL-200 features a space-saving rotating arm design combined with a multiple sensory-based gripper system to interface Prior Scientific's ProScan II series of motorized stages to ensure a safe and reliable high-speed slide transfer routine. System stability is built-in by the presence of a rigid, common baseplate for both the slide loader and the microscope, which minimizes vibration and provides a solid connection between the two. The PL-200 has a 200-slide capacity accommodated in four removable slide racks that house 50 slides each. An optional barcode reader can be calibrated to read one-dimensional and two-dimensional barcodes.

Prior Scientific
For information 781-878-8442
www.prior.com

Imaging Systems

The Leica MM AF imaging systems combine Leica's microscopy with MDS Analytical Technologies' MetaMorph software to provide a comprehensive range of solutions for wide-field imaging in life sciences research. The new product line offers tightly integrated systems in which ease of use and optical performance give researchers an efficient platform for many types of imaging experiments. An integrated journaling capability allows flexible customization of the platform for virtually any imaging application.

Leica Microsystems
For information +49 (6441) 292550
www.leica-microsystems.com

Custom Conjugation Services

The Kodak X-Sight Nanospheres and Large Stokes Shift Dyes line now features custom conjugation services. The services include labeling of proteins, peptides, and antibodies with the company's fluorescent nanoparticles and dyes, which enable researchers to conduct in vitro and in vivo imaging experiments using some of the brightest imaging agents on the market. The process can be complex and time-consuming for those unfamiliar with it and some biomolecules can be difficult to label; the new services are performed by scientists using cutting-edge techniques that enhance and standardize the performance of the conjugated biomolecule.

Carestream Health
For information 877-747-4357
www.carestreamhealth.com

Cell Observer

The Cell Observer SD (spinning disk) microscope system enables confocal observation and documentation of experiments on living cells over a long period of time and with high frame rates. Through the integration of the Yokogawa Electric Company's CSU-X1 confocal scanning unit with Carl Zeiss products, the system offers a one-stop solution. The system features confocal imaging of very fast processes with optimum specimen protection and without bleaching effects, which is particularly helpful in examining the fast, dynamic processes of molecular cell biology, developmental biology, and neurobiology. Incubation accessories enable users to observe living specimens for hours without damaging them. Major incubation parameters, such as temperature and carbon dioxide content, can be saved automatically along with the image data.

Carl Zeiss
For information +49 (0) 3641 64-2770
www.zeiss.de

Electronically submit your new product description or product literature information! Go to **www.sciencemag.org/products/newproducts.dtl** for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availabilty of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.