6 February 2009 | $10

Science

Speciation

AAAS

SPECIAL SECTION

Speciation

INTRODUCTION

REVIEWS

>> For a list of all related content, see p. 727 or go to www.sciencemag.org/darwin/

EDITORIAL

NEWS OF THE WEEK

NEWS FOCUS

LETTERS

BOOKS *ET AL.*

POLICY FORUM

page 709

PERSPECTIVES

CONTENTS continued >>

COVER

Charles Darwin's abiding interest in the classification of barnacles, illustrated here, was one of the inspirations for his investigations into the origins and diversification of species. *Science* marks Darwin's 200th birthday with a special section on speciation beginning on page 727.

Image: Illustrations by George Sowerby, reproduced with permission from* The Complete Work of Charles Darwin Online*, John van Wyhe, Ed. (http://darwin-online.org.uk); background: iStockphoto.com

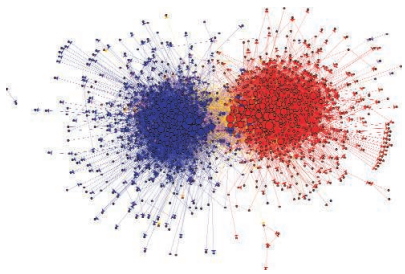
DEPARTMENTS

Published by AAAS

BREVIA

REPORTS

CONTENTS continued >>



page 721

page 737

page 771

Published by AAAS

SCIENCEONLINE

SCIENCEXPRESS
www.sciencexpress.org

Molecular and Evolutionary History of Melanism in North American Gray Wolves
T. M. Anderson et al.

Black coat color in wolves derives from a mutation that originated in domestic dogs.
10.1126/science.1165448
>> Science *Podcast*

Blue or Red? Exploring the Effect of Color on Cognitive Task Performances
R. Mehta and R. Zhu

Blue favors creativity in humans, whereas red improves attention to detail.
10.1126/science.1169144

Dynamic Order-Disorder in Atomistic Models of Structural Glass Formers
L. O. Hedges et al.

The liquid-to-glass transition of a simple mixture is controlled by parameters that drive the system away from equilibrium.
10.1126/science.1166665

Implications of a VLBI Distance to the Double Pulsar J0737-3039A/B
A. T. Deller et al.

Tests of gravity theories that use this double pulsar, now recognized as being more distant, will be more precise.
10.1126/science.1167969

TECHNICALCOMMENTS

Comment on "Human-Specific Gain of Function in a Developmental Enhancer"
L. Duret and N. Galtier

full text at www.sciencemag.org/cgi/content/full/323/5915/714c

Response to Comment on "Human-Specific Gain of Function in a Developmental Enhancer"
S. Prabhakar et al.

full text at www.sciencemag.org/cgi/content/full/323/5915/714d

SCIENCENOW
www.sciencenow.org
Highlights From Our Daily News Coverage

Test Tube Babies Shed Light on Nature Versus Nurture
Children born to genetically unrelated mothers provide clues to origins of antisocial behavior.

Turtles Island-Hopped Their Way Across a Warm Arctic
Widespread volcanism and tropical temperatures helped ancient turtles migrate to America.

Titan's Methane Mystery
Scientists suspect liquid hydrocarbon reservoirs lurk beneath the Moon's surface.

SCIENCESIGNALING
www.sciencesignaling.org
The Signal Transduction Knowledge Environment

PERSPECTIVE: Extracellular ATP in the Immune System—More Than Just a "Danger Signal"
A. Trautmann

Extracellular ATP is an important modulator of immune responses.

PERSPECTIVE: Bistable Switches for Synaptic Plasticity
H. Ogasawara and M. Kawato

Nearly a decade intervened between theoretical predictions and experimental verification of the mechanisms underlying synaptic plasticity.

PODCAST
P. P. Pandolfi and A. M. VanHook

Inactivating mTOR signaling in the adult mouse prostate decreases the incidence of cancer.

E-LETTER: Microarrays Need Phylogenetics
M. Abu-Asab

Interpreting microarray experiments might benefit from evolution-compatible analytical tools.

SCIENCECAREERS
www.sciencecareers.org/career_magazine
Free Career Resources for Scientists

February 2009 Funding News
J. Fernández

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

Darwin's Legacy 1: Rich Collections, Deep Expertise
S. Carpenter

Natural history museums have reasserted their scientific relevance in recent years.

Darwin's Legacy 2: Keeping Order
K. Travis

London entomology curator Erica McAlister is responsible for everything from museum collections to fieldwork.

Darwin's Legacy: Resources
Science *Careers Staff*

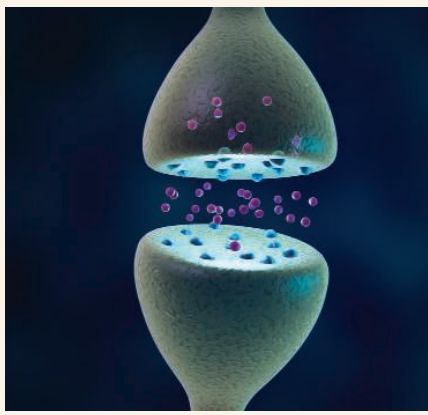
Here is a collection of resources related to careers in the *H.M.S. Beagle*'s wake.
>> *Speciation section p. 727; www.sciencemag.org/darwin/*

SCIENCEPODCAST
www.sciencemag.org/multimedia/podcast
Free Weekly Show

Download the 6 February *Science* Podcast to hear about the biochemistry behind better short-term memory, the genetics of wolf coat color, the origin of art, and more.

ORIGINSBLOG
blogs.sciencemag.org/origins
A History of Beginnings
>> *Speciation section p. 727; www.sciencemag.org/darwin/*

SCIENCEINSIDER
blogs.sciencemag.org/scienceinsider
Science Policy News and Analysis



SCIENCESIGNALING
Dynamic synapses.

SCIENCECAREERS
Career opportunities via Darwin.

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. 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

CREDITS: (*SCIENCE SIGNALING*) CHRIS BICKEL; (*SCIENCE CAREERS*) SHAI MEIRI

<< Beetle Drive

Males of some animal species can adopt alternative reproductive strategies. For example, species may have two types of male, one of large size and horns that guard females, and one of smaller size that lacks horns and sneaks copulation. Such facilitative alternative phenotypes are thought to be regulated by a single developmental threshold. However, **Rowland and Emlem** (p. 773) show that dung beetles may have several qualitatively different threshold mechanisms that each regulate the expression of the same trait and affect the reproductive tactic used. Furthermore, in a surprising number of beetles, two thresholds influenced a single phenotype—the horn—suggesting that facultative male trimorphism occurs among a wide range of beetle species.

Protective Film Production?

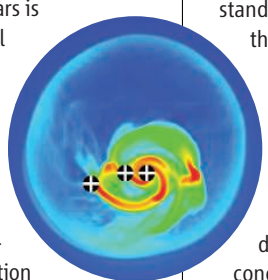
For crystallization controlled by nucleation events, impurities and boundaries can influence the free energy of a system, and thus drive the crystallization toward specific morphologies or crystalline forms. The crystallization of polymers occurs through the repeated folding of chains into lamellae and so boundaries also confine chain motion. **Wang *et al.*** (p. 757; see the Perspective by **Lemstra**) made a series of polymer films through a multiple folding process generating extremely thin, alternating layers of two different polymers, only one of which will crystallize. As the layer thickness was decreased, a preferred orientation for the crystals developed. Initially, this showed fold surfaces parallel to the interface layers, but, for very narrow layers, the lamellae aligned parallel to the interfaces. The crystals in these layers looked like single crystals, and presented a much tougher barrier for gas diffusion. This approach could potentially generate excellent protective films from common polymeric materials.

Legacy of Mass Extinctions

Mass extinctions restructure taxonomic composition of biota on a global scale. However, their legacy, in terms of subsequent patterns of evolutionary diversification, is poorly understood. **Krug *et al.*** (p. 767; see the Perspective by **Crame**) investigated how the end-Cretaceous (KT) mass extinction altered temporal spatial diversification dynamics. Marine bivalves provide excellent fossil records and a well-documented modern biogeography. Long-term shifts in the rate of species origination were identified for the cohort of bivalve genera residing in modern oceans, which show a clear and sustained increase in diversification following the KT event. For bivalves, virtually all of the world's bioprovinces today bear the signature of the KT mass extinction despite the climatic, tectonic, and biotic events of the intervening 65 million years.

No Size Too Large

Radiation pressure is thought to halt the formation of stars with masses greater than 20 times that of the Sun, yet, somehow, such stars are known to exist. The radiation pressure produced by the light emitted by these stars is stronger than their gravitational attraction, preventing them from accreting gas from their natal cloud. **Krumholz *et al.*** (p. 754, published online 15 January; see the Perspective by **Whitney**) present three-dimensional radiation-hydrodynamic simulations of the formation of massive stars, which show that the instabilities driven by the interaction between gravity and radiation allow accretion to continue up to unlimited masses. These instabilities also produce fragmentation, leading to a high proportion of multiple star systems among massive stars.



Oxygen Reduction at Doped Carbon Nanotubes

Fuel cells catalyze the reduction of oxygen with platinum, but cost and scarcity of platinum have led to a search for other catalysts with similar characteristics. Carbon-based catalysts are one alternative, and **Gong *et al.*** (p. 760) report that carbon nanotubes, when doped with nitrogen atoms and grown as vertically aligned arrays, catalyze the oxygen reduction reaction in highly alkaline solutions with activities comparable to that of platinum. The high activity is attributed to the electron-accepting ability of the nitrogen centers creating a net positive charge on adjacent carbon atoms.

Awesome Ores

Ore deposits represent unusually high concentrations of metals or other minerals, often deposited from a fluid. A major question toward understanding their origins is the concentrations in the fluids prior to deposition. Until recently, it has not been possible to measure metal concentrations in the tiny fluid inclusions directly within most ore minerals. **Wilkinson *et al.*** (p. 764; see the Perspective by **Bodnar**) have now measured this concentration directly in two lead-zinc deposits. Metal concentrations were orders of magnitude higher than predicted from solubility arguments or from data on accessory minerals.

Great-Appendage Arthropods

The "great appendage" arthropods, which lie near the base of the arthropod phylogenetic tree, are characterized by a large clawed limb at the front of the head and have only been found in the Burgess Shale–type deposits of the Cambrian. **Kühl *et al.*** (p. 771) report the discovery of a much more recent great appendage arthropod from 100 million years after the Cambrian, preserved in the Hunsrück Slate, the famous pyritized fossil deposit near Bundenbach in Germany. This arthropod is a kind of chimera, which retains

Continued on page 685

CREDITS (TOP TO BOTTOM): J. M. ROWLAND AND D. J. EMLEN; KRUMHOLZ *ET AL.*

Continued from page 683

primitive characters in an otherwise more advanced arthropod morphology. Thus, great appendage arthropods persisted for much longer than previously assumed, and their absence in the fossil record probably reflects a lack of suitable preservation sites.

Counting Centrosomes

Accurate chromosome segregation during eukaryotic cell division requires a microtubule spindle, formation of which in animals is driven by centrosomes. In order to maintain genome stability, centrosomes must be duplicated once every cell division cycle along with the genomic DNA. **Hemerly *et al.*** (p. 789) show that Orc1, a component of the Origin Recognition Complex (ORC) essential for DNA replication, also acts at centrosomes to control their copy number. Orc1 inhibits Cyclin E–dependent reduplication of the centrosome, and its localization to centrosomes is directed by the Orc1-binding partner Cyclin A.

Social Climbing

Many ant colonies are infiltrated by specialized social parasites that penetrate unharmed to feed on the rich resources concentrated within ant nests. **Barbero *et al.*** (p. 782) show that the larvae and pupae of a large blue (*Maculinea*) butterfly social parasite mimic the distinctive sounds made by their host ant's queen, eliciting similar enhanced care from the workers. The queen ants make different sounds from the workers in ant societies, and the workers behave more protectively when they hear a queen's sounds. Thus, the parasite is afforded better care and higher social status.

Little Killers

Platelets plug damaged blood vessels when activated by exposure to collagen, and can also adhere to and promote the sequestration of malaria parasites within narrow peripheral blood vessels, contributing to cerebral pathology. Their generally sticky nature also means that platelets are valuable adjuncts to front-line defenses against early stages of pathogen attack. **McMorran *et al.*** (p. 797) now show that, despite their pathological role later in infection, platelets are important to the innate host responses against the initial stages of malaria. Platelet inhibitors, including aspirin, remove the lethal effects of platelets on the malaria parasite *Plasmodium falciparum* in vitro. Mice engineered for platelet deficiency succumbed more rapidly to infection with a similar parasite *Plasmodium chabaudi*. In both cases the platelets stuck to the infected red blood cells and killed the parasite.

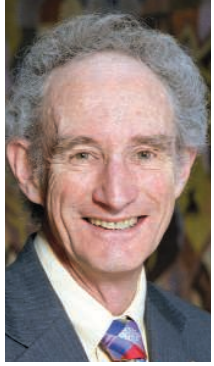
Working Memory and Dopamine

Working memory—the ability to retain information for short periods of time—is important for a wide range of cognitive functions. Dopaminergic neurotransmission plays a central role in working memory. **McNab *et al.*** (p. 800) investigated whether cognitive training changes the density of cortical dopamine D1 receptors and subcortical dopamine D2 receptors. Intensive training in volunteers induced an increase in working memory capacity, which correlated with changes in D1 but not in D2 receptor binding potential. These training-induced changes indicate an unexpectedly high level of plasticity of the human cortical dopamine D1 system and emphasize the mutual interdependence of behavior and the underlying brain biochemistry.

Axonal Regeneration in Nematodes

During nervous system development, a huge number of axons reach out toward their connections. In the adult, however, such outgrowth is much less common. When an adult nerve is injured, neurons are sometimes able to grow out new axons, but in humans, this imperfect process occurs more readily for nerves of the peripheral nervous system. **Hammarlund *et al.*** (p. 802, published online 22 January) now shed some light on the molecular tools that support axon regeneration in the nematode, *Caenorhabditis elegans*, and differentiate the process from normal development. In worms in which a genetic fault leads to excessively brittle axons, or in which a laser is used to cut axons, regeneration of the axon from the remaining stump depended upon a key kinase pathway. This kinase functions early in the road to regeneration, but does not seem to be involved during initial development of the axons.

CREDIT: BARBERO *ET AL.*



Robert M. May and Paul H. Harvey are in the Department of Zoology, University of Oxford, Tinbergen Building, Oxford OX1 3PS, UK.

Species Uncertainties

APPROPRIATELY FOR A YEAR CELEBRATING DARWIN, THIS ISSUE OF *SCIENCE* HAS A SPECIAL section that surveys our current understanding of speciation processes (p. 727). The section illustrates Dobzhansky's memorable injunction that nothing in biology makes sense except in the light of evolution.

You cannot, however, even begin to make sense of biodiversity until you have some systematic sense of what is actually there. This task began, astonishingly, a full century after Newton and his predecessors, with the 10th edition of Linnaeus' *De Rerum Naturae* (1758). The legacy of this unhappy quirk lingers with us today. Take the U.S. Library of Congress—its Web site indicated that, as of December 2008, the library held 20,854,816 cataloged books. Yet we do not know, to within 10% or so, how many distinct eukaryotic species have been named and recorded (the species concept arguably needs to be interpreted differently for prokaryotes). This derives partly from the lack of synoptic and coordinated catalogs for many invertebrate taxa and partly from resulting unresolved synonymies: the same species separately identified and named on two or more occasions. The total number of known eukaryotic species is currently estimated as ~1.8 to 1.9 million, but with all the synonymies removed it may be 1.6 million or fewer. Bird and mammal species are known very well and most other vertebrates reasonably well. Somewhere around 90% of plant species are probably known. But our knowledge of invertebrate species (insects, helminths, and others) is woefully inadequate. So credible estimates of the true eukaryotic species total run around 5 to 10 million, but suggestions as low as 3 million or as high as 100 million can be defended.

Given these uncertainties, we can say even less about the numbers of species likely to be threatened with extinction. Assessments by the International Union for Conservation of Nature (IUCN) suggest that ~20% of all mammal species and 12% of all birds are seriously endangered. But our ignorance about even those insect species that have been recorded is highlighted by the fact that only 0.06% of them feature in the IUCN Red List of Threatened Species. It is nevertheless possible to estimate the rate of species extinctions over the past century as compared with average rates over the 550-million-year span of the fossil record by assuming that well-known groups, such as birds and mammals, are representative. This suggests that recent extinction rates have been 100 to 1000 times faster than average. And under further stresses from habitat destruction, alien introductions, and overexploitation, all exacerbated by climate change, a further 10-fold increase is foreseen over the coming century. This puts us on the breaking tip of a sixth great wave of mass extinctions, differing from the previous Big Five in that it is associated with our activities rather than with environmental events.

A better understanding of speciation processes, and of species themselves and their ecological roles, is not just an academic exercise. The recent Millennium Ecosystem Assessment made a comprehensive appraisal of the condition of, and trends in, the world's ecosystem services: benefits provided to humans as a result of species' interactions within ecosystems. It identified 24 categories of such services, finding that 15 are being degraded or used unsustainably, 4 have been enhanced, and the remaining 5 are unevaluatable. A separate study assessed the value of these services as being comparable with conventionally assessed gross domestic product. Most of these ecosystem services are delivered by invertebrate species, which are largely unknown to us. One consequence is that our growing understanding of how ecosystems are structured and function remains insufficient to answer the question of whether we will be able to engineer repairs to, or substitutes for, degraded ecosystem services.

Other initiatives, long overdue, are speeding up the rate at which we are cataloging the library of living species. But concomitant growth in humanity's ecological footprint, arguably to unsustainable levels, requires all these advances and more. Best of times, worst of times.

– Robert M. May and Paul H. Harvey

10.1126/science.1170937

Published by AAAS

Bond crisis halts California projects
698

Security concerns change U.S. biology
699

U.S. BUDGET

Agencies Sweat the Details of Spending Billions More on Science

It is a problem U.S. science agencies are delighted to have: how to spend billions of dollars on basic research over the next 18 months in a way that does not cause headaches down the road.

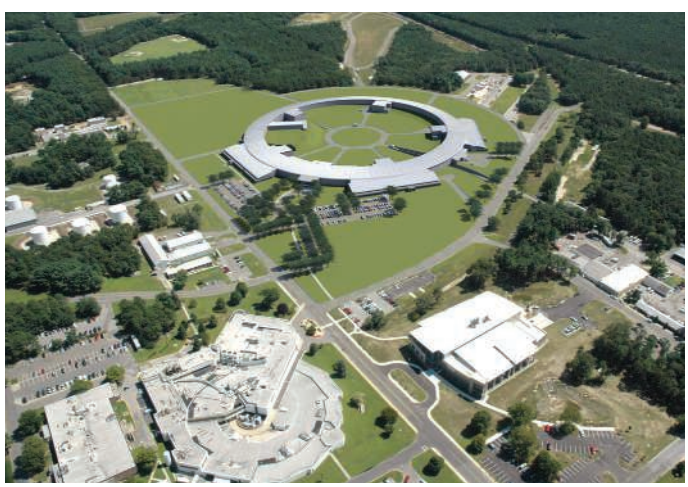
Although tax breaks and aid to beleaguered states are getting most of the media attention, President Barack Obama's plan to stimulate the U.S. economy includes an unprecedented infusion of money into science. It reflects bipartisan support for the idea that research not only puts people to work but also increases the chances of creating new industries that will sustain economic growth. Most of the money would be spent on either bread-and-butter research grants to individuals and small groups or larger awards to rejuvenate labs, buy equipment, and build or renovate facilities at universities and federal laboratories.

Agency officials say there are more than enough worthy ideas for putting the money to immediate use. But the trick will be to avoid a crash when the stimulus money dries up. For infrastructure projects, that means not letting capacity outstrip the government's ability to support the researchers who will occupy the space. The challenge is even greater for grants, requiring a delicate balancing of veteran and first-time grantees, consideration of the impact on underrepresented groups, and avoiding a bolus of new grants that expire at the same time and trigger a flood of application renewals. Any mistakes in managing either pot, say scientists from both the previous and current Administrations, could destabilize the overall scientific enterprise.

"I do think that money of the magnitude being proposed can be spent on useful things," says John Marburger, who headed the White House Office of Science and Technology Policy for the entire tenure of President George W. Bush. "But it's short-term money. The great danger is creating facilities that no one can afford to operate."

Harold Varmus, who advised the Obama campaign and who was named co-chair of the President's Council of Advisers on Science and Technology in December, makes the same point about research grants. "Not everybody understands that grants create an obligation," says Varmus, a former U.S. National Institutes of Health (NIH) director. "So the base is crucial. Obama talked repeatedly during the campaign about gradual and consistent funding for science. Maybe part of this [stimulus] should go into the base."

Although the package is still being debated, the eventual figures for science promise to be huge. The $819 billion package approved last week by the House of Representatives contains some $20 billion for research-related activities, including a $6 billion dollop for the Department of Education to distribute to states for all manner of repairs and renovations at colleges and universities. The plan now before the Senate offers a slightly different mix—more for research and less for infrastructure at NIH, less for both pots at the U.S. National Science Foundation (NSF), and less for Department of Energy (DOE) facilities.



A leading light. The stimulus bill could speed up construction of Brookhaven's $900 million National Synchrotron Light Source II.

CREDIT: DOE

Obama, who has promised that three-fourths of the new, one-time spending will occur by the fall of 2010, has urged legislators to work out their differences in the next few weeks so that the money can start flowing.

NSF, NIH, and DOE officials are taking different tacks in preparing for the windfall. NSF plans to do more of the same within both research and infrastructure. NIH hopes to add new twists to its research portfolio. And DOE is expected to accelerate construction, letting the future take care of itself. Here are highlights from each agency.

■ **NSF.** The House bill would give NSF $2 billion for new research grants and $900 million for three different types of infrastructure programs, including reviving and expanding one for university projects. The Senate is offering less in both categories—$1 billion and $350 million.

Rather than asking for new proposals, Director Arden L. Bement says that program managers will dip into the pool of existing applications submitted since 1 October 2008, the start of the 2009 fiscal year, and fund those rated most highly by reviewers. "We have enough in our backlog to spend that [$2 billion] right now," he says. The expansion will also allow NSF "to take care of young investigators," he adds.

Money for a multiyear grant (typically 3 years) would be committed up front, at the time of the award, rather than parceled out year by year. That will put pressure on investigators "to spend it quickly rather than putting it into the bank," says Bement. NSF estimates that the House package translates into 3000 additional awards that would employ 12,750 scientists, postdocs, and graduate and undergraduate students. Then there's the impact on the economy of what's bought with the grant, everything from reagents to high-end instrumentation.

NSF also plans to stay the course on infrastructure spending. Bement says he will dust off and update the last solicitation for an identical but smaller academic modernization program that lapsed in 1997 if the

Published by AAAS

The challenges facing polio eradication
702

What triggers a superorganism?
706

$200 million proposed by the House makes it through the legislative gauntlet.

■ **NIH.** Acting NIH Director Raynard Kington says the agency has drawn up three ways to spend the grants' portion of the windfall, which is $1.5 billion in the House version and $2.7 billion in the Senate. (Senator Arlen Specter (R–PA) has proposed bumping the Senate number to $9.2 billion.)The first would be a new call for proposals involving "topics in which there have been scientific or technical challenges" that might yield quick results with a blast of cash. Researchers would apply for accelerated review for awards of up to $500,000 a year for 2 years. These challenge grants would fit well with NIH's emphasis on new investigators and high-risk science, Kington says.

NIH would spend some of the stimulus money on standard investigator grants (R01s) that scored well in peer review last year but didn't get funded. The catch: The money would last only 2 years instead of the usual 4 years. Finally, Kington says, NIH might also add to the size of the awards made to investigators whose requested budgets were cut or who can identify "related research areas that might be meritorious." Again, however, the supplement would probably be for only a couple of years.

All of these strategies are aimed at avoiding what Kingston calls the "hard lesson" of the 5-year doubling that left grantees with what Marburger describes as "a gargantuan appetite" that NIH hasn't been able to fill. "There will be great pressure [to use the stimulus money] to fund more R01s just as we currently fund them," Kington says. "But we have to be careful that we do not do anything imprudent, and that includes setting ourselves up to have big commitments 2 years out" that can't be met.

The House bill would also give NIH $1.5 billion to spend on repairs and improvements of extramural research facilities and for "shared instrumentation." (The Senate would provide $300 million.) NIH is all set to handle awards like these and can do it quickly, Kington says. Both houses have proposed $500 million for renovations to the NIH campus.

■ **DOE.** The big gap between what the House and Senate have proposed for DOE's Office of Science—$2 billion versus $430 million—is forcing DOE officials to hold their fire. Acting office head Patricia Dehmer would say only that her staff has done "detailed homework" on how to implement the House's orders to "support improvements to DOE laboratories and scientific facilities."

Her former boss, Raymond Orbach, says he assumes legislators "will split the difference." The American Physical Society has drawn up a list of $1.6 billion in infrastructure or equipment improvements that could be jump-started within 4 months, and Orbach says he's not afraid of accruing long-term mortgages. "The actual running cost of a facility is only 10% of its construction," says Orbach. "These investments will create platforms that don't exist anywhere else in the world. That's an opportunity you can't afford to pass up."

The wait is agonizing for individual project directors. Steven Dierker, who manages the just-approved National Synchrotron Light Source II at Brookhaven National Laboratory in Upton, New York, says adding $150 million to DOE's request for $103 million this year would put hundreds more people to work, sooner. It would also save money, he adds, by shortening the 7-year construction schedule for the $912 million facility.

–JEFFREY MERVIS

With reporting by Eli Kintisch and Eliot Marshall.

Growth industry? Congress may revive and expand an NSF program that funded research facilities like these greenhouses at the University of California, Davis.

HIGHER EDUCATION

New Ph.D.s to Teach Harvard Undergrads

Harvard University plans to hire up to 20 recent Ph.D.s to teach undergraduate courses in a move that officials say will improve instruction and help students facing a tough job market.

The new College Fellows Program was announced last week in an e-mail to faculty and will go into effect this fall. Fellows will be paid $48,000 with full benefits to work in some 20 academic departments throughout the Faculty of Arts and Sciences. The program is open to all recent—since 2005—Ph.D. graduates. The awards are for 1 year, with a second year possible, and the money will come from the university's instructional budget.

"A large part of the goal was to support graduates and mentor excellent teaching among recent Ph.D.s, and [another] was to meet essential teaching needs in Harvard College," says Allan Brandt, dean of Harvard's Graduate School of Arts and Sciences. "We wanted to develop strong teaching for our Harvard College students and to make sure our teaching needs were met."

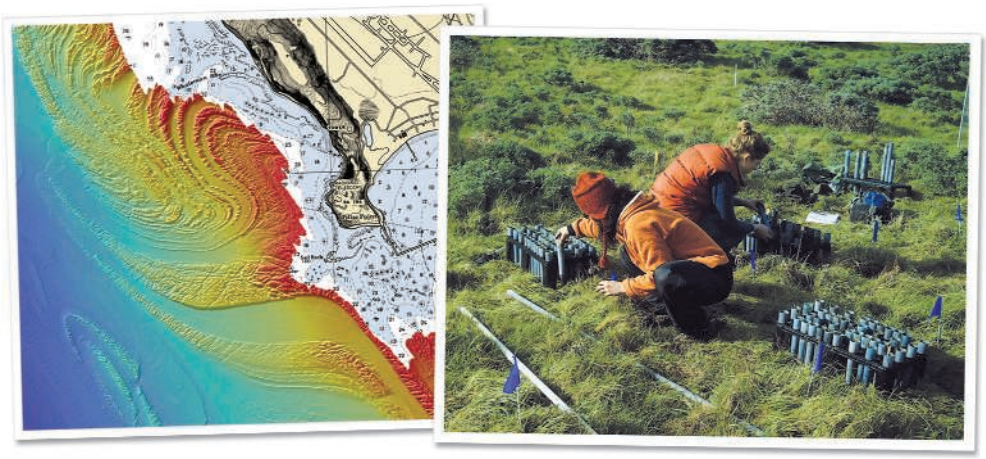
Each fellow will be assigned a faculty mentor, and teaching-focused seminars are planned. The program, Brandt says, "is designed for people who have a deep interest in university teaching." Fellows will be expected to carry 70% of the teaching load of a faculty member, leaving them some time to pursue their research. "At this career stage," says Brandt, "it's very important that they have some protected time to continue their research endeavors."

James Hanken, Alexander Agassiz professor of zoology, says the program offers new Ph.D. recipients "a sort of a temporary hold" in a tough job market. "If it's a teaching postdoc that doesn't consume all of your waking hours and leaves you time to do some research, I think it can be a good deal," he says. **–SUSAN GAIDOS**

Susan Gaidos writes for *Science* Careers.

CREDIT: RENEE C. BYER/*SACRAMENTO BEE*/ZUMA PRESS

Published by AAAS



On hold. A sea floor–mapping study (*left*) and coastal prairie restoration project are among the victims of California's bond fund freeze.

FINANCIAL CRISIS

California Researchers Chilled by Sudden Freeze on Bond Funds

For many California researchers, the bad news came just before the holidays: As a result of the state's deteriorating financial situation, the department of finance was freezing billions of dollars in funding tied to the sale of state bonds. These funds had been slated for a vast variety of projects statewide, everything from supporting schools and public housing to building libraries and fixing freeways. But the freeze also pulled the plug—at least temporarily—on thousands of environmental research and conservation projects. In the ensuing weeks, scientists, graduate students, and nonprofit organizations have faced a mad scramble to replace what most had assumed was a secure source of funding.

When the decision was announced on 19 December, Rikk Kvitek, a marine ecologist at California State University, Monterey Bay, got the call right before heading to his lab holiday party. "I had to walk in and say 'Merry Christmas, and by the way, there's no work tomorrow because virtually all of our funding is tied up in this.'" Kvitek is a principal investigator on a $20 million state-funded seafloor–mapping project that was just hitting full stride. "For the last 8 years, I've been really pushing the state to map all the state waters because we knew that it would really transform coastal marine research and how resource management was done," he says. The high-resolution digital maps are being used, among other things, to help select sites for the state's network of marine protected areas. Because of its scale, the project has become the primary focus of Kvitek's work, and he says for the first time in 20 years he had all of his funding eggs in one basket.

Like others, Kvitek was instructed to stop all work immediately on bond-funded projects. The next day, a 180-foot ship surveying the northern California coast pulled into port and has been out of commission ever since. Kvitek has managed to find temporary funding from a foundation to keep his 15 students and staff members working on data analysis, but he says that money will last only 2 or 3 months.

California's governor and legislature are at loggerheads about how to narrow a projected $40 billion budget deficit, and the state's credit rating has tanked, scuttling its ability to sell bonds to raise capital. In more typical times, the state loans money to approved projects and subsequently recoups the money by selling bonds. But now that the bonds aren't selling, the state decided that the loans had to stop. Environmental projects are feeling the pinch in part because of the public support they've received in recent years as California voters have passed propositions authorizing the state to sell bonds to fund projects to study and manage the state's natural resources. (The state's bond-funded stem cell institute has not yet felt the pain as acutely but is girding itself; see sidebar.)

"This came totally out of the blue," says Susan Williams, the director of the Bodega Marine Laboratory in Bodega Bay, administered by the University of California (UC), Davis. Many of the frozen projects were intended to help support California's coastal economy, Williams says. "These are not pie-in-the-sky projects," she says, citing ocean-monitoring projects used by fisheries managers and the U.S. Coast Guard, efforts to combat invasive species, and research on the impact of climate change.

Some projects have been put on hold at a critical juncture. One example, Williams says, is a coastal prairie-restoration project run by the Bodega marine lab. This postcard-pretty habitat along the Sonoma County coast is being taken over by invasive grasses. The eradication plan called for a series of mowing and herbicide treatments followed by plantings of native species. The first treatment was completed last fall, but the funds are now frozen for the second round of grass removal, which should be happening now. "Unless the funding is restored and we can do the second removal, the seed bank of these weeds will germinate," causing "an even bigger problem" than existed before the ▶

Stem Cell Institute Looks for New Ways to Raise Cash

In 2004, California voters authorized the state to raise $3 billion for stem cell research through the sale of bonds. Now, the state's finances are in disarray and its bonds aren't selling, forcing the California Institute for Regenerative Medicine (CIRM) to look for new ways to raise cash. At a meeting last week in Burlingame, California, the institute's leaders discussed alternatives and provisionally approved funding for $58 million in new training grants.

Although the state's recent decision to freeze funding on projects tied to bond sales has had a chilling impact on many researchers in the state (see main text), the freeze has not had a big immediate impact on CIRM, says Vice President for Operations John Robson. As of 1 January, the institute had $158 million on hand, thanks to proceeds from bond sales that took place before the recent economic nosedive. That should be enough to maintain operations and honor previous funding commitments for several months, Robson says: "We feel like we have enough money to carry us to about November."

But then what? CIRM estimates it would need to raise close to $136 million just to continue funding ongoing commitments through the end of 2010; it would need a total of $377 million to also fund projects that have already been given preliminary approval from its board and anticipated new programs. At the meeting, CIRM Chairman Robert Klein proposed raising money by selling private placement bonds to wealthy investors interested in the potential societal benefits of stem cell research. **–G.M.**

CREDITS (LEFT TO RIGHT): CALIFORNIA STATE UNIVERSITY, MONTEREY BAY; PETER ALPERT

Published by AAAS

program started, Williams says.

The restoration project was also paying tuition, fees, and a stipend for Tawny Mata, an ecology graduate student at UC Davis. The funding freeze "has pretty much left me up in the air about how I'll finish my Ph.D.," Mata says. A teaching assistantship is paying her bills this quarter, but beyond that Mata isn't sure how she'll manage. "I've been contacting any professor I've ever worked with to see if they have any money lying around." She's not alone: A recent e-mail survey found that at least 24 out of about 160 students in her program had lost at least some funding.

Across UC Davis, 60 projects received stop-work orders, says Jan Hopmans, chair of the university's Department of Land, Air and Water Resources—20 in his department alone. "Many of these grants are for a few hundred thousand to a few million dollars," Hopmans says. "We have 50 employees just in my department for whom we have in principle no funding at this time." Researchers, students, and technicians have been reassigned to projects with other sources of funding where possible, Hopmans says, but so far 13 people have received layoff notices.

Nonprofit groups are also feeling the pain. "A lot of our grantees are relatively small organizations, and some of them will go out of business if this goes on too long," says Samuel Schuchat, executive officer of the Coastal Conservancy, the state agency charged with administering bond-funded grants for coastal research and conservation. One such program, the Invasive Spartina Project, an effort to eradicate invasive *Spartina* cordgrass from the San Francisco Bay, would be especially painful to lose, say Schuchat and others. The state has already invested nearly $10 million in the project, which has reduced the area covered by the grass by 90% since 2006 and is on course to eradicate it by 2012, says Peggy Olofson, the project's director. Olofson has cobbled together money to run a scaled-down operation this year, but beyond that the future is uncertain.

How long the bond funds will remain frozen is unclear, but all eyes are on Sacramento, where the governor and state legislators are wrangling over how to close the budget gap—a necessary first step toward restoring the state's credit rating and restoring its ability to sell bonds. Only then will those affected by the freeze be able to start thinking of a thaw.

–GREG MILLER

BIOSECURITY

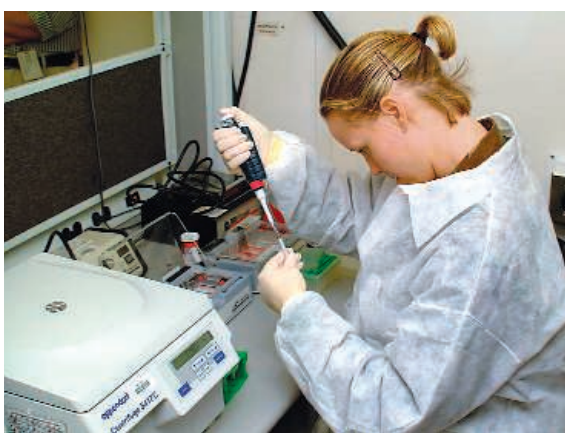
Life Scientists Cautious About Dual-Use Research, Study Finds

Some life scientists are changing the way they do business because of security concerns, according to a U.S. survey released this week.

Researchers and policymakers in the United States have been hotly debating the need for new government regulations to prevent the misuse of life sciences research by terrorists and other bad actors. Even without such regulations, according to the survey, a few scientists are avoiding "dual use" research projects with the potential for harm; some are shying away from international collaborations; others are excluding foreign graduate students and postdocs from certain lines of work and censoring themselves while talking about their research.

In all, 15% of the nearly 2000 life scientists who responded to the survey, conducted in late 2007 by the National Research Council and AAAS (publisher of *Science*), reported having changed their behavior in one or more of those ways. "It is a surprisingly high number," says study chair Ronald Atlas, a microbiologist at the University of Louisville in Kentucky. He finds it worrisome that security concerns may be impinging on the traditional openness of research in the life sciences. "What's not clear is whether the community is overreacting or if this is an appropriate response," Atlas says.

The finding is also an implicit endorsement of the popular argument among academics for letting scientists police themselves on dual-use research rather than imposing government-mandated rules. The National Science Advisory Board for Biosecurity endorsed that self-governance approach in recommendations to the government in 2007, but federal officials have not yet decided what the policy should be.



Self-review. Some researchers are avoiding certain projects because of security concerns.

Richard Ebright, a chemist at Rutgers University, New Brunswick, who has argued in favor of tougher regulations, says he finds the survey results "hard to believe," given that previous studies have shown that most scientists in the community aren't even aware of dual-use concerns. Ebright suspects that the survey, which was e-mailed to 10,000 life scientists who are members of AAAS, attracted an overwhelming proportion of responses from individuals who would "prefer not to see [government] regulations." Atlas agrees that the survey may have captured "a biased group that had been thinking about this topic" and says that the findings "would require further verification from broader surveys."

The study authors say the survey results point to the need for clearer guidelines on what kinds of research might have the potential for dual use. "It's possible that some life scientists are being overcautious because there is no good definition of dual-use research," Atlas says. Panelist Robert Cook-Deegan, a biosecurity expert at Duke University in Durham, North Carolina, says biosafety committees at some institutions are already working with their scientists to help evaluate the dual-use potential of research projects and respond accordingly.

As an example, he cites a project led by Mark Denison of Vanderbilt University in Nashville, Tennessee, and Ralph Baric of the University of North Carolina, Chapel Hill, that set out to make a SARS-like virus using synthetic biology techniques. The researchers "thought about dual use with their biosafety committees all along, and we did a half-day workshop before their publication to talk about what should not be included in the final publication and why," Cook-Deegan says. The paper was published in the 16 December 2008 issue of the *Proceedings of the National Academy of Sciences*, with minor modifications to the language and no data withheld. "It's a really nice example of scientists taking dual use seriously," he says.

–YUDHIJIT BHATTACHARJEE

CREDIT: STEVE FERENDO/AMERICAN FORCES PRESS SERVICE/DOD

Published by AAAS

GLOBAL HEALTH

Some Neglected Diseases Are More Neglected Than Others

In a bird's-eye view of who's fighting diseases among the poor, Ireland and the United States stand out for largess; Germany and Japan seem skimpy. As for targets, the battle against dengue is relatively well-funded, whereas pneumonia, meningitis, and diarrhea get mere table crumbs. The "big three"—HIV/AIDS, malaria, and tuberculosis—together soak up some 80% of the money.

Those are some conclusions from the first global study tracking spending on research and development for "neglected diseases." The authors, led by Mary Moran of the George Institute for International Health in Sydney, Australia, hope that the study will help grantmakers decide where best to put their money and spur countries into action.

With the help of an expert panel, the team first decided which diseases currently suffered scientific neglect; to qualify, they had to occur disproportionately in low- or middle-income countries, and there had to be a need for new vaccines, drugs, or diagnostics that the market fails to address. The team then painstakingly tried to trace every grant or investment made in 2007 by governments, institutes, charities, and companies to address those needs. The paper, published in *PloS Medicine* this week, documents a total of $2.5 billion spent on 30 diseases.

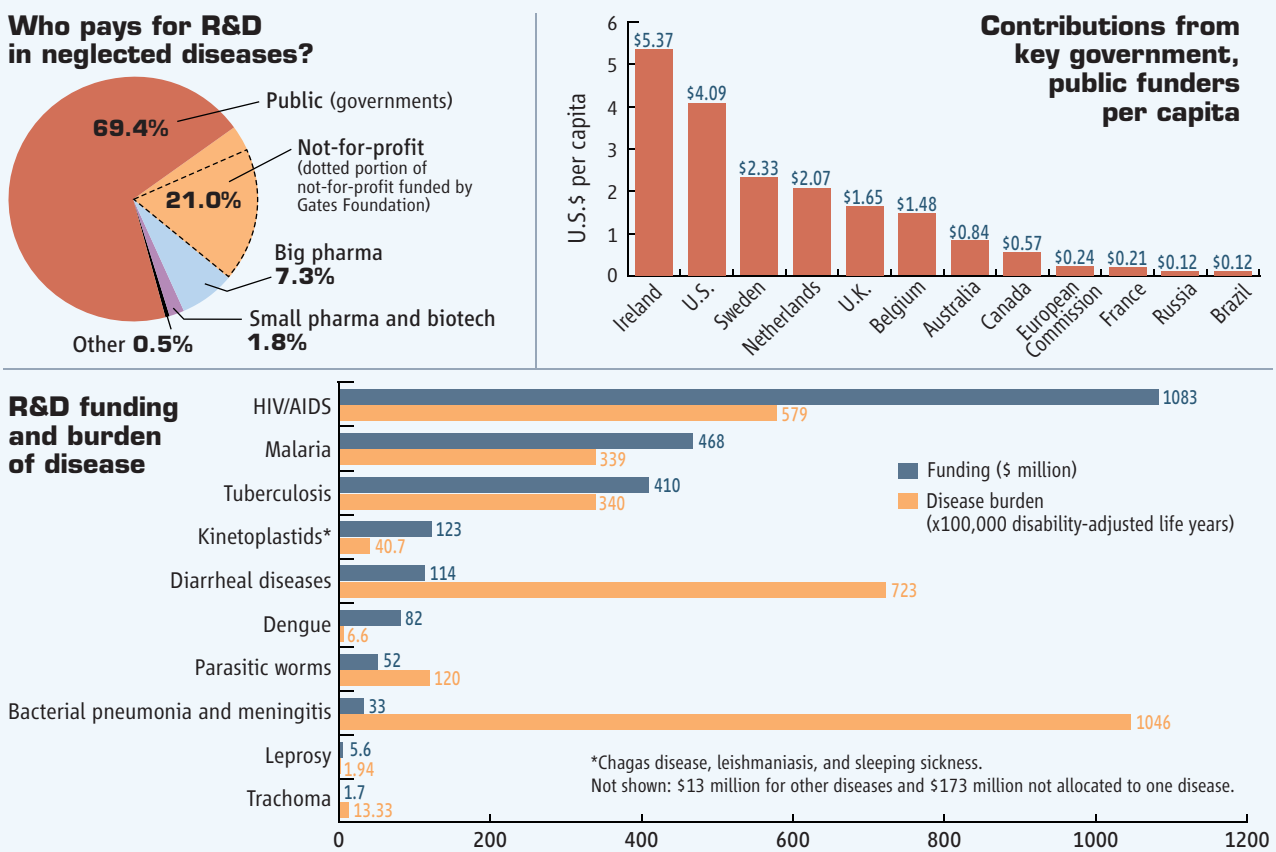
Government agencies pick up the biggest part of the tab, with the U.S. National Institutes of Health good for 42% of the research. (The Bill and Melinda Gates Foundation comes in second at 18%.) Per capita, however, Ireland is the biggest public spender.

A comparison of the estimated global burden of each disease as expressed in disability-adjusted life years (DALYs), a measure that aims to capture both disease and death, shows that R&D funders have flocked to the big three, while other diseases with an even bigger burden get little R&D attention. The accuracy of the DALY estimates, which the World Health Organization last updated in 2004, is under constant debate. Moran explains that funders may have other reasons to pick a disease, such as the expectation that investments will pay off quickly; that's why the team mentioned DALYs only briefly in its report. Still, it's clear that some diseases are underfunded, she says.

Indeed, the data show that the big three no longer deserve the term "neglected," says David Molyneux of the Liverpool School of Tropical Medicine in the United Kingdom, who calls the paper "impressive." Molyneux says policymakers must "shift resources" to respiratory infections, diarrhea, and other underfunded diseases. Some can be controlled without new research, by scaling up current, and often cheap, methods. (See Three Q's on p. 695.)

The team, which has funding from the Bill and Melinda Gates Foundation to do the survey annually through 2013, is already working on next year's edition. Next time, Moran hopes to include the contributions of four significant funders—India, China, and pharma giants Merck and Wyeth—that failed to submit data this year. But she worries that the economic downturn may lead some funders to spend less where more is needed.

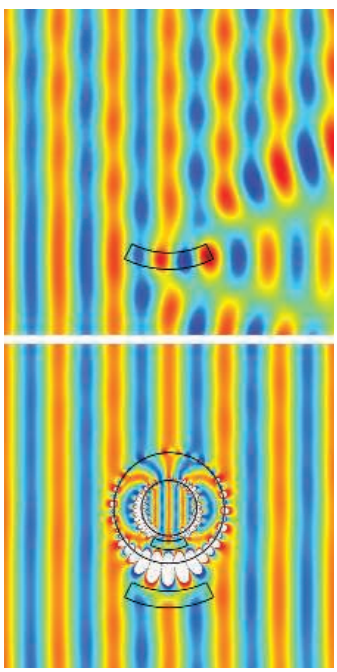
–MARTIN ENSERINK



PHYSICS

Invisibility Umbrella Would Let Future Harry Potters See the Light

Two years ago, physicists blurred the distinction between science and fiction by producing a shell-like "invisibility cloak" that made both itself and an object inside it undetectable—albeit when viewed with microwaves of a specific frequency. Now, a team from Hong Kong has gone one better with a theoretical scheme for an "invisibility umbrella" that can make both itself and an object placed *beside* it disappear. The previous cloak ferried incoming light around the enclosed object, rendering it blind to the outside world. In contrast, the umbrella leaves the hidden thing out in the open, so it can "see" its surroundings.



See? An object scatters light waves traveling from left to right (*top*). With its embedded antiobject, the invisibility umbrella counteracts the scattering, rendering both things invisible.

It's still a far cry from the magic garment that enables Harry Potter to sneak around Hogwarts Castle unseen and spy on others, but the concept draws rave reviews from other researchers. "It's an absolutely brilliant idea," says physicist Ulf Leonhardt of the University of St. Andrews in the U.K., one of the pioneers of cloaking theory.

In devising the scheme, Yun Lai, Che Ting Chan, and colleagues at the Hong Kong University of Science and Technology in Kowloon melded two earlier approaches. In 2005, Andrea Alù and Nader Engheta, electrical engineers at the University of Pennsylvania, predicted that researchers could make an object nearly invisible by coating it with a tailored layer of "metamaterial"—an assemblage of metallic rods and C-shaped rings that interacts with electromagnetic radiation in novel ways—counteracting the thing's tendency to scatter light.

In contrast, in May 2006, Leonhardt and, independently, John Pendry of Imperial College London imagined stretching space so that the paths of light rays bend around an object. They then calculated how they might mimic that impossible stretching by sculpting the properties of a shell of metamaterial. Building on Pendry's implementation of such "transformation optics," experimenters David Schurig and David Smith at Duke University in Durham, North Carolina, produced a microwave cloak just 5 months later (*Science*, 20 October 2006, p. 403).

Chan and colleagues struck a middle ground. In theory and simulations, they first used transformation optics to perfectly cancel the scattering from a cylindrical post of ordinary material by coating it in metamaterial. The researchers then realized that they could make an object near the now-invisible post disappear, too.

The trick, they report in a paper to be published in *Physical Review Letters*, is to embed a matching "antiobject"—the metamaterial equivalent of a voodoo doll—in the outer layer of the post. The scattering from the embedded antiobject exactly cancels the scattering from the object, Chan says, "so it looks as if there is nothing there." Because the hidden object remains outside the post or umbrella, it can detect light from its surroundings.

The scheme does have limitations, Pendry notes. The umbrella works for only a single frequency, has to be specifically tailored to the object to be hidden, and won't completely hide something that absorbs light. Still, Pendry says, "that's carping on my part—it's really a neat idea."

Making an invisibility umbrella even for microwaves may be challenging. It requires "left-handed" metamaterials that bend light in an especially strange way and are difficult to make. Still, Leonhardt says, "it's clear that someone will do this in the future." Given current progress, don't be surprised if some wizard of an experimenter does it sooner rather than later. **–ADRIAN CHO**

CREDIT: YUN LAI/HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

*Science*Insider

From the *Science* Policy Blog

As the global economy continues to melt down, scientists saw some bright spots. But the news wasn't all good, and some of it was downright odd. Here are some highlights from *Science*'s science policy blog, *Science*Insider:

U.S. scientists saw more dollar signs this week when the U.S. Senate debated billions for federal research as part of a massive economic stimulus package. *Science*Insider has a breakdown of who would get what. Still, some argue that money alone won't be sufficient to help young students who want to pursue scientific careers; an overhaul of the U.S. research system is needed. Things weren't as rosy north of the border: On 27 January, the Canadian prime minister unveiled a budget that makes some big cuts to the country's main source of research grants. Given the prospects for flush funding in the United States, Canadian scientists worry about a brain drain. Scientists in France are even more riled up, thanks to an incendiary speech by President Nicolas Sarkozy, which lambasted the country's research system as "infantilizing and paralyzing."

Also making headlines: two chemical elements. Officials at Los Alamos National Laboratory are scratching their heads over a mysterious case of **beryllium contamination.** Adding to the intrigue is the source of the harmful chemical: a part of the lab known as Technical Area 41. Meanwhile, off dry land, scientists have started dumping 6 tons of iron into the Southern Ocean as part of a controversial geoengineering project that aims to increase the ocean's ability to absorb CO_2.

And finally, in a bit of **bizarre news**, an unnamed senior official at the U.S. National Science Foundation apparently viewed pornography on a work computer over 2 years. The incident has caught the attention of Senator Charles Grassley (R–IA), who now questions whether the agency deserves the $3 billion the U.S. House of Representatives has proposed giving it as part of the economic stimulus package.

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

Published by AAAS

Polio: Looking for a Little Luck

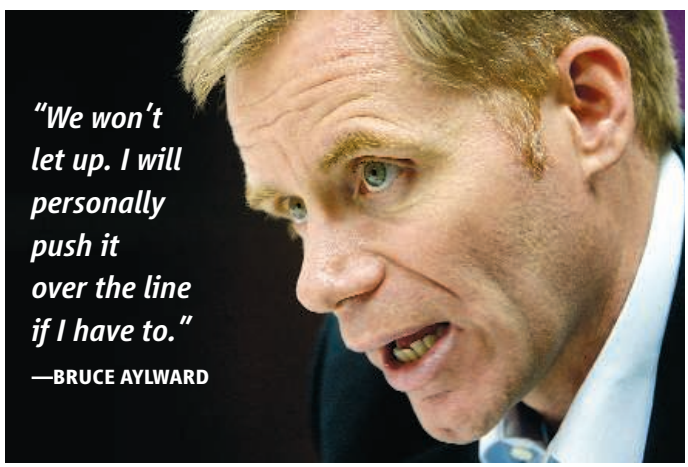
After another year of setbacks—and some real gains—the global polio eradication initiative continues with more support than ever

GENEVA, SWITZERLAND—Soon after she took office 2 years ago, Margaret Chan called an urgent meeting. The time had come, said the new director-general of the World Health Organization (WHO), to take a hard look at the global polio-eradication initiative, which by then was 6 years past deadline, a couple of billion dollars over budget, and facing increasing questions about its feasibility from scientists and tapped-out donors. She wanted no more grand promises of when the virus would be vanquished from the planet. Instead, Chan and the "major stakeholders"—the partner organizations, donors, and countries—launched an "intensified" 2-year program, setting measurable milestones by which to judge progress. The leaders of the global initiative, a collaborative effort based at WHO, were to report back in February 2009, at which time the world could reassess its massive investment in the biggest global health program ever.

That moment of reckoning is here, and the initiative has met only one of the milestones set 2 years ago. At 1643, global polio cases in 2008 were actually higher than the 1315 total in 2007, and the virus remains entrenched in the last four countries where, for reasons both social and biological, it refuses to budge.

Still, no one is talking about pulling the plug. If anything, the beleaguered program has garnered more support and more money than it did several years ago. Just last month, the Bill and Melinda Gates Foundation, Rotary International, and the U.K. and German governments pledged $635 million for polio eradication. "Polio *has* to succeed" is the widely voiced sentiment among Chan and other global health leaders, not only because of the huge investment—20-plus years and nearly $6 billion—but also because of the unsettling realization that there is no palatable way out (*Science*, 20 April 2007, p. 362). Stopping now, so close to the finish, would mean losing the spectacular gains of the past 20 years—a defeat that would certainly be the death knell for other potential global eradication projects, like those for malaria or measles, says Peter Wright, an infectious-disease expert at Dartmouth College who advises the eradication initiative. And that is a decision no one is yet ready to make.

That leaves the Global Polio Eradication Initiative (GPEI)—a collaboration led by WHO, Rotary International, UNICEF, and the U.S. Centers for Disease Control and Prevention—trying every trick in the book to beat the virus into submission. GPEI has a new 5-year plan that calls for reaching those unmet milestones—and far more. The program is investing heavily in research on improved vaccines that earlier program leaders swore would never be necessary; the virus would be gone by the time one could be developed. In a major departure, they are rethinking whether the world can ever safely stop vaccinating against polio, the fundamental assumption on which polio eradication was sold some 20 years ago (see sidebar, p. 705). And perhaps most of all, they are hoping for a little luck.



Stalemate

For a long time, it looked like the polio warriors had the virus licked. Soon after the program was launched in 1988, with the confident prediction that polio would be gone by 2000, the program began dispatching virus from more than 100 countries in quick succession, using the tried-and-true approach that had already eradicated polio in the Western Hemisphere: supplementing routine polio immunizations with huge countrywide campaigns several times a year to deliver drops of Albert Sabin's oral polio vaccine (OPV) to every child under age 5.

By 2000, global cases had fallen 99% from 350,000 to 791, reaching an all-time low of 483 in 2001. In the process, one of the three wild poliovirus serotypes, type 2, was eradicated almost inadvertently—providing a proof of concept that the ambitious plan was indeed feasible. By 2006, type 1 and type 3 virus were cornered in just four "endemic" countries—India, Nigeria, Afghanistan, and Pakistan—where transmission has never been interrupted (*Science*, 26 March 2004, p. 1960).

But there the initiative stalled, with the four endemic countries periodically erupting and reinfecting other polio-free countries and the global case count hovering between about 1000 and 2000.

As skepticism mounted among scientists and weary donors, a few advocated throwing in the towel on eradication and concentrating instead on keeping the virus in check (*Science*, 12 May 2006, p. 832). Meanwhile, Bruce Aylward, the peripatetic and unfailingly optimistic M.D./MPH who has led the effort since 1998, kept insisting that success was just around the corner—just another year away.

That was the context in which Chan launched the intensified program, pouring in more money and resources to determine once and for all whether eradication could be achieved. The answer, everyone agreed, depended on progress in the four endemic countries.

India seesaws

In India, more than in any other country, the polio fighters were banking on a win in 2007–08. The plan was to deal a "mortal blow" to poliovirus type 1, considered the worst player because it causes more paralytic disease and spreads faster than the other remaining wild virus, type 3. Once type 1 was dispatched, mopping up type 3 in India would be easy, they predicted, and would show the world the program was on track.

"If we can do it in India, the toughest place in the world, we can do it anywhere," says Aylward. Early on, polio experts realized India was different; instead of the three to four

CREDIT: TANUSHREE PUNWANI/REUTERS/LANDOV

Published by AAAS

doses of OPV that had sufficed to stop poliovirus transmission elsewhere, in some parts of northern India, children needed eight or 10 doses to be protected—and still some became paralyzed. In the other endemic countries, GPEI says the problem is "a failure to vaccinate"; in India, by contrast, the problem is compounded by "a failure of the vaccine." The country's huge population, explosive birthrate, and crowded, squalid conditions combine to create an ideal environment for the virus, which is transmitted by feces.

To tackle this "pernicious transmission," as Aylward calls it, in early 2005 GPEI helped rush into use a new, more immunogenic version of OPV—a monovalent form that focused all its firepower on just type 1 (mOPV1). Later that year, mOPV3 was introduced (*Science*, 14 January 2005, p. 190). With the new vaccines in hand, northern India launched its sequential strategy, vowing to wipe out type 1 by the end of 2008.

Volunteers flooded the country with oral polio drops, upping rounds from every 2 months to every 4 weeks and focusing on the toughest districts in the crowded, impoverished states of Uttar Pradesh, outside Delhi, where circulation was most intense, and Bihar, some 800 kilometers east.

The results were stunning. Type 1 cases across the country dropped from 648 in 2006 to 73 in 2008. Most remarkable, Uttar Pradesh, which Aylward calls the wellspring of polio in the country—"every virus in India since 1999 has been linked to that area," he says—went 12 months without a case. "It is really a hallmark [achievement]," says Samuel Katz, a polio expert at Duke University in Durham, North Carolina, who directs GPEI's newly reconstituted research advisory committee.

But in early 2008, India was blindsided by a walloping epidemic of type 3 polio. Although the program had continued to use occasional rounds of mOPV3 and the trivalent formulation, tOPV, to forestall just such an event, "we didn't get the balance right," concedes David Heymann, who oversees the polio program as WHO's assistant director-general for infectious diseases.

Then in June 2008, type 1 came back to Uttar Pradesh. Genetic analyses showed it wasn't a local Uttar Pradesh virus, with its distinct genetic signature—instead, it was "imported" from Bihar. To scientists, the distinction was important—it meant transmission in Uttar Pradesh had indeed stopped for the first time ever—but that still left the country battling an epidemic on two fronts, with cases in 2008 down from 2007 but still, at 556, alarmingly high.

CREDIT: WHO/T. MORAN

Backsliding. Nigeria again has a runaway polio epidemic.

Published by AAAS

At a November 2008 meeting, the India Expert Advisory Group, which oversees the country's effort, vowed to continue the fight into 2009, again focusing on type 1 but adding more doses of mOPV3 to keep that serotype in check. As a contingency plan, WHO and partners are testing a higher potency mOPV1, and the country is exploring whether adding doses of inactivated polio vaccine can help boost immunity in young children.

All bets are still on India to be the first of the four endemic countries to stop transmission of the wild virus. India is the "key to donor confidence," says Heymann. "We need a victory."

Nightmare in Nigeria

In contrast, few expected much progress in northern Nigeria, where opposition, apathy, political instability, and corruption have stymied the program for years. But even realists didn't necessarily expect an outbreak of the magnitude that struck in 2008, in which cases in some areas were 10 times higher than in 2007. "Polio in Nigeria remains a nightmare," says Oyewale Tomori of Redeemer's University near Lagos, head of the country's expert polio advisory group.

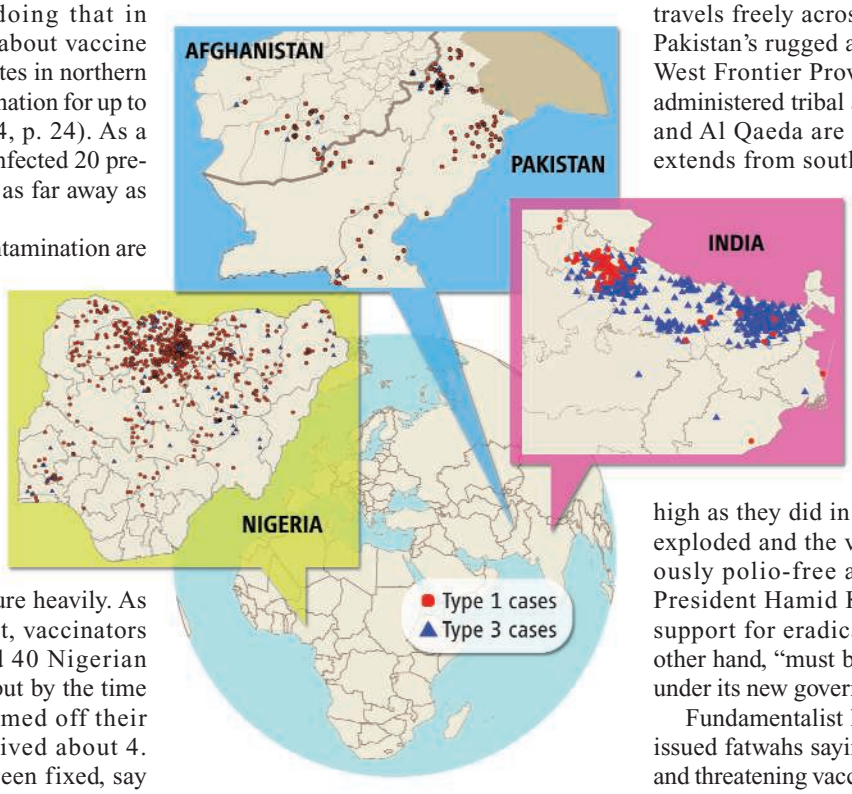
In May 2008, WHO issued a blunt warning that Nigeria posed a risk to the rest of the world, threatening to derail the entire global effort. It came close to doing that in 2003–04, when suspicions about vaccine safety led several Muslim states in northern Nigeria to stop all polio vaccination for up to a year (*Science*, 2 July 2004, p. 24). As a result, virus from Nigeria reinfected 20 previously polio-free countries, as far away as Indonesia.

Rumors about vaccine contamination are no longer the major impediment to eradication; instead, Nigeria's problems are largely "operational," say GPEI officials, citing a lack of political will and the government's failure to provide even the most rudimentary health services. Others more bluntly refer to "gross incompetence" and say graft and corruption figure heavily. As Tomori explains, in the past, vaccinators might have been promised 40 Nigerian nairas a day for their work, but by the time government officials skimmed off their share, each may have received about 4. Those problems have now been fixed, say GPEI officials.

At the epicenter of the epidemic in Kano state, 68% of all children have received fewer than three doses of OPV, and up to 30% are "zero-dose." With 791 cases in 2008, Nigeria accounted for almost 50% of the global total. That number is especially frustrating to polio experts because stopping transmission in Nigeria should be a cinch compared with India. Studies by Nicholas Grassly and colleagues at Imperial College London have shown that vaccine efficacy is high there; that means that transmission of the virus should stop when population immunity reaches roughly 80%.

Now, as in 2003, WHO and world leaders are trying to shame Nigeria into action. Last May, the World Health Assembly passed a resolution singling out Nigeria and calling on the country to quickly stop its runaway outbreak.

The public humiliation may be having the desired effect. In July, President Umaru Yar'Adua vowed to redouble the effort. The ineffective head of the national polio program has been replaced, the third such change in 3 years. "This time it is different; the president is on board," asserts Aylward. Tomori is more circumspect, saying he will wait to see whether this high-level commitment actually translates to action on the ground. Meanwhile, in 2008, poliovirus from Nigeria spread to seven West African countries. Other war-torn countries, including Chad and Sudan, are still grappling with epidemics sparked by earlier "importations" from Nigeria.



Mixed bag. Despite significant advances, polio cases in 2008 remained high in the four endemic countries.

SOURCE: WHO

Perils in Afghanistan and Pakistan

In the other two hot spots, Afghanistan and Pakistan, violence, political turmoil, religious opposition, and the fierce autonomy of local leaders render eradication all but impossible. Large swaths of both countries are "no-go" zones where WHO and other United Nations personnel are not allowed to operate. National polio teams can still get in but are justifiably leery of doing so. Even in "accessible" areas, a "climate of fear" prevails, and vaccination teams may report going more often than they actually do, says epidemiologist Rudi Tangermann, who oversees efforts in the two countries from Geneva.

In mid-2007, GPEI thought it had made significant headway; a "third party" had brokered an agreement with the Taliban to let polio vaccinators work unimpeded. Despite that agreement, in March 2008, two polio workers and their driver were killed by a suicide bomber in southern Afghanistan, where they were traveling to prepare for a vaccination campaign.

Surveillance and monitoring are compromised as well. "We are peering in from the outside," concedes Aylward. The countries constitute one epidemiologic block, with two transmission zones where the wild virus travels freely across the border. One is in Pakistan's rugged and inhospitable North-West Frontier Province and the federally administered tribal areas, where the Taliban and Al Qaeda are resurgent. The second extends from southern Afghanistan, near Kandahar, through Baluchistan, and then stretches all the way to northern Sindh in central Pakistan.

Of the two, Pakistan is the bigger worry, says Tangermann. The number of cases rose in Afghanistan in 2008, but nowhere near as high as they did in Pakistan, where type 1 exploded and the virus spread into previously polio-free areas. In Afghanistan, President Hamid Karzai has pledged his support for eradication; Pakistan, on the other hand, "must become more committed under its new government," says Heymann.

Fundamentalist leaders in Pakistan have issued fatwahs saying the vaccine is unsafe and threatening vaccinators. "Refusals" have risen considerably. In February 2007, a Pakistani doctor and his driver were killed by a

Published by AAAS

RETHINKING THE POLIO ENDGAME

One of the toughest conundrums for the long-running Global Polio Eradication Initiative has been whether and when it would be safe to stop vaccinating once they deem the virus gone. Now, the thinking is undergoing a major shift.

The first big complication came in 1999 when scientists realized that the weakened virus used in the live oral polio vaccine (OPV) could revert to its neurovirulent form in rare cases and spark an epidemic. Thus was born the "OPV paradox": OPV was necessary to eradicate the virus, but as long as OPV was in use, eradication could never be achieved. As a solution, World Health Organization (WHO) scientists proposed a plan: After the world was certified polio-free, all countries would stop using OPV simultaneously, as if at the stroke of midnight.

"The world will be a much safer place if more countries use [inactivated polio vaccine]" when the virus is deemed gone.
—ROLAND SUTTER

Some scientists dismissed the idea as folly and instead advocated universal use of the inactivated polio vaccine (IPV), already widely used in developed countries. That would be the only way to ensure the world was really safe, they argued, and the only way to prevent a gross inequity in which poor countries bore all the risk of polio.

For years, WHO maintained that such a switch wasn't feasible: IPV was too expensive for poor countries, it must be injected, and its effectiveness is unproven in tropical settings. Now, experts such as Roland Sutter, who heads WHO research in Geneva, Switzerland, concede that IPV does have a role after all. "The world will be a much safer place if more countries use IPV," he says.

To make that possible, WHO is now dusting off some earlier studies and investing heavily in new research into a cheaper, more effective version of IPV. WHO is looking at "dose sparing" strategies that could bring down the cost. In Cuba and Oman, it is testing the efficacy of using one-fifth the normal dose, delivered intradermally with an injection gun instead of intramuscularly with a needle. Other projects are trying to "stretch" the antigen with new adjuvants.

Another big push is for what is called a "Sabin IPV." One of the drawbacks of the standard Salk IPV is that production starts with the dangerous wild virus, which is then killed. To reduce the chances of an accidental release, IPV is manufactured only in facilities that operate under strict biocontainment procedures and only in countries that maintain a very high population immunity against polio. Both requirements rule out transferring the technology to developing-country manufacturers, which would bring down the cost.

A Sabin IPV would use the less infectious attenuated strain from the oral vaccine as its seed stock, providing "a margin of safety" should an accident occur, says Sutter. Several clinical trials of a Sabin IPV are ongoing; if all goes well, it could be introduced within 5 to 8 years.

Ultimately, says Bruce Aylward, who runs WHO's eradication initiative, "I want something much better than Sabin IPV." Several groups are working on manipulating the virus to make safer seed strains that could be handled under less stringent safeguards. It's still early days, but there are several promising leads, including a virus that can't survive at body temperature.

Sutter and Aylward say each country will decide whether to continue vaccinating. But if countries do choose to continue, they want to have in place a "cost neutral" vaccine that delivers the same protection as OPV at the same price—without the risk. **–L.R.**

remote bomb while they were returning from a village where they were trying to persuade parents to let their children be vaccinated.

Equally unsettling, the Geneva team suspects that the program in Pakistan is weaker than they imagined and that the viral foe may be tougher. Earlier reports that vaccination teams reached 95% of the target children seem to have been fabricated, says Tangermann. And recent studies by Grassly and colleagues at Imperial suggest that viral transmission is much more efficient in Pakistan than previously believed, closer to that of India than that of Nigeria. "Pakistan is the only place we really have questions about what we are dealing with," says Aylward.

To get a better fix on the biology, WHO and Pakistani partners are planning studies to measure antibodies to the virus in children in Karachi, Peshawar, and Lahore. "We have to see how effective the vaccine is and how well the program is working," says Aylward.

On the political front, Aylward has been trying to work his magic. At a high-level meeting last December in Islamabad, he and other partners got assurances from Minister of Health Mir Aijaz Hussain Jakhrani that Pakistan would make eradication a priority.

For now, says Heymann, the most the program can hope to achieve there is to show it can stop transmission in conflict-free regions, like Punjab, where despite repeated campaigns, circulation remains intense. For the other areas, they wait. "We may be quite slow in areas with security problems," he says.

Vote of confidence

Despite the upsurge of cases in 2008, Aylward insists that the world is much closer to eradication than it was a year ago. Chan has declared polio eradication WHO's "top operational priority," saying in a speech in June, "The credibility of not just WHO but of many other health initiatives is on the line." She is organizing an independent review to figure out what went wrong in each country and what the program could do better.

The global oversight body, the Advisory Committee on Polio Eradication, is on board as well; in December, it endorsed GPEI's strategic plan for 2009 to 2013. Although Aylward is leery of firm deadlines, the plan calls for interrupting type 1 transmission in India by the end of 2009 and type 3 the following year. They hope to wipe out both types in Afghanistan and Pakistan in 2010, but Nigeria might take a year longer. All that depends, of course, on donors keeping their checkbooks open and countries putting their mind and muscle behind eradication.

The donors have stepped up. With the $635 million infusion from Gates and others, funding looks better than it has in years. The $255 million Gates grant is the largest single donation since Rotary International kicked off the effort in 1988 with $240 million.

Aylward keeps up his unrelenting schedule, visiting endemic and reinfected countries to spur or prod them into action. For the toughest spots, the big guns go too, such as Chan, Heymann, and the newest advocate, Bill Gates, who is also championing malaria eradication and who visited India in December 2008 and Nigeria just last week.

For now, the global health community seems willing to give the eradication initiative more time. There are still skeptics who say it will never be finished. But most agree with Wright. "It's terribly hard," he says. "All the models suggest it is not a good idea to give up on the program."

"We won't let up," said Aylward in an interview from the noisy Brazzaville airport in Congo en route to Islamabad. "I will personally push it over the line if I have to. We still have very long sleeves and lots of tricks up them if we need them." **–LESLIE ROBERTS**

Published by AAAS

RETHINKING THE POLIO ENDGAME

One of the toughest conundrums for the long-running Global Polio Eradication Initiative has been whether and when it would be safe to stop vaccinating once they deem the virus gone. Now, the thinking is undergoing a major shift.

The first big complication came in 1999 when scientists realized that the weakened virus used in the live oral polio vaccine (OPV) could revert to its neurovirulent form in rare cases and spark an epidemic. Thus was born the "OPV paradox": OPV was necessary to eradicate the virus, but as long as OPV was in use, eradication could never be achieved. As a solution, World Health Organization (WHO) scientists proposed a plan: After the world was certified polio-free, all countries would stop using OPV simultaneously, as if at the stroke of midnight.

"The world will be a much safer place if more countries use [inactivated polio vaccine]" when the virus is deemed gone.
—ROLAND SUTTER

Some scientists dismissed the idea as folly and instead advocated universal use of the inactivated polio vaccine (IPV), already widely used in developed countries. That would be the only way to ensure the world was really safe, they argued, and the only way to prevent a gross inequity in which poor countries bore all the risk of polio.

For years, WHO maintained that such a switch wasn't feasible: IPV was too expensive for poor countries, it must be injected, and its effectiveness is unproven in tropical settings. Now, experts such as Roland Sutter, who heads WHO research in Geneva, Switzerland, concede that IPV does have a role after all. "The world will be a much safer place if more countries use IPV," he says.

To make that possible, WHO is now dusting off some earlier studies and investing heavily in new research into a cheaper, more effective version of IPV. WHO is looking at "dose sparing" strategies that could bring down the cost. In Cuba and Oman, it is testing the efficacy of using one-fifth the normal dose, delivered intradermally with an injection gun instead of intramuscularly with a needle. Other projects are trying to "stretch" the antigen with new adjuvants.

Another big push is for what is called a "Sabin IPV." One of the drawbacks of the standard Salk IPV is that production starts with the dangerous wild virus, which is then killed. To reduce the chances of an accidental release, IPV is manufactured only in facilities that operate under strict biocontainment procedures and only in countries that maintain a very high population immunity against polio. Both requirements rule out transferring the technology to developing-country manufacturers, which would bring down the cost.

A Sabin IPV would use the less infectious attenuated strain from the oral vaccine as its seed stock, providing "a margin of safety" should an accident occur, says Sutter. Several clinical trials of a Sabin IPV are ongoing; if all goes well, it could be introduced within 5 to 8 years.

Ultimately, says Bruce Aylward, who runs WHO's eradication initiative, "I want something much better than Sabin IPV." Several groups are working on manipulating the virus to make safer seed strains that could be handled under less stringent safeguards. It's still early days, but there are several promising leads, including a virus that can't survive at body temperature.

Sutter and Aylward say each country will decide whether to continue vaccinating. But if countries do choose to continue, they want to have in place a "cost neutral" vaccine that delivers the same protection as OPV at the same price—without the risk. **–L.R.**

remote bomb while they were returning from a village where they were trying to persuade parents to let their children be vaccinated.

Equally unsettling, the Geneva team suspects that the program in Pakistan is weaker than they imagined and that the viral foe may be tougher. Earlier reports that vaccination teams reached 95% of the target children seem to have been fabricated, says Tangermann. And recent studies by Grassly and colleagues at Imperial suggest that viral transmission is much more efficient in Pakistan than previously believed, closer to that of India than that of Nigeria. "Pakistan is the only place we really have questions about what we are dealing with," says Aylward.

To get a better fix on the biology, WHO and Pakistani partners are planning studies to measure antibodies to the virus in children in Karachi, Peshawar, and Lahore. "We have to see how effective the vaccine is and how well the program is working," says Aylward.

On the political front, Aylward has been trying to work his magic. At a high-level meeting last December in Islamabad, he and other partners got assurances from Minister of Health Mir Aijaz Hussain Jakhrani that Pakistan would make eradication a priority.

For now, says Heymann, the most the program can hope to achieve there is to show it can stop transmission in conflict-free regions, like Punjab, where despite repeated campaigns, circulation remains intense. For the other areas, they wait. "We may be quite slow in areas with security problems," he says.

Vote of confidence

Despite the upsurge of cases in 2008, Aylward insists that the world is much closer to eradication than it was a year ago. Chan has declared polio eradication WHO's "top operational priority," saying in a speech in June, "The credibility of not just WHO but of many other health initiatives is on the line." She is organizing an independent review to figure out what went wrong in each country and what the program could do better.

The global oversight body, the Advisory Committee on Polio Eradication, is on board as well; in December, it endorsed GPEI's strategic plan for 2009 to 2013. Although Aylward is leery of firm deadlines, the plan calls for interrupting type 1 transmission in India by the end of 2009 and type 3 the following year. They hope to wipe out both types in Afghanistan and Pakistan in 2010, but Nigeria might take a year longer. All that depends, of course, on donors keeping their checkbooks open and countries putting their mind and muscle behind eradication.

The donors have stepped up. With the $635 million infusion from Gates and others, funding looks better than it has in years. The $255 million Gates grant is the largest single donation since Rotary International kicked off the effort in 1988 with $240 million.

Aylward keeps up his unrelenting schedule, visiting endemic and reinfected countries to spur or prod them into action. For the toughest spots, the big guns go too, such as Chan, Heymann, and the newest advocate, Bill Gates, who is also championing malaria eradication and who visited India in December 2008 and Nigeria just last week.

For now, the global health community seems willing to give the eradication initiative more time. There are still skeptics who say it will never be finished. But most agree with Wright. "It's terribly hard," he says. "All the models suggest it is not a good idea to give up on the program."

"We won't let up," said Aylward in an interview from the noisy Brazzaville airport in Congo en route to Islamabad. "I will personally push it over the line if I have to. We still have very long sleeves and lots of tricks up them if we need them." **–LESLIE ROBERTS**

Published by AAAS

Heave ho. Weaver ants work together to build nests for colonies with millions of members.

EVOLUTIONARY BIOLOGY

Agreeing to Disagree

Two friends debate the relative importance of kinship in the evolution of complex social systems in insects

When Bert Hölldobler and Edward O. Wilson wrote their first book, *The Ants*, it was a labor of love. Passionate about the insects they studied, they produced a Pulitzer Prize–winning work that brought worldwide attention to the sophisticated social world of ants. But when they sat down to write the sequel more than a dozen years later, the honeymoon was over.

For these two biologists, ants, bees, wasps, and termites hold a special place in nature, having achieved the ultimate in altruistic behavior. In these so-called eusocial species, one or a few females produce young that other individuals in the colony care for in lieu of producing their own offspring. Twenty years ago, Wilson, based at Harvard University, and Hölldobler, now based at Arizona State University, Tempe, attributed this seemingly self-sacrificing behavior to the unusually close kinship among colony members.

Yet, over the past 4 years, the duo has struggled to reconcile new information indicating that some social insects are exceptions to the rules that guided their thinking when they wrote *The Ants*. Although their new book, *The Superorganism*, tackles the evolution of eusociality, the two researchers are at odds over how these social systems got started.

"We agree on everything except what happens on that one magic step," Wilson says. He no longer argues that kinship is the underlying driving force, suggesting instead that what tips the scales in favor of eusociality is a particular set of environmental conditions that leads to the guarding of the young in nests close to dependable food sources. Over the past 2 years, he has pushed hard to get the field to see his new perspective. And Hölldobler has pushed back. "There is nothing wrong with kin selection," Hölldobler insists. "It's a very important concept to understanding the early evolution of eusociality—how it begins."

Others wonder what the fuss is about. "It's a tempest in a teapot," says Hudson Kern Reeve, an evolutionary biologist at Cornell University. "It's distracting us from the really interesting questions." But Wilson disagrees: "If we can't understand the beginning of eusociality, our whole view of it is wobbly."

Living large

Eusociality represents the ultimate in successful living, say Hölldobler and Wilson. Summed together, the 20,000 species of ants, bees, wasps, and termites represent two-thirds of the insect biomass, even though they account for just 2% of the insect species. Ants in tropical rainforests outweigh all the vertebrate inhabitants combined. (All ants and termites are eusocial; some bees and wasps are.)

Their colonies, nests, and hives actually represent "superorganisms," a concept introduced 80 years ago by William Morton Wheeler and reemphasized by Hölldobler and Wilson in their new book. Many organisms form groups, but in the superorganism, the group operates with a minimum of internal strife; most group members work toward the common good with little apparent mind to individual self-interest.

That attitude had Charles Darwin flummoxed. He called social insects a "special difficulty" capable of toppling his theory of evolution. Natural selection should favor individuals that always put themselves—and the propagation of their own genes—first. Yet that didn't seem to be happening with worker bees or soldier ants, which often have no reproductive capability at all and instead strive to keep their queen and her offspring well-fed and protected. Darwin suspected that family ties might explain this apparent paradox—with selection operating at the family and not the individual level.

In the 1960s, William Hamilton, a British biologist building on work by a fellow countryman, geneticist J. B. S. Haldane, put forth a very convincing theory, often called kin selection, that explained eusociality: Such altruism could evolve when a queen propagated enough of her workers' genes to make it worth the workers' while to help her rather than reproduce independently.

In the 1970s, Wilson became one of kin selection's greatest proponents. Textbooks embraced the idea, and theoreticians worked hard to refine their understanding of how kin selection worked. "It became dogma," says Wilson. Yet when Wilson first started to put together the new book's chapter discussing the origin of eusociality, he ran into trouble. "In the last few years, the picture had changed considerably," he says. Since the publication of *The Ants*, field biologists had discovered many variations on the eusociality theme. Eusocial termites, for example, do not have the so-called haploid-diploid genetic system that was critical to Hamilton's explanation of kin selection, and some species that do have those genetics haven't evolved eusociality.

CREDIT: BERT HÖLLDOBLER

Published by AAAS

Moreover, it became clear "there are counterforces that oppose close-pedigree kinship in the early evolution of social insects," Wilson points out. Genetic diversity leads to greater resistance to disease and other destructive forces, and outbreeding does occur in these systems. He cites the example of primitive wasps, in which individuals in the same colony tend to become less related as time passes. He reasons that if close relatedness were so important, then these groups should continue to be made up of close kin. "Yet counterforces against the favoring of close kin in the origin of eusociality have not been taken into account by the theoreticians," he criticizes.

In 2004, Wilson wrote a manuscript on his new view and sent it to a half-dozen colleagues, working with them one on one to refine his ideas. At the time, Hölldobler was on board, and together they co-authored a perspective in the 20 September 2005 *Proceedings of the National Academy of Sciences*. They pointed out that three forces were at work in shaping an insect society: group, individual, and "collateral" kin selection (involving relatives other than offspring). Group selection occurs when one cooperating hive or nest outcompetes another one; individual selection involves the survival and reproductive output of a particular ant, wasp, or bee; and kin selection has an impact when relatives other than offspring help spread shared genes.

In that paper, Wilson and Hölldobler analyzed the conditions under which each type of selection tended to enhance or undermine cooperation, and how strongly. They concluded that the degree of relatedness primarily affects how quickly a social system changes but not whether cooperation was favored in the first place. "Eusociality … can, in theory at least, be initiated by group selection in either the presence or absence of close relatedness," they wrote. But it could not arise without group selection.

What was important was not relatedness, they argued, but the evolution of a gene variant that caused offspring to linger at the nest rather than disperse and an environment that favored cooperation. Researchers have yet to identify such an allele, but several researchers are looking, particularly in species that sometimes cooperate and sometimes live solitary lives, depending on the circumstances. "The preadaptation for eusociality will always [involve] a family and in that family, the individuals will be closely related," notes Wilson. "But that's very different from saying that close relatedness brought them to take that last step [into eusociality]."

As evidence that kinship is not critical, they cited examples of species that appear to be on the first steps toward eusociality, such as xylocopine carpenter bees. These bees form temporary semisocial groups that begin when one member of a pair bullies the other into becoming a worker while the first becomes the queen. Workers that are related to the queen tend to leave, but nonrelatives tend to stick around, suggesting that kinship has little to do with the cohesiveness of the group. Likewise, primitively eusocial wasps will set up group living arrangements with either kin or nonkin.

Wilson and Hölldobler also argued that because eusociality is quite rare—these altruistic societies have emerged only about a dozen times and exist in just 15 out of the 2600 known insect and arthropod families—it takes just the right combination of factors to set the stage for it. For insects, the appropriate setting involves maternal behavior in which the female builds a nest, gathers food, and feeds the young. In that situation, it helps to have someone guarding the young while the queen is foraging.

Moreover, if the environment has ample food but a lot of competition, "that kind of pressure would make a group superior over an individual," says Wilson. Consider primitive termites that nest in dead trees in California. Many king-queen pairs settle into a tree, but eventually one colony will take over the entire trunk, depending on how strong a cooperative unit that pair establishes. "It's solid group selection," says Wilson.

Conflict resolution. Hölldobler (*top, left*) and Wilson debated how to cover the origins of eusociality in their new book, which discusses social insects including termites. In dampwood termites (*bottom*), workers and nymphs feast on a queen killed when two colonies converged.

CREDITS (TOP TO BOTTOM): KATHLEEN HORTON; BARBARA L. THORNE

The rift

Wilson continued to expand on these ideas. In the January 2008 issue of *BioScience*, he reiterated that collateral kin selection does not play a significant role and reemphasized that the road to eusociality entails setting up cooperative housekeeping and foraging from a local, dependable food source.

The numbers support his view, he says. No eusocial species exist among the 70,000 species of parasitoid wasps and their relatives that travel from prey to prey to lay their eggs; whereas it arose seven times in the 55,000 aculeate (stinging) wasps that keep paralyzed prey in nests. Among the 9000 species of aphids and thrips, only those few that induce plant hosts to form galls—a nest of sorts—have become eusocial even though most of them form groups. And out of 10,000 decapod crustaceans, only snapping shrimp, which live in sponges, are eusocial, he reported.

In these settings, species with the flexibility to live in groups can win out. Wilson thinks today's eusocial species likely had ancestors similar to a Japanese stem-nesting xylocopine bee, *Ceratina flavipes*: Most of the time females make do on their own, but every once in a while, they pair up and divide the labor, setting the stage for a "group" to outdo individuals and for the tendency to form groups to be favored.

And, he argues, it may not take many genes to get a group to form. Studies of bacteria and fungi show that a single gene can have a profound effect on dispersal. A loss of the dispersal instinct positions individuals to become helpers.

About this time, Wilson began collaborating closely with David Sloan Wilson of Binghamton University in New York. After organizing a discussion group with about 20 others, they published an expanded explanation of these ideas in the December 2007 issue of

The Quarterly Review of Biology and in the September-October 2008 issue of *American Scientist*. Kin-selection theory, they argue, is ineffective. "The theory that traditionalists use leads them anywhere they want to go" and fails to make useful predictions, Wilson asserts. "To make [a theory] really stand [up], you have to show that that's the only result that can come from your theory, and they haven't done that."

Meanwhile, Hölldobler became convinced that kin selection really is critical for the origin of eusociality after all. He became increasingly unhappy with what Wilson and he had written in 2005, concluding that they had not made a clear enough distinction between cooperation and eusociality and that their evidence included supposedly primitively eusocial species that were really more advanced than that. "You don't need to be closely related for the evolution of cooperative groups, but eusociality is very special," he notes.

"Reeve and I set the record straight," he says, in the 5 June 2007 issue of the *Proceedings of the National Academy of Sciences.* They emphasized that because, mathematically, group and kin selection were equivalent, the argument that group rather than kin selection lay at the heart of the origin of eusociality was spurious. "The fuss about kin selection versus group selection is quite pointless," says Hölldobler.

He sees the evolution of eusociality as a two-step process activated by ecological conditions and operating on closely related individuals. First, a subsocial species must work out an arrangement in which offspring that are capable of reproducing nonetheless help the mother raise siblings. Next, that arrangement becomes cast in stone, with sterile workers.

In primitive ants, workers are almost the size of the queen, still have functional ovaries, and can mate, but as long as the mother remains fertile, they usually refrain from laying eggs. Evolution to eusociality often stalls at this first primitive stage, Hölldobler notes. Colonies can never get bigger than what the queen and worker nestmates can police because individuals might otherwise "cheat" the system and reproduce on their own.

But under certain circumstances, bigger, more harmonious groups outdo rivals and evolve full-fledged eusocial networks, in which millions of sterile specialized individuals work together as a superorganism. In these cases, competition between groups favors cooperation within a group, and competition between individuals in the group dissipates. "At that point, it's not important to have high relatedness," says Hölldobler. And, as he and Wilson agree, it can be healthier for a colony to have genetic diversity.

Hölldobler's opinion seems to hold the majority. "On this one issue, it is largely Wilson on one side and most of the rest of us on the other," says David Queller, an evolutionary biologist at Rice University in Houston, Texas. "Like many researchers, I have become convinced that variation in costs and benefits is probably more important than variation in relatedness favored by Hamilton, but relatedness is still an essential part of it, and it all fits into Hamilton's kin-selection paradigm."

Yet there are dissenters other than Wilson and Wilson. "Relatedness is not needed to explain the origin of worker behavior," says James Hunt, a biologist who studies wasps at North Carolina State University in Raleigh. "A mountain of data [has shown] that neither high relatedness nor asymmetry of relatedness among nestmates is necessary for the origin of sociality."

This rift slowed progress on *The Superorganism*. Originally, Wilson was supposed to write the chapter on the genetics and evolution of eusociality, but after 5 years, it was still not done, with the two authors at odds about how to proceed. Finally in July 2007, Wilson handed his pen over to Hölldobler, who labored over a version that Wilson eventually said he could live with as long as there were references to Wilson's perspective. The chapter "represents the best compromise that Bert and I could reach," says Wilson. And now, on the book-signing circuit for their new book, Wilson and Hölldobler downplay their points of departure. "We were worried that that chapter would distract the reader from the wonderful natural history," says Hölldobler. So they beg off questions related to the controversy.

And they remain fast friends. In the end, says Wilson, "we agreed to disagree."

–ELIZABETH PENNISI

Promising beginnings. At their nests, a female paper wasp (*top*) and a female sweat bee set up shop where early offspring may become workers that help raise the next generation.

CREDITS (TOP TO BOTTOM): JAMES H. HUNT; NICOLAS J. VEREECKEN

Published by AAAS

On the Origin of Art and Symbolism

Since their discovery by French spelunkers in 1994, the magnificent lions, horses, and rhinos that seem to leap from the walls of Chauvet Cave in southern France have reigned as the world's oldest cave paintings. Expertly composed in red ochre and black charcoal, the vivid drawings demonstrate that the artistic gift stretches back more than 30,000 years. These paintings are almost sure to be mentioned in any article or paper about the earliest art. But what do they really tell us about the origins of artistic expression?

The prehistoric humans who decorated Chauvet's walls by torchlight arrived at the cave with their artistic genius already in full flower. And so, most researchers agree that the origins of art cannot simply be pegged to the latest discovery of ancient paintings or sculpture. Some of the earliest art likely perished over the ages; much remains to be found; and archaeologists don't always agree on how to interpret what is unearthed. As a result, instead of chasing after art's first appearance, many researchers seek to understand its symbolic roots. After all, art is an aesthetic expression of something more fundamental: the cognitive ability to construct symbols that communicate meaning, whether they be the words that make up our languages, the musical sounds that convey emotion, or the dramatic paintings that, 30,000 years after their creation, caused the discoverers of the Chauvet Cave to break down in tears.

While sites like Chauvet might be vivid examples of what some researchers still consider a "creative explosion" that began when modern humans colonized Europe about 40,000 years ago, an increasing number of prehistorians are tracing our symbolic roots much further back in time—and in some cases, to species ancestral to *Homo sapiens*. Like modern humans themselves, symbolic behavior seems to have its origins in Africa. Recent excavations have turned up elaborate stone tools, beads, and ochre dating back 100,000 or more years ago, although researchers are still debating which of these finds really demonstrate symbolic expression. But there's widespread agreement that the building blocks of symbolism preceded full-blown art. "When we talk about beads and art, we are actually talking about material technologies for symbolic expression that certainly postdate the origins of symbolic thought and communication, potentially by a very wide margin," says archaeologist Dietrich Stout of University College London.

The evolution of symbolism was once thought to have been as rapid as "flicking on a light switch," as archaeologist Clive Gamble of the Royal Holloway, University of London, put it some years ago. But given new evidence that symbolic behavior appears long before cave paintings, Gamble now says that his much-cited comment needs to be modified: "It's a dimmer switch now, a stuttering candle."

As they more precisely pinpoint when symbolic behavior began, scientists are hoping they might one day crack the toughest question of all: What was its evolutionary advantage to humans? Did symbols, as many researchers suspect, serve as a social glue that helped tribes of early humans to survive and reproduce?

THE YEAR OF DARWIN

This essay continues our monthly series. See more on humans' evolutionary journey online at **blogs.sciencemag.org/origins**.

Venus, phallus, or pebble?

"I don't know much about Art, but I know what I like," quipped the humorist and art critic Gelett Burgess back in 1906. For archaeologists, distinguishing art from nonart is still quite a challenge. Take the 6-centimeter-long piece of quartzite known as the Venus of Tan-Tan. Found in Morocco in 1999 next to a rich trove of stone tools estimated to be between 300,000 and 500,000 years old, it resembles a human figure with stubby arms and legs. Robert Bednarik, an independent archaeologist based in Caulfield South, Australia, insists that an ancient human deliberately modified the stone to make it look more like a person. If so, this objet d'art is so old that it was created not by our own species, which first appears in Africa nearly 200,000 years ago, but by one of our ancestors, perhaps the large-brained *H. heidelbergensis*, thought by some anthropologists to be the common ancestor of modern humans and Neandertals. That would mean that art is an extremely ancient part of the *Homo* repertoire. "Ignoring the few specimens we have of very early paleoart, explaining them away, or rejecting them out of hand does not serve this discipline well," Bednarik wrote in a 2003 analysis of the Venus of Tan-Tan in *Current Anthropology*.

Symmetry in stone. Some stone tools require a mental image to create.

Yet many archaeologists are skeptical, arguing that the stone's resemblance to a human figure might be coincidence. Indeed, the debate over the Tan-Tan "figurine" is reminiscent of a similar controversy over a smaller stone discovered in 1981 at the site of Berekhat Ram in the Israeli-occupied Golan Heights. To some archaeologists, this 250,000-year-old object resembles a woman, but others argue that it was shaped by natural forces, and, in any case, looks more like a penguin or a phallus. Even after an exhaustive microscopic study concluded that the Berekhat Ram object had indeed been etched with a tool to emphasize what some consider its "head" and "arms," many researchers have rejected it as a work of art. For some, proof of symbolic behavior requires evidence that the symbols had a commonly understood mean-

CREDITS (TOP TO BOTTOM): KATHARINE SUTLIFF/*SCIENCE*; THE BOXGROVE PROJECT; GEORGE RICHMOND/BRIDGEMAN ART LIBRARY, LONDON (SUPERSTOCK)

Published by AAAS

A roaring start. Researchers agree that Chauvet Cave's magnificent paintings, including these lions, are full-blown art.

ing and were shared within groups of people. For example, the hundreds of bone and stone "Venus figurines" found at sites across Eurasia beginning about 30,000 years ago were skillfully carved and follow a common motif. They are widely regarded not only as symbolic expression, but full-fledged art.

Thus many researchers are reluctant to accept rare, one-off discoveries like the Tan-Tan or Berekhat Ram objects as signs of symbolic behavior. "You can imagine [an ancient human] recognizing a resemblance but [the object] still hav[ing] no symbolic meaning at all," says Philip Chase, an anthropologist at the University of Pennsylvania. Thomas Wynn, an anthropologist at the University of Colorado, Colorado Springs, agrees: "If it's a one-off, I don't think it counts. It's not sending a message to anyone."

Tools of the imagination

Given how difficult it is to detect the earliest symbolic messages in the archaeological record, some researchers look instead for proxy behaviors that might have required similar cognitive abilities, such as toolmaking. Charles Darwin himself saw an evolutionary parallel between toolmaking and language, probably the most sophisticated form of symbolic behavior. "To chip a flint into the rudest tool," Darwin wrote in *The Descent of Man*, demands a "perfect hand" as well adapted to that task as the "vocal organs" are to speaking.

To many researchers, making sophisticated tools and using symbols both require the capacity to hold an abstract concept in one's head—and, in the case of the tool, to "impose" a predetermined form on raw material based on an abstract mental template. That kind of ability was probably not needed to make the earliest known tools, say Wynn and other researchers. These implements, which date back 2.6 million years, consist mostly of rocks that have been split in two and then sharpened to make simple chopping and scraping implements.

Then, about 1.7 million years ago, large, teardrop-shaped tools called Acheulean hand axes appeared in Africa. Likely created by *H. erectus* and probably used to cut plants and butcher animals, these hand-held tools vary greatly in shape, and archaeologists have debated whether creating the earliest ones required an abstract mental template. But by about 500,000 years ago, ancient humans were creating more symmetrical Late Acheulean tools, which Wynn and many others argue are clear examples of an imposed form based on a mental template. Some have even argued that these skillfully crafted hand axes had symbolic meanings, for example to display prestige or even attract members of the opposite sex.

The half-million-year mark also heralded the arrival of *H. heidelbergensis*, which had a much larger brain than *H. erectus*. Not long afterward, our African ancestors began to create a wide variety of finely crafted blades and projectile points, which allowed them to exploit their environment in more sophisticated ways, and so presumably enhance their survival and reproduction. Archaeologists refer to these tools as Middle Stone Age technology and agree that they did require mental templates. "The tools tell us that the hominid world was changing," says Wynn.

As one moves forward in time, humans appear able to imagine and create even more elaborate tools, sharpening their evolutionary edge in the battle for survival. By 260,000 years ago, for example, ancient humans at Twin Rivers in what is now Zambia could envision a complex finished tool and put it together in steps from different components. They left behind finely made blades and other tools that had been modified—usually by blunting or "backing" one edge—to be hafted onto handles, presumably made of wood. These so-called backed tools have been widely regarded as evidence of symbolic behavior when found at much younger sites. "This flexibility in stone tool manufacture [indicates] symbolic capabilities," says archaeologist Sarah Wurz of the Iziko Museums of Cape Town in South Africa.

Similar cognitive abilities were possibly required to make the famous 400,000-year-old wooden spears from Schöningen, Germany. One recent study concludes that these spears' creators—probably members of *H. heidelbergensis*—carried out at least eight preplanned steps spanning several days, including chopping tree branches with hand axes and shaping the spears with stone flakes.

The idea that sophisticated toolmaking and symbolic thought require similar cognitive skills also gets some support from a surprising quarter: brain-imaging studies. Stout's team ran positron emission tomography scans on three archaeologists—all skillful stone knappers—as they made pre-Acheulean and

> "If it's a one-off, I don't think it counts. It's not sending a message to anyone."
>
> —THOMAS WYNN, UNIVERSITY OF COLORADO, COLORADO SPRINGS

Symbolic start. Some scientists argue that this 77,000-year-old engraved ochre shows symbolic capacity.

CREDITS (TOP TO BOTTOM): FRENCH MINISTRY OF CULTURE AND COMMUNICATION/DRAC RHONE-ALPES/DEPARTMENT OF ARCHAEOLOGY; CHRIS HENSHILWOOD AND FRANCESCO D'ERRICO

Published by AAAS

Late Acheulean tools. Both methods turned on visual and motor areas of the brain. But only Late Acheulean knapping turned on circuits also linked to language, the team reported last year.

Color me red

At Twin Rivers, it's not just the tools that hint at incipient symbolic behavior. Early humans there also left behind at least 300 lumps of ochre and other pigments in a rainbow of colors: yellow, red, pink, brown, purple, and blue-black, some of which were gathered far from the site. Excavator Lawrence Barham of the University of Liverpool in the United Kingdom thinks they used the ochre to paint their bodies, though there's little hard evidence for this. Most archaeologists agree that body painting, as well as the wearing of personal ornaments such as bead necklaces, was a key way that early humans symbolically communicated social identity such as membership in a particular group, much as people today declare social allegiances and individual personalities by their clothing and jewelry.

Yet while the Twin Rivers evidence is suggestive, it's hard to be sure how the ochre was actually used. There's little sign that it was ground into powder, as needed for decoration, says Ian Watts, an independent ochre expert in Athens. And even ground ochre could have had utilitarian uses, says archaeologist Lyn Wadley of the University of Witwatersrand in Johannesburg, South Africa. Modern-day experiments have shown that ground ochre can be used to tan animal hides, help stone tools adhere to bone or wooden handles, and even protect skin against mosquito bites.

"We simply don't know how ancient people used ochre 300,000 years ago," Wadley says. And since at that date the ochre users were not modern humans but our archaic ancestors, some experts are leery of assigning them symbolic savvy.

Yet many archaeologists are willing to grant that our species, *H. sapiens*, was creating and using certain kinds of symbols by 75,000 years ago and perhaps much earlier. At sites such as Blombos Cave on South Africa's southern Cape, people left sophisticated tools, including elaborately crafted bone points, as well as perforated beads made from snail shells and pieces of red ochre engraved with what appear to be abstract designs. At this single site, a number of what many archaeologists consider diagnostic elements of symbolic behavior came together. And in work now in press, the Blombos team reports finding engraved ochre in levels dating back to 100,000 years ago (*Science*, 30 January, p. 569).

There are other hints that the modern humans who ventured out of Africa around this time might also have engaged in symbolic behavior. At the Skhul rock shelter in Israel, humans left 100,000-year-old shell beads considered by some to be personal ornaments (*Science*, 23 June 2006, p. 1731). At the 92,000-year-old Qafzeh Cave site nearby, modern humans apparently strongly preferred the color red: Excavators have studied 71 pieces of bright red ochre associated with human burials. Some researchers argue that this represents an early case of "color symbolism," citing the universal importance of red in historical cultures worldwide and the apparently great lengths to which early humans went to gather red ochre. "There is very strong circumstantial evidence for the very great antiquity of the color red as a symbolic category," says anthropologist Sally McBrearty of the University of Connecticut, Storrs.

These finds of colorful ochre, fancy tools, and beads have convinced many researchers that the building blocks of symbolism had emerged by at least 100,000 years ago and possibly much earlier. But why? What selective advantages did using symbols confer on our ancestors?

To some scientists, the question is a no-brainer, especially when it is focused on the most sophisticated form of symbolic communication: language. The ability to communicate detailed, concrete information as well as abstract concepts allowed early humans to cooperate and plan for the future in ways unique to our species, thus enhancing their survival during rough times and boosting their reproductive success in good times. "What aspects of human social organization and adaptation wouldn't benefit from the evolution of language?" asked Terrence Deacon, a biological anthropologist at the University of California, Berkeley, in his influential book *The Symbolic Species: The Coevolution of Language and the Brain*. Deacon went on to list just some of the advantages: organizing hunts, sharing food, teaching toolmaking, sharing past experiences, and raising children. Indeed, many researchers have argued that symbolic communication is what held groups of early humans together as they explored new environments and endured climatic shifts.

As for art and other nonlinguistic forms of symbolic behavior, they may also have been key to cementing these bonds, by expressing meanings that are difficult or impossible to put into words. In that way, artistic expression, including music, may have helped ensure the survival of the fittest. This may also explain why great art has such emotional force, because the most effective symbols are those that convey their messages the most powerfully—something the artists at Chauvet Cave seem to have understood very well.

–MICHAEL BALTER

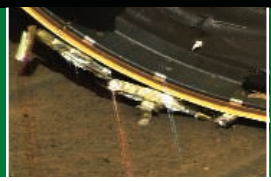
Eye of the beholder. Archaeologists debate whether this modified stone was meant to represent a woman.

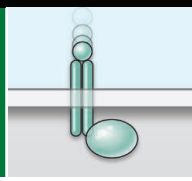
CREDIT: FRANCESCO D'ERRICO AND APRIL NOWELL

Online
sciencemag.org

Hear author Michael Balter discuss the roots of art at **www.sciencemag.org/darwin**.

Published by AAAS





LETTERS

edited by Jennifer Sills

Keeping Raw Data in Context

WE AGREE WITH THE PREMISE STATED IN "MAKING CLINICAL data widely available" (Special Section on Clinical Trials, J. Kaiser, 10 October 2008, p. 217): Sharing patient-level data from human studies would help investigators make more and better discoveries more quickly and with less duplication. However, this happy circumstance will occur only if investigators interpret the raw data properly within the context of the study that generated that data. This implies that any initiative to share raw clinical research data must also pay close attention to sharing clear and complete information about the design of the original studies. Relying on journal articles for study design information is problematic, for three reasons. First, journal articles often provide insufficient detail when describing key study design features such as randomization (*1*) and intervention details (*2*). Second, some data sets may come from studies with no publications [only 21% of oncology trials registered in ClinicalTrials.gov before 2004 and completed by September 2007 were published (*3*)]. Finally, investigators cannot reliably search journal articles for methodological concepts like "double blinding" or "interrupted time series," crucial concepts for proper interpretation of the data. A mishmash of nonstandardized databases of raw results and unevenly reported study designs is not a strong foundation for clinical research data sharing.

Data sharing. Shared data will only be useful if complete information about the original study design is also available.

We believe that the effective sharing of clinical research data requires the establishment of an interoperable federated database system that includes both study design and results data. A key component of this system is a logical model of clinical study characteristics in which all the data elements are standardized to controlled vocabularies and common ontologies to facilitate cross-study comparison and synthesis (*4*). As a first step toward this vision, the Human Studies Database Project is developing a standardized representation of study designs for all ongoing human studies at NIH-funded Clinical and Translational Science Award (CTSA) institutions, with the ultimate goal of using these standards to integrate local institutional databases of human studies for large-scale data sharing and reuse (*5*). This approach would also support legally mandated trial registration and results reporting (*6*), NIH-related data sharing initiatives (*7*, *8*), and requests for raw data by biomedical journals (*9*). To complement work on defining and standardizing study design descriptions, national and international policies are still urgently needed to ensure effective, ethical, and coordinated sharing of tightly coupled study design and raw patient-level data (*10*).

IDA SIM,[1]* CHRISTOPHER G. CHUTE,[2] HAROLD LEHMANN,[3] RAKESH NAGARAJAN,[4] MEREDITH NAHM,[5] RICHARD H. SCHEUERMANN[6]

[1]Division of General Internal Medicine, Department of Medicine, University of California, San Francisco, CA 94143–0320, USA. [2]Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN 55905, USA. [3]Division of Health Sciences Informatics, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA. [4]Division of Laboratory and Genomic Medicine, Department of Pathology and Immunology, Washington University in St. Louis, St. Louis, MO 63110, USA. [5]Duke Translational Medicine Institute, Duke University Health System, Durham, NC 27715, USA. [6]Department of Pathology and Division of Biomedical Informatics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA.

*To whom correspondence should be addressed. E-mail: ida.sim@ucsf.edu

References

1. D. Moher, A. Jones, L. Lepage, for the CONSORT Group, *JAMA* **285**, 1992 (2001).
2. P. Glasziou, E. Meats, C. Heneghan, S. Shepperd, *Br. Med. J.* **336**, 1472 (2008).
3. S. Ramsey, J. Scoggins, *Oncologist* **13**, 925 (2008).
4. I. Sim, B. Olasov, S. Carini, *J. Biomed. Inform.* **37**, 108 (2004).
5. Human Studies Database Project (2008); available at http://rctbank.ucsf.edu/home/hsdb.html.
6. D. A. Zarin *et al.*, *JAMA* **297**, 2112 (2007).
7. NIH Data Sharing Policy (2007); available at http://grants.nih.gov/grants/policy/data_sharing/.
8. NCI's Clinical Trials Reporting Program, National Cancer Institute (2008); available at www.cancer.gov/ncictrp.
9. C. Laine, S. Goodman, M. Griswold, H. Sox, *Ann. Intern. Med*. **146**, 450 (2007).
10. D. Ghersi *et al.*, *Bull. World Health Org*. **86**, 492 (2008).

The IRB Is Key

IN HIS RESPONSE TO A LETTER ON HIS PERspective "*Homo experimentalis* evolves" (*1*) (31 October 2008, p. 672), J. A. List replies that the economic research he conducted did not require informed consent by the unwitting participants because the study yielded interesting results and did no harm. The response concludes, "Those cases in which there are minimal benefits of informed consent but large costs are prime candidates for relaxation of informed consent."

There are certainly grounds for dispensing with informed consent in social sciences research, and Institutional Review Boards (IRBs) have guidelines for doing so. However, there are no general rules that cover every instance, and the investigator does not have the authority to make the decision about when the guidelines apply. Instead, exemptions are granted by IRBs only on a case-by-case basis. In this case, the appropriate question is, did List submit his research protocol to his supervising IRB and request an exemption? The decision to dispense with informed consent is not the author's.

WILLIAM R. LOVALLO

Department of Psychiatry and Behavioral Sciences, University of Oklahoma Health Sciences Center, VA Medical Center, Oklahoma City, OK 73104, USA. E-mail: bill@mindbody1.org

Reference

1. J. A. List, *Science* **321**, 207 (2008).

Response

AS MY LETTER SUGGESTS, INSTITUTIONAL REVIEW Boards (IRBs) are a necessary condition and serve an invaluable role; I wrote, "Local

Published by AAAS

Research Ethics Committees and Institutional Review Boards in the United States serve an important role in monitoring such activities." All of my research has IRB approval, and I suspect from W. R. Lovallo's concerns that he and I agree on all aspects of subject approval. In my own research, I have been even more stringent than IRB requirements—I do not deceive subjects and always ensure that they are better off due to my experiment. These conditions are certainly not a constraint in all, or even many, IRBs that I am aware of.

On another note, some researchers do not have Local Research Ethics Committees and Institutional Review Boards. Outside the United States, researchers in the social sciences must rely largely on their own moral principles. I view my letter as also speaking to these scholars and letting them know that strict standards need to be followed.

JOHN A. LIST

Department of Economics, University of Chicago, and National Bureau of Economic Research, Chicago, IL 60637, USA. E-mail: jlist@uchicago.edu

Implications of Ancient Ice

IN THEIR BREVIA "ANCIENT PERMAFROST AND a future, warmer Arctic" (19 September 2008, p. 1648), D. G. Froese *et al.* reported the discovery of an ice wedge more than 740,000 years old within the discontinuous permafrost zone of central Yukon Territory, Canada. Permafrost in this area is warm (with an average temperature greater than –2°C) and strongly controlled by local site conditions. It is generally sparse or absent on south-facing slopes and in areas lacking insulating vegetation cover.

Permafrost is very sensitive to climate change: If the average annual air temperature changes by several degrees, then the temperature of soil and permafrost soon will change by the same amount (*1*–*3*). Therefore, a 2°C warming there would lead to permafrost degradation in the Yukon Territory. By assuming that temperatures during at least two previous interglacials were substantially warmer than those today, Froese *et al*. conclude that the persistence of that ice wedge implies that permafrost can be more stable than previously thought.

Their conclusion may not be valid, for two reasons. First, Froese *et al*. provide no evidence that average annual temperatures in the Yukon Territory were higher during previous interglacials than they are today. Second, there is no physically substantiated explanation of how permafrost could be stable across such climate warming. Permafrost stability during warmer interglacials would seem to require that the permafrost was covered with a massive layer of organic insulator, such as moss or peat, in order to prevent summer heating of soils. Such a condition seems to be contradicted by the facts that such a cover does not exist there today and that there is no evidence that widespread fires might have burned any such cover had it once existed. Consequently, a simpler explanation of why the ice wedge did not melt during the past 740,000 years is that this territory was never substantially warmer then it is today. **SERGEY ZIMOV**

North East Scientific Station, Pacific Institute for Geography, Russian Academy of Sciences, Cherskii, Yakutia Republic, 678830, Russia. E-mail: sazimov55@mail.ru

References

1. V. A. Kudryavtsev, Ed., *Obshee Merzlotovedenie (Geokryologiya)* (Moscow State Univ. Press, Moscow, 1978) (in Russian).
2. D. M. Lawrence, A. G. Slater, *Geophys. Res. Lett*. **32**, L24401 (2005).
3. D. V. Khvorostyanov *et al*., *Tellus Ser. B* **60**, 250 (2008).

Response

WE THANK S. ZIMOV FOR HIS INTEREST IN our Brevia. Zimov does not agree that the persistence of ice wedges in Yukon Territory for more than 700,000 years reflects thermal resilience of permafrost more than a few meters below the surface. He instead interprets our finding as evidence that tempera-

CORRECTIONS AND CLARIFICATIONS

Letters: "Unsung hero Robert C. Gallo" by G. Abbadessa *et al*. (9 January, p. 206). Author Riccardo Dalla-Favera's name was spelled incorrectly. The misspelling has been corrected online.

Random Samples: "Getting there" (2 January, p. 19). The credit should have been Andrew Nelson/Joint Research Center of the European Commission.

News Focus: "Shortfalls in electron production dim hopes for MEG solar cells" by R. F. Service (19 December 2008, p. 1784) contained several inaccuracies. A 2004 paper by Victor Klimov and colleagues reported multiple exciton generation (MEG) in lead selenide nanocrystals, not lead sulfide as reported. The maximum efficiency at the time was 220% for a photon energy equal to 3.8 times the nanocrystals' band gap. This was later increased to 268% at 4.3 times the band gap. (The 700% efficiency reported was produced using photons with 7.8 times the nanocrystal band gap.) More recent results for stirred solutions saw the efficiency decline not to 40% as stated, but to 150% for photons 4.3 times the bandgap, only slightly higher than the 123% seen by Moungi Bawendi's group at MIT for the same ratio of the photon energy to energy gap. The difference in the MEG effect between different nanocrystal samples in the case of stirred solutions was 30%, not 300% as reported.

The sentence "Synthetic differences between samples may have left some with surfaces that enhance the effect" should have stated simply that the surface variation between samples likely accounts for some of the variation of the effect. The Los Alamos team stirred the samples to keep the nanoparticles from absorbing more than one photon from sequential laser pulses, not "more than one photon at a time." Klimov does not consider the new results "bad news," but agrees that the effect is too small to make a practical difference for current solar cells. The MEG effort at the National Renewable Energy Laboratory is led by Arthur Nozik; the group treated solar cells made with nanocrystal films with ethanedithiol, not hydrazine.

An E-Letter with further clarifications has been published online at www.sciencemag.org/cgi/eletters/322/5909/1784a.

TECHNICAL COMMENT ABSTRACTS

Comment on "Human-Specific Gain of Function in a Developmental Enhancer"

Laurent Duret and Nicolas Galtier

Prabhakar *et al*. (Reports, 5 September 2008, p. 1346) argued that the conserved noncoding sequence *HACNS1* has undergone positive selection and contributed to human adaptation. However, the pattern of substitution in *HACNS1* is more consistent with the neutral process of biased gene conversion (BGC). The reported human-specific gain of function is likely due to the accumulation of deleterious mutations driven by BGC, not positive selection.

Full text at www.sciencemag.org/cgi/content/full/323/5915/714c

Response to Comment on "Human-Specific Gain of Function in a Developmental Enhancer"

Shyam Prabhakar, Axel Visel, Jennifer A. Akiyama, Malak Shoukry, Keith D. Lewis, Amy Holt, Ingrid Plajzer-Frick, Harris Morrison, David R. FitzPatrick, Veena Afzal, Len A. Pennacchio, Edward M. Rubin, James P. Noonan

Duret and Galtier argue that human-specific sequence divergence and gain of function in the *HACNS1* enhancer result from deleterious biased gene conversion (BGC) with no contribution from positive selection. We reinforce our previous conclusion by analyzing hypothesized BGC events genomewide and assessing the effect of recombination rates on human-accelerated conserved noncoding sequence ascertainment. We also provide evidence that AT → GC substitution bias can coexist with positive selection.

Full text at www.sciencemag.org/cgi/content/full/323/5915/714d

tures during previous interglacials were no higher in the Yukon than those today.

Zimov states that the temperature of permafrost will increase soon after surface air temperature increases. However, unlike the elements of the cryosphere that lack insulating surface cover, such as sea ice or glaciers, permafrost is buffered from air temperatures by soil properties, snow cover, vegetation, and ground ice. For example, peat, a common ground cover in discontinuous permafrost, promotes rapid winter cooling and slows summer warming, while ice lenses at the top of permafrost retard thaw due to latent heat effects. Field evidence for the resilience of permafrost is demonstrated from the early Holocene warm interval at a site in the discontinuous permafrost zone of central Yukon, where depth of seasonal thaw (active layer) was only 70 cm greater than today (*1*), despite several centuries of warmer-than-present temperatures (*2*).

Zimov presents an alternative interpretation for the preservation of ancient ice, suggesting that middle and late Pleistocene interglaciations in our study region were not substantially warmer than present. However, northward expansion of latitudinal tree-line (*3*), fossil plant and insect assemblages (*4*), and stratigraphic evidence for widespread degradation of shallow permafrost (*5*) indicate that temperatures were warmer by up to several degrees during the last interglaciation.

In light of the physical properties of permafrost, proxy paleoenvironmental indicators from the last interglacial, and other evidence in our Report, we maintain that the persistence of 700,000-year-old ice wedges in the discontinuous permafrost zone reflects the thermal resilience of deep permafrost. We note, however, that our observations of relict ground ice in Yukon Territory are consistent with the thaw of shallow permafrost to a depth of a few meters, which we think occurred during the last interglaciation (*6*). Higher air temperatures certainly will warm and eventually thaw shallow permafrost and associated carbon reservoirs, but permafrost more than a few meters below the surface is likely stable through the 21st century.

DUANE G. FROESE,[1]* ALBERTO V. REYES,[1] JOHN A. WESTGATE,[2] SHARI J. PREECE[2]

[1]Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, AB T6G 2E3, Canada. [2]Department of Geology, University of Toronto, Toronto, ON M5S 3B1, Canada.

*To whom correspondence should be addressed. E-mail: duane.froese@ualberta.ca

References

1. C. R. Burn, F. A. Michel, M. W. Smith, *Can. J. Earth Sci.* **23**, 794 (1986).
2. D. S. Kaufman *et al.*, *Quat. Sci. Rev.* **23**, 529 (2004).
3. M. E. Edwards, T. D. Hamilton, S. A. Elias, N. H. Bigelow, A. P. Krumhardt, *Arct. Antarct. Alp. Res.* **35**, 460 (2003).
4. J. V. Matthews, C. E. Schweger, J. A. Janssens, *Géogr. Phys. Quat.* **44**, 341 (1990).
5. T. L. Péwé, G. W. Berger, J. A. Westgate, P. M. Brown, S. W. Leavitt, *Geol. Soc. Am. Spec. Pap.* **319** (1997).
6. A. V. Reyes, D. G. Froese, B. J. L. Jensen, *Eos* **87**, Fall Meet. Suppl. PP51B-1142 (2006).

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

Comment on "Human-Specific Gain of Function in a Developmental Enhancer"

Laurent Duret[1]* and Nicolas Galtier[2]

Prabhakar *et al.* (Reports, 5 September 2008, p. 1346) argued that the conserved noncoding sequence *HACNS1* has undergone positive selection and contributed to human adaptation. However, the pattern of substitution in *HACNS1* is more consistent with the neutral process of biased gene conversion (BGC). The reported human-specific gain of function is likely due to the accumulation of deleterious mutations driven by BGC, not positive selection.

The characterization of functional noncoding regulatory elements positively selected during human evolution is of major importance for understanding the genetic basis of human-specific adaptations. One possible approach to identify such elements entails searching for genomic regions that are highly conserved across nonhuman vertebrates but strongly divergent, that is, rapidly evolving, in humans (*1*, *2*). These regions, called HACNSs (human-accelerated conserved noncoding sequences) or HARs (human-accelerated regions), are good candidates for being regulatory elements under positive selection.

Prabhakar *et al.* (*3*) reported the detailed analysis of the first of these candidates, the 546–base pair (bp) long *HACNS1*. *HACNS1* has accumulated 16 human-specific changes since the human/chimpanzee divergence, which represents a substitution rate four times as high as would be expected given the average neutral substitution rate in the human genome. Thirteen of the 16 changes are clustered in a small region (81 bp) of *HACNS1*. Using a mouse model, the authors demonstrated that human *HACNS1* could drive the expression of a reporter gene in the mesenchyme of the early-developing forelimb and later-developing hindlimb in embryos. This pattern of expression was very different from the one observed when the chimpanzee or macaque *HACNS1* sequences were assayed. Prabhakar *et al.* further showed through directed mutagenesis that the 13 changes in the 81-bp region are responsible for the difference in enhancer activity between human *HACNS1* and its orthologs in apes. Accelerated sequence evolution is a hallmark of positive selection. The authors therefore concluded that these 13 substitutions have been driven by positive selection. They suggest that these changes may have contributed to the evolution of human-specific digit and limb patterning (*3*).

Positive selection, however, is not the only possible explanation for accelerated sequence evolution. Biased gene conversion (BGC) is a neutral process associated with meiotic recombination, which favors the fixation of AT → GC mutations (*4*). Given that recombination often occurs in hotspots (< 2 kb), BGC can create strong substitution hotspots, thereby leading to spurious signatures of positive selection (*5*–*7*). BGC was identified through its effect on neutral sites (*6*–*9*), but it can also drive the fixation of weakly deleterious mutations in functional elements (*5*, *10*). Noteworthy features of BGC are that its prevalence is particularly high in subtelomeric regions (*7*) and that it is much more strongly associated with the rate of crossovers in male than in female germlines (*6*, *7*, *9*). We have previously shown that among the HARs that were proposed as candidates for positive selection, many show the hallmarks of BGC: There is an excess of HARs in regions of high recombination rate and the pattern of substitution in HARs is strongly biased toward GC (*5*). This is precisely the pattern observed in *HACNS1*. First, this element is located in a subtelomeric region of chromosome 2, where the rate of male crossover is particularly elevated [2.77 cM/Mb, compared with 0.98 cM/Mb on average for autosomes; regions with a male crossover rate higher than 2.77 cM/Mb represent only 7% of the genome (*11*)]. Second, among the 16 substitutions in *HACNS1* there are 14 AT → GC substitutions, 2 GC → CG substitutions, but not a single GC → AT substitution. Functional elements (coding or noncoding) in mammalian genome are not particularly GC-rich, so there is a priori no reason why selection should systematically favor AT → GC over GC → AT mutations. Conversely, this pattern of substitution is exactly the one expected under the BGC model.

Prabhakar *et al.* (*3*) reject neutral hypotheses for three reasons: (i) the substitution rate in *HACNS1* is four times the local neutral rate; (ii) the authors claim that under the neutral BGC hypothesis "one would expect an increase in the overall substitution rate across the entire region of increased AT → GC substitution" but that the human-specific substitution rate is only elevated in a narrow 81-bp region of *HACNS1* and is close to the local average outside this region; and (iii) their experiments demonstrate that these human-specific substitutions have a substantial functional impact.

The first point is not an argument against the BGC model: As with selection, strong BGC episodes can result in substitution hotspots (*5*–*7*, *12*). The second point is a misinterpretation of the cited article by Galtier and Duret (*5*). In this paper, we indicate that a BGC-driven substitution hotspot must be GC-biased. However, the reverse proposal is not true: Weak BGC can lead to an excess of AT → GC substitutions without strongly affecting the substitution rate, as shown theoretically and empirically (*7*). Recombination hotspots vary in strength, are evolutionarily unstable, and tend to move rapidly (*13*). The pattern presented by Prabhakar *et al.* [figure 4 in (*3*)] suggests that the evolution of the 81-bp segment has been driven by a strong episode of BGC due to an intense and/or long-lived recombination hotspot, in a region otherwise affected by weaker recombination hotspots. This is in agreement with current knowledge about the spatiotemporal distribution of recombination in humans (*14*).

Finally, the third point is the most important argument raised by the authors: The strong functional changes associated with the human-specific substitutions in HACNS1 imply that its rapid evolution was driven by adaptation. This logical implication, however, does not necessarily hold; not only advantageous substitutions have a phenotypic effect. We know that BGC can overcome natural selection and drive the fixation of weakly deleterious AT → GC mutation (*5*, *10*). The fact that *HACNS1* is under very strong purifying selective pressure in nonhuman vertebrates indicates that, in most species, mutations in this element have some deleterious effect. Hence, the most likely interpretation of the observed substitution pattern is that the human-specific changes correspond to weakly deleterious mutations, driven to fixation by BGC. In other words, the strong functional shift in human *HACNS1* enhancer activity probably results from the accumulation of numerous weakly deleterious substitutions.

In conclusion, we contend that the substitution pattern in *HACNS1* does not support the hypothesis of positive selection. Although we cannot formally exclude that *HACNS1* somehow contributed to human adaptation, the most parsimonious interpretation is that the evolution of *HACNS1* merely reflects the maladaptive consequences of recombination hotspots—the Achilles' heel of our genome.

[1]Université de Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Évolutive, F-69622, Villeurbanne, France. [2]Université Montpellier 2, CNRS UMR 5554, Institut des Sciences de l'Evolution, Place E. Bataillon, CC64-34095 Montpellier, France.

*To whom correspondence should be addressed. E-mail: duret@biomserv.univ-lyon1.fr

References and Notes

1. K. S. Pollard *et al*., *Nature* **443**, 167 (2006).
2. S. Prabhakar, J. P. Noonan, S. Paabo, E. M. Rubin, *Science* **314**, 786 (2006).
3. S. Prabhakar *et al*., *Science* **321**, 1346 (2008).
4. G. Marais, *Trends Genet.* **19**, 330 (2003).
5. N. Galtier, L. Duret, *Trends Genet.* **23**, 273 (2007).
6. T. R. Dreszer, G. D. Wall, D. Haussler, K. S. Pollard, *Genome Res.* **17**, 1420 (2007).
7. L. Duret, P. F. Arndt, *PLoS Genet.* **4**, e1000071 (2008).
8. J. Meunier, L. Duret, *Mol. Biol. Evol.* **21**, 984 (2004).
9. M. T. Webster, N. G. Smith, L. Hultin-Rosenberg, P. F. Arndt, H. Ellegren, *Mol. Biol. Evol.* **22**, 1468 (2005).
10. N. Galtier, L. Duret, S. Glemin, S. Ranwez, *Trends Genet.* **25**, 1 (2009).
11. A. Kong *et al*., *Nat. Genet.* **31**, 241 (2002).
12. J. I. Montoya-Burgos, P. Boursot, N. Galtier, *Trends Genet.* **19**, 128 (2003).
13. W. Winckler *et al*., *Science* **308**, 107 (2005).
14. S. Myers, L. Bottolo, C. Freeman, G. McVean, P. Donnelly, *Science* **310**, 321 (2005).
15. The authors are supported by the Centre National de la Recherche Scientifique and by the Agence Nationale de la Recherche.

12 September 2008; accepted 12 January 2009
10.1126/science.1165848

Response to Comment on "Human-Specific Gain of Function in a Developmental Enhancer"

Shyam Prabhakar,[1]* Axel Visel,[1] Jennifer A. Akiyama,[1] Malak Shoukry,[1] Keith D. Lewis,[1]† Amy Holt,[1] Ingrid Plajzer-Frick,[1] Harris Morrison,[2] David R. FitzPatrick,[2] Veena Afzal,[1] Len A. Pennacchio,[1,3] Edward M. Rubin,[1,3]‡ James P. Noonan[1]‡§

Duret and Galtier argue that human-specific sequence divergence and gain of function in the *HACNS1* enhancer result from deleterious biased gene conversion (BGC) with no contribution from positive selection. We reinforce our previous conclusion by analyzing hypothesized BGC events genomewide and assessing the effect of recombination rates on human-accelerated conserved noncoding sequence ascertainment. We also provide evidence that AT → GC substitution bias can coexist with positive selection.

Duret and Galtier (*1*) suggest that biased gene conversion (BGC) alone drove the rapid human-specific evolution of human-accelerated conserved noncoding sequence 1 (*HACNS1*) (*2*, *3*) and that positive selection played no role. BGC is a nonadaptive, recombination-driven mechanism hypothesized to increase the rate of AT → GC substitutions in a given interval (*4*, *5*). In support, it has been shown that the AT → GC substitution rate is more strongly correlated with male recombination rate than with GC content (*6*). Clusters [100 to 1000 base pairs (bp)] of human-specific substitutions occurring in regions of elevated male recombination rate show an AT → GC bias, which suggests that BGC may have contributed to some substitution hotspots in the genome (*5*). These correlations are stronger at the 1 Mb scale than at finer scales, likely because fine-scale recombination rates are less stable over time than coarse-scale rates (*5*, *7*, *8*). In addition, human-accelerated regions (HARs) identified in a whole-genome screen of conserved genomic elements have an excess of AT → GC substitutions and are enriched in broad regions of high average recombination (*4*, *9*, *10*).

In light of these findings, Duret and Galtier's observation that some HARs and HACNSs could be due solely to BGC is reasonable (*4*). However, their present attempt to extrapolate this argument to *HACNS1* requires several unsupported assumptions regarding the ascertainment of HACNSs compared with HARs and the known distribution of recombination hotspots in the genome. We examine the authors' claims below.

The robust gain of limb expression in *HACNS1* is not in itself evidence of adaptation. However, the most likely effect of deleterious substitution in the enhancer would be to randomly distort the existing function rather than to strengthen the ancestral expression pattern while introducing a new, robust expression domain. The precise, highly spatially restricted gain of function in *HACNS1* is exactly what one would expect from an adaptive process that altered a specific functional module in the enhancer, as suggested by the pattern of substitutions we observed, while leaving the ancestral expression domains intact. Therefore, we find the adaptive hypothesis qualitatively more parsimonious than the authors' alternative (*1*).

Overrepresentation of HARs in subtelomeric regions, which display high recombination rates, has been cited as evidence that some HARs are due to BGC rather than to positive selection (*4*, *10*). However, in assuming that HACNSs suffer from a similar systematic bias, Duret and Galtier misunderstand how HACNSs were identified. In contrast to HARs, which were ascertained relative to whole-genome averaged substitution rates, HACNSs were defined based on their acceleration relative to the local neutral substitution rate (*2*, *9*). Thus, the method used to identify HACNSs mitigates the local average effect of recombination-driven BGC on substitution rates. Although some HACNSs may be due to BGC, a comparison of HACNS distribution relative to male-specific recombination rates demonstrates that HACNSs as a class show no enrichment in regions of high recombination (Table 1) (*11*). Moreover, the fraction of conserved noncoding sequences within a 10-Mb window centered on *HACNS1* that are HACNSs (0.8%) is lower than the whole-genome average (0.9%). The elevated male-specific recombination rate around *HACNS1* thus has no bearing on whether *HACNS1* is affected by BGC.

Because HACNSs show no correlation with coarse-scale recombination rates, Duret and Galtier's claim that BGC explains human-specific acceleration in *HACNS1* rests on a single point: the excess of AT → GC substitutions in the enhancer. As we discussed in (*3*), this AT → GC excess is not confined to the 546-bp *HACNS1* element; instead, it extends into the nonconserved flanking regions, covering ~5 kb in all (*3*). However, the overall human-specific substitution rate is close to the local (1 Mb) average rate across nearly this entire region, with one critical exception: the sharp spike of 13 substitutions in a highly constrained 81-bp subregion within *HACNS1*.

Duret and Galtier (*1*) propose a nonparsimonious "BGC only" scenario for the spatial profile of substitutions in and around *HACNS1*: a sharp (~80 bp), extremely strong recombination hotspot embedded in a larger (~5 kb) weak hotspot. In support, they claim that such a spatial profile is "in agreement with current knowledge about the spatiotemporal distribution of recombination in humans" and reference Myers *et al.* (*12*). However, this is a misinterpretation, because that study analyzed recombination rates at a low resolution of ~2 kb and therefore cannot provide insight on the likelihood of ~80-bp hotspots within larger "warmspots." Moreover, the authors' proposed BGC-driven cluster of 13 substitutions in 81 bp is implausibly narrow and dense, even when compared with the 200 most extreme putative BGC-driven substitution hotspots in the genome, which average ~370 bp and have an average substitution density equivalent to <5 substitutions in 81 bp (*5*).

We reject the authors' assumption that BGC and adaptation are mutually exclusive. The highly localized spike in the substitution rate in *HACNS1* is more consistent with positive selection than with BGC. However, the AT → GC substitution

Table 1. Lack of HACNS enrichment in regions of high male-specific recombination (*11*).

	All CNSs	HACNSs	Fisher two-sided *P*
Total in genome	110,549	992	
No. in regions with male recomb ≥2.0 cM	13,542 (12.3%)	135 (13.6%)	0.25
No. in regions with male recomb ≥2.5 cM	9,330 (8.44%)	87 (8.77%)	0.73

[1]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [2]MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, UK. [3]United States Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA.

*Present address: Computational and Mathematical Biology, Genome Institute of Singapore 138672, Singapore.
†Present address: Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA.
‡To whom correspondence should be addressed. E-mail: EMRubin@lbl.gov (E.M.R.); james.noonan@yale.edu (J.P.N.)
§Present address: Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA.

bias in the rapidly evolving segment of *HACNS1* and in the ~5-kb flanking sequence where the overall substitution rate is not elevated suggests that BGC may have influenced the evolution of this enhancer. The substitution pattern in *HACNS1* may thus be due to positive selection synergizing with a moderate BGC domain. A recombination hotspot of mild strength, that extended over ~5 kb and elevated the fraction of AT → GC substitutions without increasing the overall substitution rate, would explain the observed substitution pattern in the neutral sequence surrounding *HACNS1*. Within *HACNS1*, positive selection for new GC-rich transcription factor binding sites or loss of AT-rich sites likely amplified a modest AT → GC fixation bias caused by BGC, producing a sharp increase in the substitution rate within an ~80 bp cluster of binding sites that generated the human-specific gain of limb expression.

Evidence that an extreme AT → GC substitution bias can coexist with positive selection comes from a 382-bp putative BGC event (chr18:897,475-897,857 in hg18) identified by Dreszer *et al*. (*5*), which contains 19 substitutions and overlaps exon 2 of *ADCYAP1*. In parallel to *HACNS1*, the human-specific substitutions comprising this BGC event are significantly clustered in functional sequence: The 132-bp exon contains 13/19 substitutions (Fisher's exact test, $P = 0.002$), despite being conserved across terrestrial vertebrate genomes. All 13 substitutions are AT → GC and produce 11 amino acid changes. The human-specific nonsynonymous substitution rate in the exon is well above the synonymous rate (dN/dS = 3.2), a classic signature of adaptive evolution (*13*). It is difficult to imagine how neutral or deleterious BGC could selectively favor nonsynonymous substitutions. We hypothesize that accelerated evolution in *ADCYAP1* and *HACNS1* are due to a common mechanism: synergy between BGC and positive selection producing a cluster of AT → GC substitutions at functional sites.

In summary, we find no basis for the authors' claim that the gain of function in *HACNS1* is solely due to BGC. Our analysis reinforces our initial conclusion, that positive selection likely played a role in *HACNS1* evolution.

References and Notes

1. L. Duret, N. Galtier, *Science* **323**, 714 (2009); www.sciencemag.org/cgi/content/full/323/5915/714c.
2. S. Prabhakar, J. P. Noonan, S. Pääbo, E. M. Rubin, *Science* **314**, 786 (2006).
3. S. Prabhakar *et al*., *Science* **321**, 1346 (2008).
4. N. Galtier, L. Duret, *Trends Genet.* **23**, 273 (2007).
5. T. R. Dreszer, G. D. Wall, D. Haussler, K. S. Pollard, *Genome Res.* **17**, 1420 (2007).
6. L. Duret, P. F. Arndt, *PLoS Genet.* **4**, e1000071 (2008).
7. C. C. Spencer *et al*., *PLoS Genet.* **2**, e148 (2006).
8. W. Winckler *et al*., *Science* **308**, 107 (2005).
9. K. S. Pollard *et al*., *Nature* **443**, 167 (2006).
10. K. S. Pollard *et al*., *PLoS Genet.* **2**, e168 (2006).
11. A. Kong *et al*., *Nat. Genet.* **31**, 241 (2002).
12. S. Myers, L. Bottolo, C. Freeman, G. McVean, P. Donnelly, *Science* **310**, 321 (2005).
13. Z. Yang, *Mol. Biol. Evol.* **24**, 1586 (2007).
14. Research was done under Department of Energy contract DE-AC02-05CH11231, University of California, E. O. Lawrence Berkeley National Laboratory, and supported by National Heart, Lung and Blood Institute grant HL066681 and National Human Genome Research Institute grant HG003988 (L.A.P.); the Agency for Science, Technology, and Research of Singapore (S.P.); an American Heart Association postdoctoral fellowship (A.V.); and NIH National Research Service Award fellowship 1-F32-GM074367 and the Department of Genetics, Yale University School of Medicine (J.P.N.).

28 October 2008; accepted 12 January 2009
10.1126/science.1166571

EVOLUTION

The Overwhelming Evidence

Massimo Pigliucci

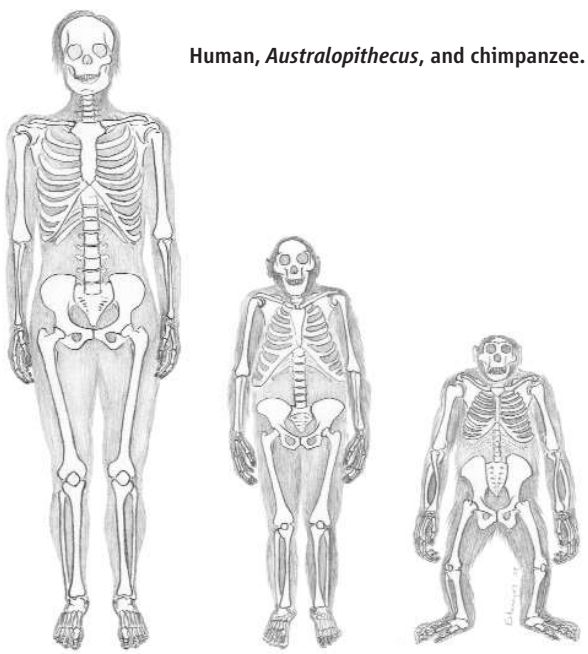
Why Evolution Is True
by Jerry A. Coyne

Viking, New York, 2009. 304 pp. $27.95, C$31. ISBN 9780670020539. Oxford University Press, Oxford. 329 pp. £14.99. ISBN 9780199230846.

It had to be done, and Jerry Coyne is unquestionably one of the most qualified people for the job. I am referring to a clear, engaging, accessible explanation of the evidence for evolution, an aspect of the so-called evolution-creation "controversy" that is too often neglected. There are, of course, plenty of books criticizing creationism and its cousin intelligent design (*1*–*2*) as well as works aiming to explain the creationist phenomenon within the broader context of American anti-intellectualism (*3*). We can also easily find plenty of superb books for the public about various aspects of evolutionary biology [e.g., (*4*–*6*)] even beyond the classical essays by Stephen Gould and Richard Dawkins. And yet, it is hard to get one's hands on a good non–college-level presentation of why evolution is, as they say, both a theory and a fact. Coyne's *Why Evolution Is True* begins to fill this obvious lacuna, even though—just like in other branches of science—additional popular writing by scientists and well-informed journalists on evolution will be welcome for many years to come.

The first eight chapters span pretty much everything one may want to know about evolution but, apparently, so few dare to explain. Coyne (an evolutionary geneticist at the University of Chicago) first introduces readers to a basic definition of evolution. With an expert hand, he then leads them through discussions of the paleontological evidence, developmental biology and vestigial organs (and molecules), biogeography, natural selection ("the engine of evolution"), sexual selection, species concepts, and speciation processes. The latter two entries should actually be read in the singular, because Coyne subscribes to one species concept (based on reproductive isolation) and one chief mode of speciation, allopatry. He and I disagree on this and on other aspects of current evolutionary theory, but this is not the place to entertain technical arguments at the cutting edge of the field. Still, readers of Coyne's book will get a fairly conservative version of evolutionary theory, with occasional hints about the many heated discussions that characterize any live science and that eventually fuel its progress toward a better understanding of the natural world.

Perhaps the best part of the book is the one that addresses the real issue: human evolution. It has been noted many times that if evolutionary biologists limited themselves to the nonhuman world, very few people would pound their fists on the pulpits of fundamentalist churches to decry "the evil doctrine." But when science strikes at our own cherished self-image, then the trouble begins. Just ask Copernicus and Galileo. Coyne is clear and convincing here, at least for those sufficiently open-minded to read the book seriously. He makes a good use of narrative and images, and one of the latter truly is worth a thousand or more words: Early on in the chapter on human evolution, Coyne simply puts side by side the complete skeletons of a modern human, an *Australopithecus afarensis*, and a chimpanzee. The message is not, obviously, that *A. afarensis* is a "missing link" between us and the chimps, because the latter two species are contemporary and modern whereas *A. afarensis* is ancient. But it takes a particularly obtuse mind to look at the figure and reject the notions that *A. afarensis* is a member of the human lineage and that we and chimps have evolved from a common ancestor. Then again, there is no dearth of obtuse minds when it comes to creationism.



Human, *Australopithecus*, and chimpanzee.

The problem with the creation-evolution issue, however, is that it is not about the evidence. The clash is not a scientific debate, it is a social controversy. Coyne understands this, and he begins his last chapter by recounting the story of a public lecture he gave about evolution and intelligent design. Afterward he was approached by someone in the audience who frankly stated: "I found your evidence for evolution very convincing—but I still don't believe it." What is a scientist to do? Coyne admits that the issue goes far beyond science, into philosophy and questions of meaning and morality. Which is why philosophers have been very helpful in this arena during the past several years. It is a matter of explaining to the public not just the power but the limits of science. Coyne is critical, for instance, of much evolutionary psychology and the facile just-so stories that have abounded of late to "explain" all sorts of human behaviors, from rape to depression. I'm with him on this. The point is not that aspects of human behavior did not evolve by natural selection, but rather that the usually high standards of behavioral genetics are simply not met by most, though not all, the evolutionary psychology literature.

Still, creationists have had a problem with evolution ever since Darwin and will continue to challenge its teaching in public schools out of the same paranoia about moral decay that gripped Victorian discussions of *The Origin of Species*. What we need is a cultural change, which notoriously takes a long time and is not just a matter of presenting the evidence for the rational position and walking away while patting oneself on the back for a job well done. Nonetheless, we must present the evidence, and Jerry Coyne's book does an excellent job of it.

The reviewer is at the Departments of Ecology and Evolution and of Philosophy, Stony Brook University, Stony Brook, NY 11794, USA. E-mail: pigliucci@genotypebyenvironment.org

References

1. B. Forest, P. R. Gross, *Creationism's Trojan Horse: The Wedge of Intelligent Design* (Oxford Univ. Press, New York, 2004); reviewed by S. Olson, *Science* **304**, 825 (2004).
2. M. Young, T. Edis, Eds. *Why Intelligent Design Fails: A Scientific Critique of the New Creationism* (Rutgers Univ. Press, New Brunswick, NJ, 2004).
3. M. Pigliucci, *Denying Evolution: Creationism, Scientism, and the Nature of Science* (Sinauer, Sunderland, MA, 2002).

Published by AAAS

4. J. Weiner, *The Beak of the Finch: A Story of Evolution in Our Time* (Knopf, New York, 1994).
5. M. W. Kirschner, J. C. Gerhart, *The Plausibility of Life: Resolving Darwin's Dilemma* (Yale Univ. Press, New Haven, CT, 2005); reviewed by B. Charlesworth, *Science* **310**, 1619 (2005).
6. N. Shubin, *Your Inner Fish: A Journey into the 3.5-Billion-Year History of the Human Body* (Pantheon, New York, 2008).

10.1126/science.1168718

EVOLUTION

On the Metazoan Tree

Wallace Arthur

Unlike many books, *Perspectives in Animal Phylogeny and Evolution* has a remarkably accurate title. Alessandro Minelli (an evolutionary biologist at the University of Padova, Italy) presents perspectives rather than "facts," because our view of the structure of the animal kingdom is still changing, and he discusses both that structure (phylogeny) and the mechanisms of evolution through which it has come about.

The book deals separately with evolutionary pattern and process, in that order. The very different flavors of the two sets of chapters suggest a parallel between the organization of the book and the organization of animals that Minelli calls "the double animal." That refers to the quasi-independence of parts derived from ectoderm or mesoderm and from endoderm (e.g., the many cases where the ectodermal or mesodermal structures are segmented while endodermal structures, such as the gut, are not).

Two commendably brief and effective introductory chapters indicate how Minelli sees the relation between evolutionary pattern and process. One theme that emerges here is that the reasons for the evolutionary origin of a structure can rarely be found in the functions of its more elaborate, much later form. For example, feathers did not begin to evolve because of selection for the ability to fly. To put it another way (one that would have appealed to Stephen Jay Gould), exaptation is everywhere.

Minelli makes it clear that he sees phylogeny as a backcloth against which evolutionary processes can be better understood. This "pattern before process" argument is familiar to us from cladistics. But Minelli subsequently goes much further than many cladists into evolutionary processes, and I confess that I find this side of things (both in the book and in general) the more interesting.

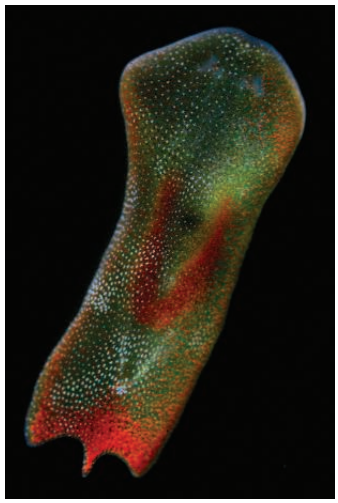
That's not to say that there is nothing of interest in the phylogenetic chapters. At their outset, Minelli reminds us that there is little evidence in the fossil record for animal life before about 550 million years ago (Ma) and nothing at all before the base of the Vendian period (about 650 Ma). This reminder is timely, as it is becoming less and less likely that the 1990s molecular clock estimates of billion-year-old divergences between major animal groups (such as protostomes and deuterostomes) are correct.

Perspectives in Animal Phylogeny and Evolution
by Alessandro Minelli
Oxford University Press, Oxford, 2009.
360 pp. $150, £75.
ISBN 9780198566205.
Paper, $70, £34.95.
ISBN 9780198566212.
Oxford Biology.

With regard to the branching pattern of basal metazoans, Minelli discusses various possibilities, including the recent, unexpected placement of comb jellies (Ctenophora)—rather than the sponges (or some of them)—as the sister group to all other animals. I am skeptical of this suggestion, and, as Minelli emphasizes, the jury is still out on it. Nonetheless, the surprising frequency with which such new possibilities appear in the literature reminds us of how important a cautious, perspectives approach is.

As Minelli pushes on toward what we might want to call (although he doesn't) more advanced animals, the book is readable enough to keep right on going. However, I suspect that most readers will instead use these chapters, especially "A gallery of the major bilaterian clades," for reference. In any event, these chapters lay the foundation for what follows, the three chapters on processes.

These offer lots of interesting ideas, all eloquently explained. I don't accept all of them, but I don't imagine that Minelli would expect me to. There are also instances where I agree with his main point but not with his underlying reason. For example, he states that "we still have no satisfactory and comprehensive theory of development," a situation he largely attributes to "the near-universal adultocentric attitude adopted until now." I accept the second statement—many biologists regard developmental stages as functioning simply "to produce an adult" rather than as organisms having (especially in the case of larvae) their own lives. But I believe the main reason we lack a comprehensive theory of development is that we don't understand the quantitative dynamics of gene-protein interactions; we are still largely restricted to a more qualitative view in which interactions are described in a binary way, as activations or inhibitions.



On the basal Bilateria branch? The acoel *Convolutriloba longifissura*.

Similarly, I like very much Minelli's concept of paramorphism, perhaps best explained in relation to limbs. As he points out, with very few exceptions segmented animals have segmented limbs and unsegmented animals have unsegmented limbs. That suggests the limbs are "axis paramorphs": the genetic machinery used to make the trunk has been reused to produce them. Minelli sees paramorphism as an alternative to gene cooption, whereas I would see it as a version of that. In any event, it is an important concept.

Minelli ends with five "guidelines for future investigations into animal evolution." One is to abandon the adultocentric view and recognize that "the fundamental unit of evolution is represented by the life cycle as an integrated whole," with each stage subject to mutationally introduced and selectively driven changes. This position has already been accepted by many, and it can be seen as a way to unite evo-devo and population genetics.

The book's one-to-five ratio of references to text suggests a wide-ranging scholarship presented in a style of writing sometimes described as "spare"—the opposite of verbose. And indeed that is exactly what Minelli offers. Whether or not one agrees with his various perspectives on phylogeny and evolutionary mechanisms, the book is easily recognized for what it is: a magnificent tour of metazoan relationships, characterized by a cautious, nondogmatic approach to both pattern and process. Minelli knows how much we don't yet know for sure, and he suggests interesting and potentially productive ways forward. This concise account easily can and should be read by all serious students of animal phylogeny and evolution.

10.1126/science.1170743

The reviewer is at the Department of Zoology, National University of Ireland, University Road, Galway, Ireland. E-mail: wallace.arthur@nuigalway.ie

PLANETARY SCIENCE

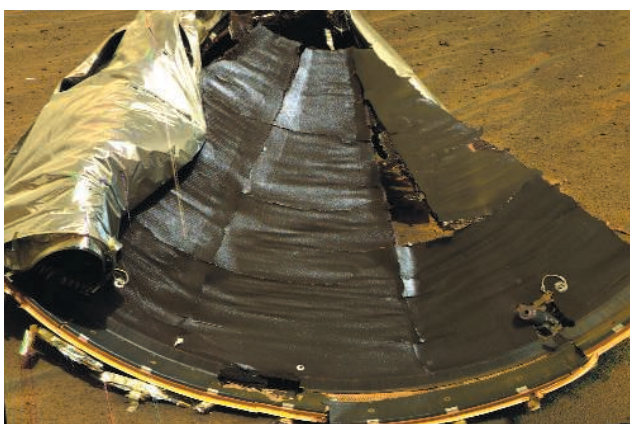
Biologically Reversible Exploration

International policies for protection of Mars and other planets from biological contamination need to be maintained and strengthened.

Christopher P. McKay

The international Committee on Space Research (COSPAR) has established a "planetary protection" policy that involves not contaminating other worlds in a way that would jeopardize the conduct of future scientific investigations. As a signatory to the 1967 Outer Space Treaty, the United States is required by article IX to avoid "harmful contamination" of the other worlds of the Solar System. However, further revisions to the policy are needed.

The two Viking landers that arrived on Mars in 1976 were heat-sterilized to comply with planetary protection. After Viking, COSPAR determined that given the inhospitable conditions on Mars, sterilization was no longer required unless the spacecraft specifically intended to search for life as one of its scientific goals. Thus, the Pathfinder spacecraft and the two Mars Exploration Rovers (Spirit and Opportunity) arrived at Mars with an estimated bioburden exceeding many hundreds of thousands of bacteria (*1*) (see photo).



The heat shield from the Mars Exploration Rover Opportunity photographed in February 2005. Debris on Mars could shield microorganisms from UV light.

In 2003, COSPAR revised its planetary protection policy for Mars from a probabilistic approach to one that was based on the mission objectives and target (*2*). It includes the notion of "special regions," regions in which biocontamination from Earth might grow and thus require special protection. Any spacecraft components that enter such a special region have to be heat-sterilized. The subsurface ice that was the target for the arm on the Phoenix lander is a special region, and the arm was therefore heat-sterilized and placed within a biobarrier bag before launch to Mars in 2007. Pieces of the rest of the Phoenix lander, which were undoubtedly contaminated, are visible on the surface of Mars.

Any hitchhiking organisms exposed to the martian environment are killed in minutes by the ultraviolet radiation from the Sun (*3*). Bacteria shielded inside the spacecraft would not be killed but would remain dormant because of the dry conditions. Over hundreds of thousands of years, they would be killed by galactic cosmic radiation (*4*).

In 2006, when the U.S. National Research Council emphasized the importance of special regions on Mars (*5*), it also recommended that National Aeronautics and Space Administration (NASA) work with COSPAR and other organizations to convene an international workshop that would focus on "ethical implications and the responsibility to explore Mars in a manner that minimizes the harmful impacts of those activities on potential indigenous biospheres." This would include discussion of whether revisions to current planetary protection policies are needed and how to involve the public in discussions about related ethical issues. This was the first official suggestion that consideration should be given to any indigenous life on Mars even if it is microbial (*6*). This workshop is planned for 2009.

What do we do if we find life on Mars? It is possible that martian life is on the same tree of life as Earth life because of the exchange of meteorites between the two planets (*7, 8*). Alternatively, it may be that life on Mars represents a second genesis—an independent origin of life (*9*). Contamination by even one Earth bacterium may be a serious issue of environmental ethics. Furthermore, if we find evidence of a second genesis, then this may open discussions of warming Mars to help that alien life to flourish (*10*, *11*). Scientists and policy-makers who consider this choice will have to deal with any contamination left on Mars by previous explorers, so that it does not flourish instead.

Sterilization of robotic spacecraft, while no longer policy, is at least possible. With human exploration, sterilization is not an option. Nor is it realistic to imagine that a human base could be so carefully engineered that it would release no microorganisms into the environment.

The spacecraft that have landed on Mars have all been surface missions. Contaminants will remain local and static and can be removed without requiring an effort vastly larger than the missions that carried the contamination. Even at the crash sites, debris from Earth extends no more than a few meters into the surface. Reversing the contamination involves recovering the spacecraft parts and exposing any contaminated dirt to the sterilizing ultraviolet (UV) sunlight. However, if, for example, robotic or human explorers drill to investigate a subsurface aquifer, biologically reversible exploration would require rigorous sterilization of any components that go down the drill hole. Similarly, if human explorers establish bases inside caves (*12*), the naturally sterilizing effect of the surface UV would be lost, and contamination would be persistent.

We should not do anything now that would close off options for the future. I propose that COSPAR, in its upcoming discussions, set a policy that all Mars exploration be biologically reversible and that this policy extend to human exploration as well.

Space Sciences, NASA-Ames Research Center, Moffett Field, CA 94035, USA; christopher.mckay@nasa.gov

References

1. J. Barengoltz, in *2005 IEEE Aerospace Conference Proceedings*, Big Sky, MT, 6 to 12 March 2005 (IEEE, Piscataway, NJ, 2005), pp. 253–261.
2. COSPAR, "Report on the 34th COSPAR Assembly," *COSPAR Inform. Bull.* no. 156, 24 (April 2003).
3. A. C. Schuerger, *Icarus*, **165**, 253 (2003).
4. G. Kminek, J. L. Bada, K. Pogliano, J. F. Ward, *Radiat. Res*. **159**, 722 (2003).
5. National Research Council, *Preventing the Forward Contamination of Mars* (National Academies Press, Washington, DC, 2006), pp. 362–372.
6. C. P. McKay, *Planet. Rep.* **21**, 4 (July/August 2001).
7. B. P. Weiss *et al*., *Science* **290**, 791 (2000).
8. N. H. Sleep, K. Zahnle, *J. Geophys. Res*. **103**, 28529 (1998).
9. C. P. McKay, *PLoS Biol* **2**, e302 (2004).
10. C. P. McKay, O. B. Toon, J. F. Kasting, *Nature* **352**, 489 (1991).
11. M. M. Marinova, C. P. McKay, H. Hashimoto, *J. Geophys. Res*. **110**, E03002 (2005).
12. P. J. Boston *et al*., *Gravit. Space Biol. Bull*. **16**, 121 (2003).

10.1126/science.1167987

CREDIT: NASA/JPL/CORNELL

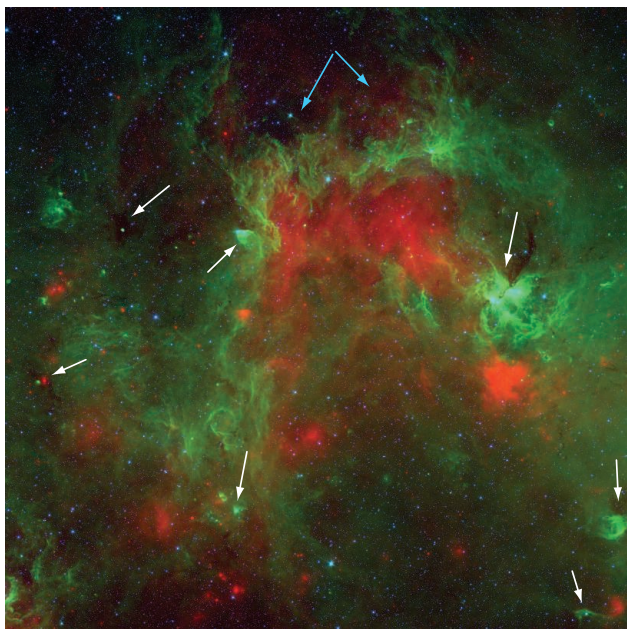
ASTRONOMY

Forming Massive Stars

Sophisticated numerical simulations can now describe the formation of supermassive stars.

Barbara Whitney

The dominant physical processes contributing to the formation of a star are gravity, rotation, and magnetic fields acting on a cloud of dense gas and dust. Magnetic fields and pressure from the gas inhibit the collapse, but once a cloud is dense enough to overcome these forces, collapse begins and gravity and rotation dominate. A starlike core forms at the center and grows as mass is accreted from the infalling gas cloud. This is just classical mechanics, mathematically formulated by Isaac Newton more than 300 years ago. For stars with a mass similar to that of our Sun, a theory for the entire sequence from a dense molecular cloud core to a hydrogen-fusion star like our current Sun has been well developed and tested with observations (*1*). For stars of about 20 to 100 solar masses, however, such a simple consideration renders them unstable, and thus a more sophisticated treatment is required. On page 754 of this issue, Krumholz *et al.* (*2*) introduce a time-dependent model that explains the physical processes involved in the formation of such massive stars.



A variety of stars. A 1.5 by 1.5 degree region on the sky (about 750 by 750 light-years in size) in our Galactic plane. Red emission shows hot dust and gas; green emission shows emission from large molecules at the surface of molecular clouds (the neutral region between the ionized hot gas and cold molecular gas). The blue sources are normal stars like our Sun. The whitish, orange, and red sources are forming stars, most of them more than 10 times as massive as our Sun. The white arrows show regions of massive (and low-mass) star formation. The blue arrows show a region where star formation appears to be inhibited, probably due to strong magnetic fields and hot temperatures caused by colliding flows of gas and dust. The structures form a ringlike shape that may have been formed by a previous generation of massive stars with strong radiation fields and winds, or even a supernova explosion. The star formation (white arrows) may be triggered by the compression in the ring caused by the previous generation of massive stars.

For stars about 20 times as massive as our Sun, the outward pressure force exerted by photons on the grains of dust in the cloud must be considered (*3*). A simple calculation balancing gravity and the radiation force finds that this radiation pressure halts the infall onto the star once it reaches 20 solar masses. Yet, stars with masses as high as 120 times that of our Sun have been observed (*4*). For most of the 20th century, the theoretical models of star formation assumed spherical geometries, which are one-dimensional. Adding rotation to the equations makes the problem two-dimensional and leads to formation of a disk. These two-dimensional numerical models found that stars up to 40 solar masses could form from accretion through the disk, whereas the radiation pressure acts more on the polar regions (*5*). Now, it turns out that newly developed three-dimensional numerical models allow another physical process into the equations, namely instabilities, which then allow even more massive stars to form.

Krumholz *et al.* describe a three-dimensional simulation of the first 50,000 years of the collapse of a 100 solar mass cloud of molecular gas and dust in slow rotation before collapse. The slow rotation at large distances becomes fast rotation as the cloud collapses as a result of conservation of angular momentum, and thus the cloud quickly collapses into a central "proto-star" and disk in about 4000 years. Because of the large pile-up of material onto the disk, two-armed spiral instabilities form after about 20,000 years; this transports angular momentum outward, allowing more material to fall inward onto the star, increasing its mass further. After about 17,000 years, when the central object has reached about 17 solar masses, the outward radiation pressure begins to exceed gravitational force, but only in certain directions, mainly perpendicular to the disk direction, forming radiation bubbles. However, infalling material can still flow around the bubbles to reach the central object. The bubbles develop deformities, and dense fingers of material continue to fall in. These fingers are analogous to the well-studied Rayleigh-Taylor instabilities in classical fluid mechanics. The net effect is that more mass falls in than is pushed out by radiation pressure. The rate at which mass falls in is variable. In their simulation, the disk instabilities cause a second smaller star to form in the disk. At the end of the simulation, the masses of the stars were 33 and 47 solar masses, with the other 20 solar masses still in the disk and infalling envelope, and their future uncertain.

In addition to overcoming previous theoretical problems for forming massive stars, the simulation of Krumholz *et al.* reproduces an observed feature of massive stars, and that is the very common occurrence of binaries. The accretion variability predicted by the simulation caused by the instabilities should be testable by searching for variation in the observed fluxes of massive protostars.

Understanding the formation of massive stars is a fundamental goal in astronomy. Massive stars form the heaviest elements from nuclear burning in their cores and in supernovae explosions after fuel in the core is exhausted. Their winds and luminosity have a profound effect on their surrounding environments, churning up the gas and dust, and both stimulating and regulating new star formation. Examples of the turbulent processes caused by the life cycle of massive star formation (birth and death) can be seen in the figure.

The simulation by Krumholz *et al*. took about 40 days of computing time on 256 processors running simultaneously. This is

Space Science Institute, 4750 Walnut Street, Suite 205, Boulder, CO 80301, USA. E-mail: bwhitney@spacescience.org

CREDIT: SPITZER SPACE TELESCOPE/GALACTIC LEGACY INFRARED MID-PLANE SURVEY EXTRAORDINAIRE

Published by AAAS

hopefully only the beginning of even more sophisticated simulations. Many intriguing open questions remain, such as, how does the central protostar evolve as it starts nuclear burning of hydrogen in its core while its surface continues to accrete from the disk and envelope? (A more compact, hotter star will ionize the surrounding gas and have a different effect on the environment than a larger, cooler star of the same mass and luminosity.) At what point do the stellar winds and radiation disrupt the disk and envelope? Can the disk survive long enough to form planets, or is the realm of planet formation available only to low-mass stars like our Sun?

References

1. F. C. Adams, C. J. Lada, F. H. Shu, *Astrophys. J.* **312**, 788 (1987).
2. M. R. Krumholz *et al.*, *Science*, **323**, 754 (2009); published online 15 January 2009 (10.1126/science.1165857).
3. F. D. Kahn, *Astron. Astrophys.* **37**, 149 (1974).
4. D. F. Figer, *Nature*, **434**, 192 (2005).
5. H. W. Yorke, C. Sonnhalter, *Astrophys. J.* **569**, 846 (2002).

10.1126/science.1169667

EVOLUTION

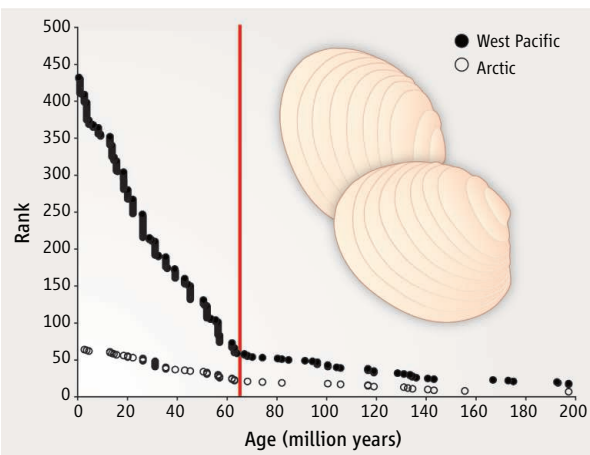
Time's Stamp on Modern Biogeography

A study of living bivalves provides clues to the evolution of biodiversity over the past 200 million years.

J. Alistair Crame

Why are there more types of plants and animals on some parts of the Earth's surface than on others, and do these patterns point to the operation of some basic laws of nature that have yet to be fully understood? Over the past three decades, answers to these questions have shifted from essentially ecological to much more evolutionary (*1*). This shift largely results from the rapid growth in phylogenetic studies, which has led to a proliferation of time-calibrated evolutionary trees, and from the advent of large, analytical databases. Using the latter, it is now possible to assemble vast amounts of spatial and temporal data on the distribution of plants and animals to establish statistically significant biogeographic patterns. On page 767 of this issue, Krug *et al.* use this approach to provide important new insights into how biodiversity evolved over the past 200 million years (*2*).

There are two critical patterns in taxonomic diversity: the trend in the total number of taxa through geologic time, and the latitudinal gradient from the tropics to the poles. These two features are undoubtedly linked. If the number of species (or genera) has increased dramatically through time (especially over the past 65 million years, referred to as the Cenozoic era), then it is almost certain that latitudinal gradients must have steepened too, because the fossil record strongly suggests that many groups originated in tropical and subtropical regions. However, if global diversity has been relatively constant for much of the past 300 million years, then the steep gradients that we see today can be regarded as essentially time-invariant features of Earth's surface. Is the strong diversity contrast between the tropics and the poles really only a feature of the youngest part of the fossil record, or has it perhaps always been there?



Age-frequency distributions of marine bivalve genera. The curves plot the ages of genera, ranked from youngest to oldest, in the tropical West Pacific and Arctic bioprovinces. The vertical line marks the K-Pg boundary at 65 million years ago. The slopes of the two provinces differ slightly before the K-Pg boundary, but diverge strongly following the mass extinction, as many more new species evolved in the tropics than in the Arctic. [From fig. S3 in (*3*)]

According to the conventional view of biodiversity through time, biodiversity rose rapidly from 542 to 445 million years ago, followed by a somewhat irregular plateau until about 250 million years ago. A steep fall at the Permian-Triassic mass extinction event was soon reversed, and the greater part of the past 250 million years was marked by a steep rise that was barely deflected by the Cretaceous–Paleogene (K-Pg) mass extinction. The increase in the number of species through the past 65 million years alone could have been by an order of magnitude (*3*, *4*).

Our faith in this view was, however, severely tested when Alroy *et al.* published a major new database survey of the fossil record of the past 540 million years in 2001 (*5*). The new diversity curve, produced by the authors, seemed to show a much flatter diversity trajectory over the past ~140 million years. Could the canonical explosive radiation of both marine and terrestrial biotas during this time frame be an artifact of the fossil record?

Closer inspection of the Alroy *et al.* (*5*) curves reveals that the younger part of the Cenozoic record (i.e., the last 23 million years) is poorly sampled, and this is particularly so of the tropics (*6*). To counter this, Alroy *et al*. later quadrupled the size of their database and produced a new curve based on substantially better coverage, both stratigraphically and geographically, of the entire Cenozoic (*7*). This curve again shows only a very modest rise over the past 65 million years; in terms of the number of genera, the peak over the last 23 million years is only 1.74 times as high as that of approximately 400 million years ago. Furthermore, Alroy *et al*. (*7*) found evidence of steep latitudinal gradients as far back in time as 475 million years ago. Although these are shallower than those of the last 23 million years, the ratio of tropical to temperate faunas seems to be very similar. These authors suggest that gradients must have developed through time by the parallel development of both tropical and temperate faunas, and not by tropical radiations alone.

Krug *et al.* now provide a somewhat different perspective on the evolution of modern

British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK. E-mail: jacr@bas.ac.uk

biogeographic patterns. Their study, too, is centered on a large database, but in this case it is entirely of living organisms, the marine bivalves. Over 28,000 records of bivalve genera and subgenera from 322 locations around the world have now been compiled by these authors, giving a global record of some 854 genera and subgenera and 5132 species. No fossils are included in the database, but because bivalves have a good fossil record, it is possible to estimate accurately the age of origin of almost all extant genera. It is then possible to plot a backward survivorship curve (*8*) for each of the 27 global bivalve provinces (*9*).

On the basis of these curves, Krug *et al.* find that origination rates of marine bivalves increased significantly almost everywhere immediately after the K-Pg mass extinction event. The highest K-Pg origination rates all occurred in tropical and warm-temperate regions. A distinct pulse of bivalve diversification in the early Cenozoic was concentrated mainly in tropical and subtropical regions (see the figure).

The steepest part of the global backward survivorship curve for bivalves lies between 65 and 50 million years ago, pointing to a major biodiversification event in the Paleogene (65 to 23 million years ago) that is perhaps not yet captured in Alroy *et al.*'s database (*5*, *7*). The jury is still out on what may have caused this event. But we should not lose sight of the fact that the steep rise to prominence of many modern floral and faunal groups in the Cenozoic may bear no simple relationship to climate or any other type of environmental change (*10*, *11*).

References

1. G. G. Mittelbach *et al.*, *Ecol. Lett.* **10**, 315 (2007).
2. A. Z. Krug, D. Jablonski, J. W. Valentine, *Science* **323**, 767 (2009).
3. P. W. Signor, *Annu. Rev. Ecol. Syst.* **21**, 509 (1990).
4. R. K. Bambach, *Geobios* **32**, 131 (1999).
5. J. Alroy *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6261 (2001).
6. A.M. Bush *et al.*, *Paleobiology* **30**, 666 (2004).
7. J. Alroy *et al.*, *Science* **321**, 97 (2008).
8. M. Foote, in *Evolutionary Patterns*, J. B. C. Jackson *et al.*, Eds. (Univ. of Chicago Press, Chicago, IL, 2001), vol. 245, pp. 245–295.
9. M. D. Spalding *et al.*, *Bioscience* **57**, 573 (2007).
10. S. M. Stanley, *Paleobiology* **33**, 1 (2007).
11. M. J. Benton, B. C. Emerson, *Palaeontology* **50**, 23 (2007).

10.1126/science.1169410

SOCIAL SCIENCE

Computational Social Science

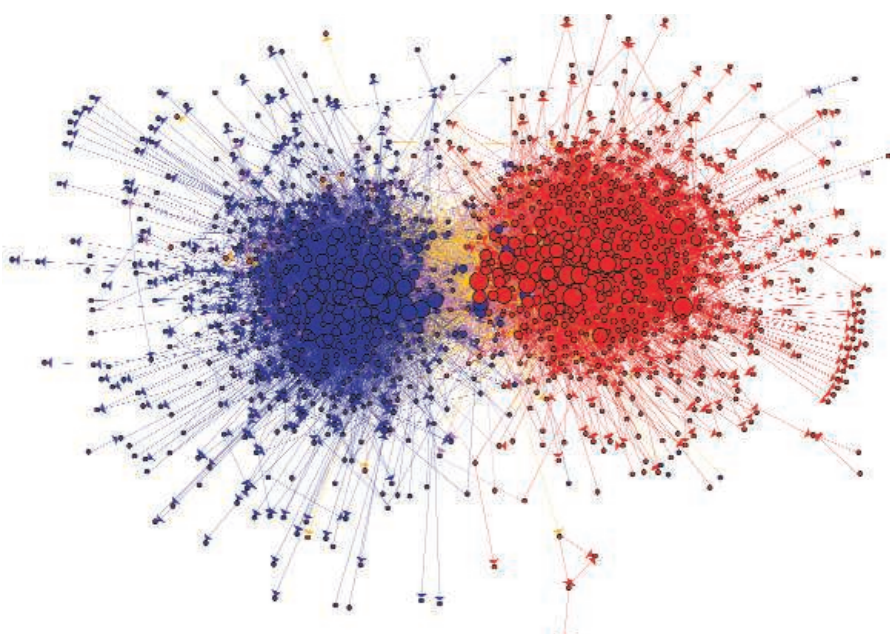
A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Barabási,[5] Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] Myron Gutmann,[3] Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne[2,11]

We live life in the network. We check our e-mails regularly, make mobile phone calls from almost any location, swipe transit cards to use public transportation, and make purchases with credit cards. Our movements in public places may be captured by video cameras, and our medical records stored as digital files. We may post blog entries accessible to anyone, or maintain friendships through online social networks. Each of these transactions leaves digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies.

The capacity to collect and analyze massive amounts of data has transformed such fields as biology and physics. But the emergence of a data-driven "computational social science" has been much slower. Leading journals in economics, sociology, and political science show little evidence of this field. But computational social science is occurring—in Internet companies such as Google and Yahoo, and in government agencies such as the U.S. National Security Agency. Computational social science could become the exclusive domain of private companies and government agencies. Alternatively, there might emerge a privileged set of academic researchers presiding over private data from which they produce papers that cannot be critiqued or replicated. Neither scenario will serve the long-term public interest of accumulating, verifying, and disseminating knowledge.

What value might a computational social science—based in an open academic environment—offer society, by enhancing understanding of individuals and collectives? What are the



Data from the blogosphere. Shown is a link structure within a community of political blogs (from 2004), where red nodes indicate conservative blogs, and blue liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it. [Reproduced from (*8*) with permission from the Association for Computing Machinery]

[1]Harvard University, Cambridge, MA, USA. [2]Massachusetts Institute of Technology, Cambridge, MA, USA. [3]University of Michigan, Ann Arbor, MI, USA. [4]New York University, New York, NY, USA. [5]Northeastern University, Boston, MA, USA. [6]Interdisciplinary Scientific Research, Seattle, WA, USA. [7]Northwestern University, Evanston, IL, USA. [8]University of California–San Diego, La Jolla, CA, USA. [9]Columbia University, New York, NY, USA [10]Cornell University, Ithaca, NY, USA. [11]Boston University, Boston, MA, USA. E-mail: david_lazer@harvard.edu. Complete affiliations are listed in the supporting online material.

obstacles that prevent the emergence of a computational social science?

To date, research on human interactions has relied mainly on one-time, self-reported data on relationships. New technologies, such as video surveillance (*1*), e-mail, and "smart" name badges, offer a moment-by-moment picture of interactions over extended periods of time, providing information about both the structure and content of relationships. For example, group interactions could be examined through e-mail data, and questions about the temporal dynamics of human communications could be addressed: Do work groups reach a stasis with little change, or do they dramatically change over time (*2*)? What interaction patterns predict highly productive groups and individuals? Can the diversity of news and content we receive predict our power or performance (*3*)? Face-to-face group interactions could be assessed over time with "sociometers." Such electronic devices could be worn to capture physical proximity, location, movement, and other facets of individual behavior and collective interactions. The data could raise interesting questions about, for example, patterns of proximity and communication within an organization, and flow patterns associated with high individual and group performance (*4*).

We can also learn what a "macro" social network of society looks like (*5*), and how it evolves over time. Phone companies have records of call patterns among their customers extending over multiple years, and e-Commerce portals such as Google and Yahoo collect instant messaging data on global communication. Do these data paint a comprehensive picture of societal-level communication patterns? In what ways do these interactions affect economic productivity or public health? It is also increasingly easy to track the movements of people (*6*). Mobile phones allow the large-scale tracing of people's movements and physical proximities over time (*7*). Such data may provide useful epidemiological insights: How might a pathogen, such as influenza, driven by physical proximity, spread through a population?

The Internet offers an entirely different channel for understanding what people are saying, and how they are connecting (*8*). Consider, for example, this past political season, tracing the spread of arguments, rumors, or positions about political and other issues in the blogosphere (*9*), as well as the behavior of individuals "surfing" the Internet (*10*), where the concerns of an electorate become visible in the searches they conduct. Virtual worlds, which by their nature capture a complete record of individual behavior, offer ample opportunities for research—experimentation that would otherwise be impossible or unacceptable (*11*). Similarly, social network Web sites offer a unique opportunity to understand the impact of a person's position in the network on everything from their tastes to their moods to their health (*12*), whereas Natural Language Processing offers increased capacity to organize and analyze the vast amounts of text from the Internet and other sources (*13*).

In short, a computational social science is emerging that leverages the capacity to collect and analyze data with an unprecedented breadth and depth and scale. Substantial barriers, however, might limit progress. Existing ways of conceiving human behavior were developed without access to terabytes of data describing minute-by-minute interactions and locations of entire populations of individuals. For example, what does existing sociological network theory, built mostly on a foundation of one-time "snapshot" data, typically with only dozens of people, tell us about massively longitudinal data sets of millions of people, including location, financial transactions, and communications? These vast, emerging data sets on how people interact surely offer qualitatively new perspectives on collective human behavior, but our current paradigms may not be receptive.

There are also enormous institutional obstacles to advancing a computational social science. In terms of approach, the subjects of inquiry in physics and biology present different challenges to observation and intervention. Quarks and cells neither mind when we discover their secrets nor protest if we alter their environments during the discovery process. As for infrastructure, the leap from social science to a computational social science is larger than from biology to a computational biology, largely due to the requirements of distributed monitoring, permission seeking, and encryption. There are fewer resources available in the social sciences, and even the physical (and administrative) distance between social science departments and engineering or computer science departments tends to be greater than for the other sciences.

Perhaps the thorniest challenges exist on the data side, with respect to access and privacy. Much of these data are proprietary (e.g., mobile phone and financial transactional information). The debacle following AOL's public release of "anonymized" search records of many of its customers highlights the potential risk to individuals and corporations in the sharing of personal data by private companies (*14*). Robust models of collaboration and data sharing between industry and academia are needed to facilitate research and safeguard consumer privacy and provide liability protection for corporations. More generally, properly managing privacy issues is essential. As the recent U.S. National Research Council's report on geographical information system data highlights, it is often possible to pull individual profiles out of even carefully anonymized data (*15*). Last year, the U.S. National Institutes of Health and the Wellcome Trust abruptly removed a number of genetic databases from online access (*16*). These databases were seemingly anonymized, simply reporting the aggregate frequency of particular genetic markers. However, research revealed the potential for de-anonymization, based on the statistical power of the sheer quantity of data collected from each individual in the database (*17*).

Because a single dramatic incident involving a breach of privacy could produce rules and statutes that stifle the nascent field of computational social science, a self-regulatory regime of procedures, technologies, and rules is needed that reduces this risk but preserves research potential. As a cornerstone of such a self-regulatory regime, U.S. Institutional Review Boards (IRBs) must increase their technical knowledge to understand the potential for intrusion and individual harm because new possibilities do not fit their current paradigms for harm. Many IRBs would be poorly equipped to evaluate the possibility that complex data could be de-anonymized. Further, it may be necessary for IRBs to oversee the creation of a secure, centralized data infrastructure. Currently, existing data sets are scattered among many groups, with uneven skills and understanding of data security and widely varying protocols. Researchers themselves must develop technologies that protect privacy while preserving data essential for research. These systems, in turn, may prove useful for industry in managing customer privacy and data security (*18*).

Finally, the emergence of a computational social science shares with other nascent interdisciplinary fields (e.g., sustainability science) the need to develop a paradigm for training new scholars. Tenure committees and editorial boards need to understand and reward the effort to publish across disciplines. Initially, computational social science needs to be the work of teams of social and computer scientists. In the long run, the question will be whether academia should nurture computational social scientists, or teams of computationally literate social scientists and socially literate computer scientists. The emergence of cognitive science offers a powerful model for the development of a computational social science. Cognitive science has involved fields ranging from neurobiology to philosophy to computer science. It has attracted the investment of substantial

Published by AAAS

resources to create a common field, and created enormous progress for public good in the last generation. We would argue that a computational social science has a similar potential, and is worthy of similar investments.

References and Notes

1. D. Roy *et al.*, "The Human Speech Project," Proceedings of the 28th Annual Conference of Cognitive Science Society, Vancouver, BC, Canada, 26 to 29 July 2009.
2. J. P. Eckmann *et al. Proc. Natl. Acad. Sci. U.S.A.* **101**, 14333 (2004).
3. S. Aral, M. Van Alstyne, "Network Structure & Information Advantage," Proceedings of the Academy of Management Conference, Philadelphia, PA, 3 to 8 August 2007.
4. A. Pentland, *Honest Signals: How They Shape Our World* (MIT Press, Cambridge, MA, 2008).
5. J.-P. Onnela *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**,7332 (2007).
6. T. Jebara, Y. Song, K. Thadani, "Spectral Clustering and Embedding with Hidden Markov Models," Proceedings of the European Conference on Machine Learning, Philadelphia, PA, 3 to 6 December 2007.
7. M. C. González *et al.*, *Nature* **453**, 779 (2008).
8. D. Watts, *Nature* **445**, 489 (2007).
9. L. Adamic, N. Glance, in Proceedings of the 3rd International Workshop on Link Discovery (LINKDD 2005), pp. 36–43; http://doi.acm.org/10.1145/1134271.1134277.
10. J. Teevan, *ACM Trans. Inform. Syst.* **26**, 1 (2008).
11. W. S. Bainbridge, *Science* **317**, 472 (2007).
12. K. Lewis *et al.*, *Social Networks* **30**, 330 (2008).
13. C. Cardie, J. Wilkerson, *J. Inf. Technol. Polit.* **5**, 1 (2008).
14. M. Barbarao, T. Zeller Jr., "A face is exposed for AOL searcher No. 4417749," *New York Times*, 9 August 2006, p. A1.
15. National Research Council, *Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data*, M. P. Gutmann, P. Stern, Eds. (National Academy Press, Washington, DC, 2007).
16. J. Felch. "DNA databases blocked from the public," *Los Angeles Times*, 29 August 2008, p. A31.
17. N. Homer, S. Szelinger, M. Redman, D. Duggan, W. Tembe, *PLoS Genet.* **4**, e1000167 (2008).
18. M.V.A. has applied for a patent on an algorithm for protecting privacy of communication content.
19. Additional resources in computational social science can be found in the supporting online material.

Supporting Online Material
www.sciencemag.org/cgi/content/full/323/5915/721/DC1

10.1126/science.1167742

CELL BIOLOGY

Moonlighting in Mitochondria

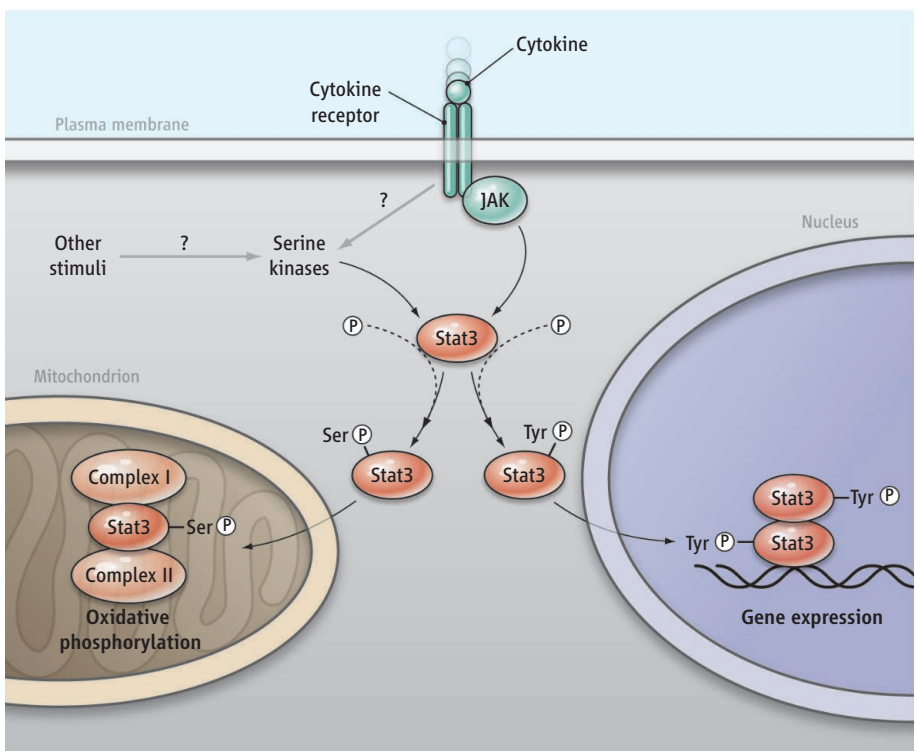
A cellular signaling pathway that responds to cytokines may coordinately control energy production by mitochondria.

Martin G. Myers Jr.

Molecules known as signal transducers and activators of transcription (STATs) regulate gene expression in the nucleus in response to cell surface receptors that are activated by cytokines. On page 793 of this issue, Wegrzyn *et al*. (*1*) reveal that the isoform Stat3 also functions in another organelle—the mitochondria—to control cell respiration and metabolism. This finding not only reveals a new role for Stat3, but implies its potential role in linking cellular signaling pathways to energy production.

Stat3 proteins represent the canonical mediators of signals elicited by type I cytokine receptors at the cell surface (*2*). For instance, the adipocytokine leptin activates Stat3 in hypothalamic neurons to promote the expression of the catabolic neuropeptide, propiomelanocortin, thereby regulating whole-body energy intake and metabolism (*3*). The binding of a cytokine to its receptor triggers an intracellular cascade of events, beginning with the activation of an enzyme, Jak kinase, which is associated with the receptor's cytoplasmic domain. The activated receptor-Jak complex then recruits and phosphorylates a tyrosine residue in cognate STAT proteins. This modification causes the STAT protein to relocate to the nucleus, where, as a dimer, it binds to specific DNA sequences and promotes gene expression (see the figure). Thus, the well-understood job of STAT proteins is to transmit a transcriptional signal from the cell surface to the nucleus. The phosphorylation of some STAT proteins on a specific serine residue may also contribute to their regulation (*2*).

Wegrzyn *et al*. have now identified another crucial role for Stat3, the isoform that responds to cytokines of the interleukin-6 and -10 families (including leptin). These cytokines act in the immune system and many other organ systems to regulate diverse cellular processes, including differentiation, proliferation, and apoptosis (*2*). Noting that GRIM-19, a mitochondrial protein, interacts with Stat3 and inhibits Stat3 transcriptional activity (*4*–*7*), the authors investigated the potential mitochondrial location of Stat3, revealing that a fraction of cellular Stat3 resides within the mitochondria of mouse myocytes and hepatocytes. Here, Stat3 associates with GRIM-19–containing



Dual deployment. The activation of a cytokine receptor at the cell surface promotes the tyrosine phosphorylation (Tyr-P) of Stat3, which dimerizes and moves to the nucleus to control gene expression. Serine phosphorylation (Ser-P) of Stat3 appears to be required for its action in mitochondria, where it promotes increased oxidative phosphorylation. Because many stimuli promote the serine phosphorylation of Stat3, many signaling pathways could regulate mitochondrial respiration via Stat3.

CREDIT: C. BICKEL/*SCIENCE*

Division of Metabolism, Endocrinology and Diabetes, Department of Internal Medicine, and Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, MI 48109, USA. E-mail: mgmyers@umich.edu

complexes I and II, which are components of the electron transport chain that generates energy by oxidative phosphorylation.

Wegrzyn *et al*. determined that in a subset of mouse B lymphocytes devoid of Stat3, oxidative phosphorylation was reduced, a defect attributable to the diminished activity of complexes I and II. That the number of mitochondria and their content of the proteins that constitute complexes I and II were not altered in Stat3-null cells suggests that Stat3 regulates the activity, as opposed to the absolute amount (as would be expected for transcriptional regulation), of complexes I and II.

The expression of Stat3 in these otherwise Stat3-null cells restored oxidative phosphorylation, and this rescue of mitochondrial function did not require the DNA binding domain, the dimerization motif, or the tyrosine phosphorylation site that controls Stat3 nuclear localization and transcriptional activity, consistent with a transcription-independent role for Stat3. By contrast, the conserved serine phosphorylation site on Stat3 was important: Expression of Stat3 with a mutation that prevented phosphorylation of this serine did not produce the rescue effect, whereas a phosphomimetic mutant at this site did. It remains to be determined whether Stat3 is unique among STAT proteins in localizing to and regulating mitochondria.

The nontranscriptional function of Stat3 in mitochondria raises many questions about its precise role in the organelle. For instance, although the results of Wegrzyn *et al*. reveal the absolute requirement for Stat3 to maintain normal mitochondrial function, they do not speak directly to the potential role for Stat3 in the physiologic regulation of cellular respiration. Presumably, specific signaling pathways, such as those that regulate the serine phosphorylation of Stat3, modulate Stat3-dependent control of cellular respiration (see the figure). Whether the control of mitochondrial localization, complex I/II association, or some other step might underlie such regulation remains unclear, however.

Although it is tempting to speculate that cytokines use Stat3 to coordinately regulate transcription and respiration, the inhibition of Stat3-dependent transcription by GRIM-19 suggests that the opposite—Stat3 performs each job at the expense of the other—is just as likely. Also, because many cytokine-independent intracellular signaling proteins (such as protein kinase C isoforms and mitogen-activated protein kinases) promote the serine phosphorylation of Stat3 (*2*), cytokines may not be the only, or even the major, controllers of Stat3-modulated oxidative phosphorylation.

Although many questions will take substantial research to work out, the newly discovered mitochondrial function for Stat3 has the potential to dramatically expand the way we think about the roles of STAT proteins, as well as how canonical cellular signaling pathways may control cell energetics.

References

1. J. Wegrzyn *et al*., *Science* **323**, 793 (2009); published online 8 January 2009 (10.1126/science.1164551).
2. C. Schindler, D. E. Levy, T. Decker, *J. Biol. Chem*. **282**, 20059 (2007).
3. S. H. Bates *et al*., *Nature* **421**, 856 (2003).
4. H. Lu, X. Cao, *Mol. Biol. Cell* **19**, 1893 (2008).
5. G. Huang *et al*., *Mol. Cell. Biol.* **24**, 8447 (2004).
6. J. Zhang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9342 (2003).
7. C. Lufei *et al*., *EMBO J.* **22**, 1325 (2003).

10.1126/science.1169660

GEOLOGY

Heavy Metals or Punk Rocks?

Meeting the world's demand for metals such as copper and zinc may be aided by recent insights into the formation of ore deposits.

Robert J. Bodnar

At the recent World Copper Congress in Santiago, Chile, Rio Tinto's chief executive for copper, Bret Clayton, reported that copper consumption is expected to double over the next two decades (*1*); demand for other metals is expected to parallel this trend. These projected metal needs cannot be satisfied with known ore bodies. To locate new deposits, minerals exploration programs require robust genetic models for the formation of economic accumulations of metals. On page 764 of this issue, Wilkinson *et al.* (*2*) elucidate one of the least understood aspects of ore formation: the concentration of metals in hydrothermal solutions that deposited the ores.

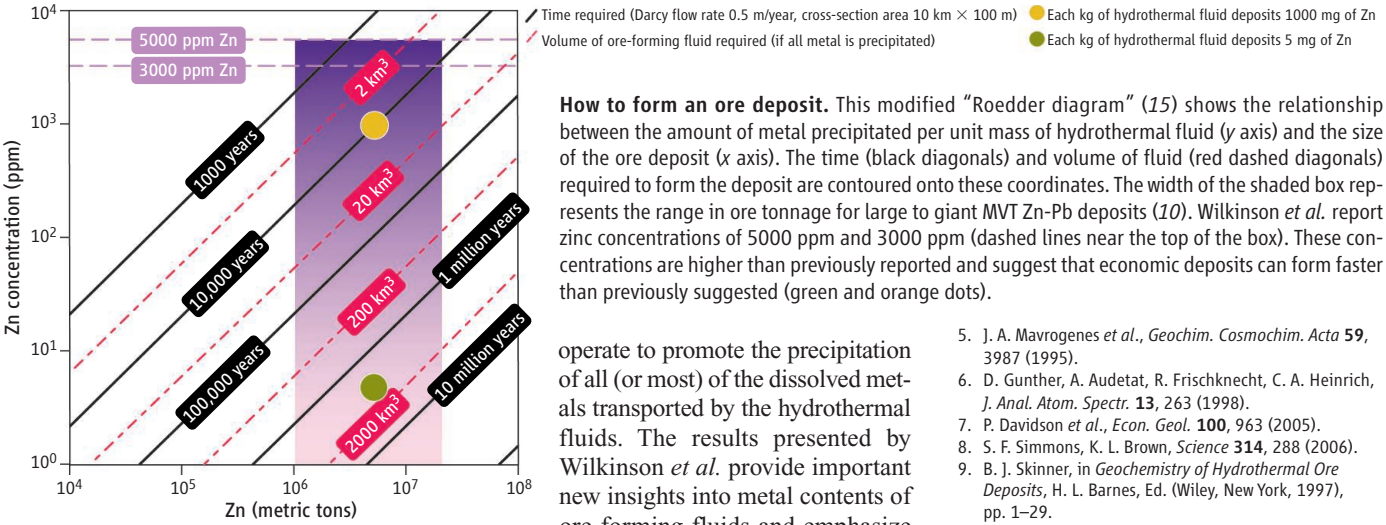
Fluid inclusions provide the best available tool for determining the physical and chemical conditions during ore formation (*3*). These microscopic samples (~5 to 50 μm in diameter) of the ore-forming fluid are trapped in ore and non-ore minerals during mineralization; they thus record the temperature, pressure, and composition of the ore-forming fluid. Advanced microanalytical techniques allow individual fluid inclusions to be analyzed to determine ore metal concentrations (*4*–*6*).

Wilkinson *et al.* now report unusually high metal contents of hydrothermal fluids from two ore districts containing sediment-hosted zinc-lead deposits, based on microanalysis of fluid inclusions. Other recent studies of fluid inclusions from copper (*7*) and gold (*8*) deposits also found much higher metal concentrations than would have been predicted on the basis of experimental data or theoretical models. If confirmed by further studies, these results have important implications for both the duration of the ore-forming process and the amounts of ore fluids needed to generate world-class ore deposits. This information is of crucial importance for understanding ore genesis, given that the duration of ore-forming systems is one of the major unknowns related to the formation of mineral deposits (*9*).

Average continental crust contains about 70 ppm zinc (Zn) and 12.5 ppm lead (Pb). In contrast, average ore grades in Mississippi Valley–type (MVT) Zn-Pb deposits similar to those studied by Wilkinson *et al.* typically are about 6% Zn and 2% Pb by weight, representing enrichment factors of about 850 and 1600, respectively. Thus, metals must be scavenged from a large volume of rock with average crustal metal values and concentrated into a much smaller rock volume to generate economic deposits. For example, large to giant MVT deposits contain on the order of 10^6 to 2×10^7 metric tons combined Zn + Pb (*10*), with an average Zn:Pb ratio of ~3.

Garven (*11*) modeled the fluid-flow history associated with formation of a large MVT Zn-Pb deposit in Pine Point, Canada, and concluded that the total hydrothermal fluid discharge through the mineralized area was 5×10^6 m^3 year^{-1}. Garven assumed that 5 mg of zinc precipitated per kilogram of solution that flowed through these rocks and concluded that it would take from 0.5 to 5.0 million years to form the deposits, with Darcy flow rates (*12*) of 1 to 5 m year^{-1}. Similar durations for ore formation (0.3 million years) have been estimated for the MVT deposits in

Department of Geosciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA. E-mail: rjb@vt.edu



How to form an ore deposit. This modified "Roedder diagram" (*15*) shows the relationship between the amount of metal precipitated per unit mass of hydrothermal fluid (*y* axis) and the size of the ore deposit (*x* axis). The time (black diagonals) and volume of fluid (red dashed diagonals) required to form the deposit are contoured onto these coordinates. The width of the shaded box represents the range in ore tonnage for large to giant MVT Zn-Pb deposits (*10*). Wilkinson *et al.* report zinc concentrations of 5000 ppm and 3000 ppm (dashed lines near the top of the box). These concentrations are higher than previously reported and suggest that economic deposits can form faster than previously suggested (green and orange dots).

the Upper Mississippi Valley district of the United States (*13*).

Flow rates and duration of the ore-forming process reported by Garven (*11*) require total hydrothermal fluid volumes ranging from 2500 to 25,000 km^3 over the lifetime of the ore-forming system. Similar volumes of fluid would be required to form other large to giant MVT deposits if each kilogram of fluid only precipitates a few milligrams of metal. However, if the metal content of the ore-forming fluid is considerably higher, as suggested by Wilkinson *et al.*, then both the amount of fluid required and the duration of the ore-forming event would be reduced by orders of magnitude (see the figure). For example, if each kilogram of hydrothermal fluid deposited 10^3 mg of Zn (orange dot in the figure), then Pine Point and similar deposits could have formed in about 10^4 years from a few cubic kilometers of hydrothermal fluid, compared to the millions of years and hundreds of cubic kilometers of fluid required assuming that each kilogram of hydrothermal fluid deposited 5 mg of Zn (green dot in the figure).

The results presented by Wilkinson *et al.* further highlight the importance of depositional processes in the formation of economic occurrences of metals. Most ore geologists now agree that fluids with metal contents sufficient to produce economic mineralization are relatively common (*14*), and that it is the lack of a suitable depositional mechanism that often limits ore formation. Temperature decrease alone cannot be the dominant mechanism, because the solubility of most metals in most hydrothermal fluids decreases by only a small amount over the temperature range determined for most deposits. Thus, other processes—such as boiling or immiscibility, fluid mixing, or fluid-rock interactions—must operate to promote the precipitation of all (or most) of the dissolved metals transported by the hydrothermal fluids. The results presented by Wilkinson *et al.* provide important new insights into metal contents of ore-forming fluids and emphasize the need for continued research to constrain the amounts of hydrothermal fluids required to form world-class ore deposits and the duration of the ore-forming events.

References and Notes

1. A pdf version of Bret Clayton's presentation at the World Copper Congress on 9 April 2008 is available at www.riotinto.com/media/speeches_7683.asp.
2. J. J. Wilkinson, B. Stoffell, C. C. Wilkinson, T. E. Jeffries, M. S. Appold, *Science* **323**, 764 (2009).
3. E. Roedder, Ed., *Fluid Inclusions* (Mineralogical Society of America, Washington, DC, 1984), vol. 12.
4. A. J. Anderson *et al.*, *Econ. Geol.* **84**, 924 (1989).
5. J. A. Mavrogenes *et al.*, *Geochim. Cosmochim. Acta* **59**, 3987 (1995).
6. D. Gunther, A. Audetat, R. Frischknecht, C. A. Heinrich, *J. Anal. Atom. Spectr.* **13**, 263 (1998).
7. P. Davidson *et al.*, *Econ. Geol.* **100**, 963 (2005).
8. S. F. Simmons, K. L. Brown, *Science* **314**, 288 (2006).
9. B. J. Skinner, in *Geochemistry of Hydrothermal Ore Deposits*, H. L. Barnes, Ed. (Wiley, New York, 1997), pp. 1–29.
10. D. L. Leach *et al.*, in *Economic Geology 100th Anniversary Volume*, J. W. Hedenquist, J. F. H. Thompson, R. J. Goldfarb, J. P. Richards, Eds. (Society of Economic Geologists, Littleton, CO, 2005), pp. 561–607.
11. G. Garven, *Econ. Geol.* **80**, 307 (1985).
12. The Darcy flow rate (also known as the specific discharge or Darcy velocity) describes how long it takes for a given volume of fluid to travel through a given cross-sectional area composed of rock plus pore space.
13. A. P. Gize, H. L. Barnes, *Econ. Geol.* **82**, 457 (1987).
14. B. W. D. Yardley, *Econ. Geol.* **100**, 613 (2005).
15. E. Roedder, in *Reports from the 21st International Geological Congress* (Copenhagen, Denmark, 1960), part 16, pp. 218–229.

10.1126/science.1166394

MATERIALS SCIENCE

Confined Polymers Crystallize

Piet J. Lemstra

Squeezing very thin polymer layers can cause them to form polymer single crystals that could make plastic films less permeable to gases.

Plastics have been very successful in replacing glass, metals, and wood, in part because they are light and easy to process into complex shapes at high speed and at low cost. However, in applications such as packaging, molded plastics can be at a disadvantage compared with steel, aluminum, and glass because of their relatively high permeability to atmospheric gases such as O_2 and CO_2. This problem arises because the synthetic polymers that are the main component of plastics are rather randomly organized in the solid state with sufficient spaces between the molecules that allow for gas diffusion. Although the problem can be solved to some extent by adding less-permeable materials to plastics, ideally it would be desirable to find a way to arrange the long-chain polymer molecules in an orderly way, namely, into crystallites in which the molecules are closely packed. On page 757 of this issue, Wang *et al.* (*1*) report that very thin layers of a commonly used polymer crystallize through special processing conditions into so-called polymer single crystals, which is surprising given the known difficulties in getting polymers to form crystals.

The focus of most polymer research is on functional properties in emerging areas such as biomedical engineering (*2*), electronics (*3*), and energy [for example, plastic solar cells (*4*)], but the vast bulk of synthetic polymers are used in plastics. Packaging materials

Eindhoven University of Technology, Den Dolech 2, 5600 MB Eindhoven, Netherlands. E-mail: p.j.lemstra@tue.nl

Published by AAAS

consume about 40% of the plastics produced. Plastics have surpassed steel in terms of production volume, with about 250 million tons of plastics produced annually worldwide.

Despite this widespread use, there are packaging applications where the gas permeablility of plastics, especially to oxygen, hampers their use, from beer bottles to coatings in advanced electronics. Plastics are quite permeable because their softening temperatures (the glass transition temperature, T_g) are low in comparison with those of silicate glass and much lower than the melt temperatures of metals. For commonly used plastics, T_g ranges from −100° to about 100°C, whereas for silicate glasses, T_g ranges from 500° to above 1000°C.

Gases or liquids permeate a plastic or glass by first dissolving in it and then diffusing through it. In the simplest theoretical description, the permeability is the product of the gas's solubility and its diffusivity. At room temperature, the diffusivity—the movement of the penetrating molecules—in silicate glass is almost zero because the silicate chains are "frozen" at ambient temperatures far below their T_g. The packing of these chains is not orderly but is still very tight and provides very little empty space and mobility for the penetrating molecules to pass through.

In contrast, the segments making up the polymer chains are quite mobile at room temperature. This movement opens up pathways for low–molar mass molecules such as O_2 and CO_2 and, in the case of polar polymers such as nylon, for water. In general, the higher the T_g of a plastic, the better its barrier properties. For that reason, the polyester poly(ethylene terephthalate) (PET), with a T_g of ~ 80°C, can be used for carbonated sodas, but it is too permeable to oxygen for storing beer.

Various options exist to improve the barrier properties of plastic systems, such as applying a metal or ceramic coating or processing alternating layers of the plastic film with deposited ceramic film. More recently, nanoscale clay particles (made from exfoliated single layers of clay) have been dispersed in a plastic film (*5*). These thin clay layers are impermeable to gases and create a tortuous path for diffusing molecules; the path length for diffusion increases and slows gas exchange.

Another approach, but one that is much more challenging, is to enhance the barrier properties of polymers by getting them to crystallize because polymer crystals are impermeable to gases. However, the long chains of synthetic polymer molecules are entangled with each other, much like cooked spaghetti. How can these long-chain and highly entangled molecules form ordered crystallites?

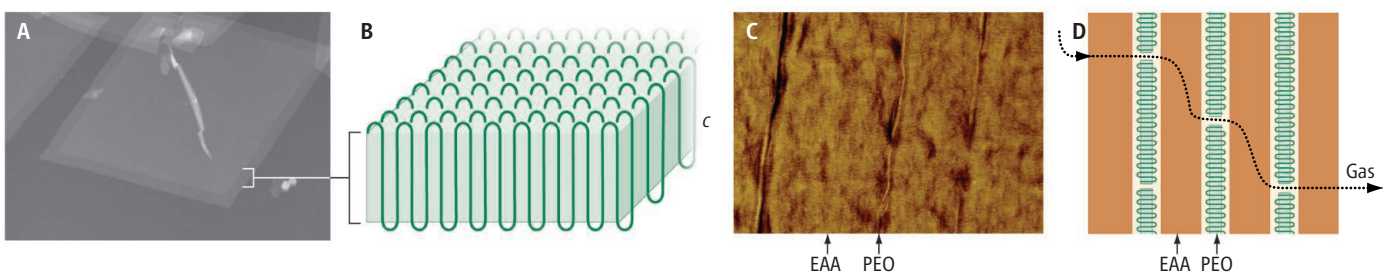
The first efforts at polymer crystallization avoided highly entangled melts and started with dilute solutions. In the 1950s, Keller (*6*), Fischer (*7*), and Till (*8*) found independently that linear polyethylene (PE) can form platelet single crystals upon cooling dilute PE solutions (see the figure, panel A). The long-chain molecules are folded in these crystals (*9*). The fold length, which corresponds to the crystal thickness, is very small, on the order of 10 to 20 nm (see the figure, panel B). The concept of folded-chain crystallization continues to raise enormous interest among polymer physicists, because it is found that all polymer-chain molecules with a regular chemical structure form these folded-chain lamellar crystals.

However, well-defined folded-chain crystals, referred to as lamellae, can only be grown from dilute solutions, which is too slow and inefficient for manufacturing. In processed products such as films, containers, and fibers, the polymers that are cooled from the molten state are only partly crystalline. Although there is still a driving force for the long-chain molecules to form folded-chain crystals, this process is hampered because the chain molecules are highly entangled polymers.

Crystallization from the polymer melt starts from nuclei (catalyst residues or purposely added nucleation agents), and chain-folded crystallites emanate from such nuclei. However, the entangled polymer chains only partly incorporate into the crystallites, creating spherical crystal aggregates called spherulites. Thus, melt-crystallized polymers are semicrystalline because those chain segments that are not folded into crystallites remain amorphous. Partial crystallization improves the barrier properties to some extent, because the crystallites are not permeable and create a tortuous path for small molecules, just as in the case of the nanoclay platelets.

Wang *et al*. now show how poly(ethylene oxide) (PEO)—a crystallizable polymer—can be processed from the melt to form very thin layers consisting of polymer single crystals. The authors use a process in which alternating layers of PEO and EAA poly(ethylene-*co*-acrylic acid) are coextruded. The pressures applied in this process force the long-chain molecules of PEO into their most compact arrangement, that of large folded-chain lamellar crystals (see the figure, panels C and D), while the EAA layers remain amorphous. The permeability of PEO to oxygen decreased by about two orders of magnitude in this crystalline form.

Showing that large lamellar crystals can be grown from the melt is not only of high academic interest. The multilayer extrusion technique is used to create higher-value products through processing [for example, flexible mirrors (*10*)]. Crystalline polymers created in this way may in future function as barrier layers for semicrystalline polymers used in packaging.



Packing polymer chains. Polymer single crystals can be obtained from very dilute solutions, e.g. polyethylene single crystals as visualized by AFM (**A**) in which the chains are folded (**B**) (*9*). Wang *et al.* have now succeeded in creating crystallized polymer layers (**C**) from the melt by extruding alternating thin layers of two different polymers. (**D**)These single-crystal layers inhibit gas diffusion by creating barriers that make the path that molecules take more tortuous.

References

1. H. Wang *et al*., *Science* **323**, 757 (2009).
2. S. Wang, L. Lu, M. Yaszemski, *Biomacromolecules* **7**, 1976 (2006).
3. G. P. Crawford, Ed., *Flat Panel Display Technology* (Wiley, New York, 2005).
4. R. A. J. Janssen, J. C. Hummelen N. S. Sariciftci, *MRS Bull.* **30**, 33 (2005).
5. H. Fischer, *Mater. Sci. Eng.* **C23**, 763 (2003).
6. A. Keller, *Philos. Mag*. **2**, 1171 (1957).
7. E. W. Fischer, *Z. Naturforsch*. **12**, 753 (1957).
8. P. H. Till, *J. Polymer Sci.* **24**, 301 (1957).
9. M. Tian, thesis, Eindhoven Univ. of Technology (2004).
10. M. F. Weber, C. A. Stover, L. R. Gilbert, T. J. Nevitt, A. J. Oudekirk, *Science* **287**, 2451 (2000).

10.1126/science.1168242

CREDIT: J. LOOS/TU-EINDHOVEN

INTRODUCTION

Happy Birthday, Mr. Darwin

THE DIVERSIFICATION AND SPECIATION OF LIVING ORGANISMS ARE THE BROAD theme for this special section, continuing our celebration of Charles Darwin's 200th birthday. The five Reviews in this section present multiple views on research on diversification at scales ranging from the macroevolutionary to the molecular. Benton (p. 728) examines the extents to which biotic and abiotic factors have shaped species diversity in the fossil record. Gavrilets and Losos (p. 732) use theoretical predictions and empirical data to identify general patterns in the temporal, spatial, and genetic/morphological properties of adaptive radiation. Schluter (p. 737) reviews how research on ecological speciation has shifted in focus from morphological evolution to reproductive isolation, tracing the links between Darwin's ideas and current thinking. Fraser *et al.* (p. 741) discuss the contentious area of microbial species formation. Finally, using examples from studies of genes and mutations involved in evolutionary change, Stern and Orgogozo (p. 746) illustrate how developmental biology and evolutionary theory might combine to reveal new predictive principles of genetic evolution.

The special section is accompanied by evolutionary coverage in all other sections of the magazine. The News section features the second in our series of monthly "origins" essays, on the origins of art and symbolic behavior. The Commentary section includes reviews of new evolutionary books. *Science* Careers carries a feature on researchers in the museum world, who play a vital role in evolutionary research. In Reports, Krug *et al.* (p. 767) reveal the legacy of the end-Cretaceous mass extinction for the subsequent diversification of bivalves [see the accompanying Perspective by Crame (p. 720)]. With a focus on conservation, Carnaval *et al.* (p. 785) model spatially explicit evolutionary processes in endemic tree-frog species in the Brazilian Atlantic Forest (the biodiversity hotspot that inspired Darwin during his South American landfall). Forbes *et al.* (p. 776) reveal how a recent host shift of the fly *Rhagoletis pomonella*, a model for sympatric speciation, has led to incipient speciation in a parasitoid wasp that attacks the fly. Rowland and Emlen (p. 773) show facultative male trimorphism in dung beetles, a hitherto unsuspected level of intraspecific variation. Tang and Presgraves (p. 779) report a cellular and molecular mechanism of hybrid sterility, which has a key role in speciation in *Drosophila*. And in *Science* Express, there is an echo of another of Darwin's central interests: Anderson *et al.* show that a melanism mutation has been selected for in gray wolves, most likely after a hybridization event with domestic dogs.

This is a sample, not a survey. Research on speciation and diversification is itself expanding and diversifying. Evolutionary topics have been covered more frequently in *Science* in the first decade of the 21st century than in any previous one (and an order of magnitude more than in the years of the Modern Synthesis in the mid-20th century). This reflects not only the continuing efforts to understand and document the selective forces leading to speciation, but also how genetic research is homing in on the molecular and cellular mechanisms that enable diversification to occur. Darwin, we hope, would be thrilled.

– ANDREW SUGDEN, CAROLINE ASH, BROOKS HANSON, LAURA ZAHN

Speciation

CONTENTS

Reviews

See also related Editorial on p. 687; News story by Pennisi; Essay by Balter; Book Reviews on pp. 716–717; Perspective on p. 720; Reports on pp. 767, 773, 776, 779, and 785; and Science *Careers,* Science *Express,* Science *Podcast, and other related online material at www.sciencemag.org/darwin/*

Science

CREDITS (TOP TO BOTTOM): A. KONINGS; M. LOSOS

Published by AAAS

REVIEW

The Red Queen and the Court Jester: Species Diversity and the Role of Biotic and Abiotic Factors Through Time

Michael J. Benton

Evolution may be dominated by biotic factors, as in the Red Queen model, or abiotic factors, as in the Court Jester model, or a mixture of both. The two models appear to operate predominantly over different geographic and temporal scales: Competition, predation, and other biotic factors shape ecosystems locally and over short time spans, but extrinsic factors such as climate and oceanographic and tectonic events shape larger-scale patterns regionally and globally, and through thousands and millions of years. Paleobiological studies suggest that species diversity is driven largely by abiotic factors such as climate, landscape, or food supply, and comparative phylogenetic approaches offer new insights into clade dynamics.

There are two ways of viewing evolution, through the spectacles of either the Red Queen or the Court Jester. The Red Queen model (*1*) stems from Darwin, who viewed evolution as primarily a balance of biotic pressures, most notably competition, and it was characterized by the Red Queen's statement to Alice in *Through the Looking-Glass* that "it takes all the running you can do, to keep in the same place." The Court Jester model (*2*) is that evolution, speciation, and extinction rarely happen except in response to unpredictable changes in the physical environment, recalling the capricious behavior of the licensed fool of Medieval times. Neither model was proposed as exclusive, and both Darwin and Van Valen (*1*) allowed for extrinsic influences on evolution in their primarily biotic, Red Queen views.

Species diversity in a Red Queen world depends primarily on intrinsic factors, such as body size, breadth of physiological tolerance, or adaptability to hard times. In a Court Jester world, species diversity depends on fluctuations in climate, landscape, and food supply. In reality, of course, both aspects might prevail in different ways and at different times, what could perhaps be called the multilevel mixed model. Traditionally, biologists have tended to think in a Red Queen, Darwinian, intrinsic, biotic factors way, and geologists in a Court Jester, extrinsic, physical factors way.

Much of the divergence between the Red Queen and Court Jester world views may depend on scale (*2*) (Fig. 1): Biotic interactions drive much of the local-scale success or failure of individuals, populations, and species (Red Queen), but perhaps these processes are overwhelmed by substantial tectonic and climatic processes at time scales above 10^5 years (Court Jester). It is important not to export organism-level processes to regional or global scales, and it is likely that evolution operates in a pluralistic way (*3*).

There are two broad methodologies for studies of species diversity through time, taxic and phylogenetic (*4*). The taxic approach involves treating species, genera, or families as independent entities and counting their occurrences against time and other factors. The phylogenetic approach uses cladograms or molecular trees to disentangle key aspects of clade histories. Clades are monophyletic, including all descendants of an ancestor, whereas taxa may be monophyletic or paraphyletic, excluding some descendants of the ancestor. Comparative macroecological studies add rigor to analyses showing that sister clades may vary in rate of evolution, timing of increases in species richness and morphospace occupation, and distributions of evolutionary novelties across lineages and subclades. Here, I will explore the largest-scale global, taxic investigations, provide an outline of how these and other studies correspond to the predictions of the Red Queen, Court Jester, and multilevel mixed models (Table 1), and outline some phylogenetic studies of the macroevolution of species diversity.

The Global Pattern of Diversification Through Time

A key question about the origin of modern biodiversity is how today's 10 million species arose from a single ultimate species of microbial life 3500 million years ago (Ma) (Fig. 1). Two models for global diversification are termed the saturation/equilibrium model (*5*–*7*) and the expansion model (*8*–*11*). The equilibrium model has prevailed, among marine paleobiologists at least, for a long time, and represents a classic Red Queen viewpoint because it implies primarily biotic controls (density dependence) on global diversity.

There are two versions of the equilibrium model, differing in the time when the global

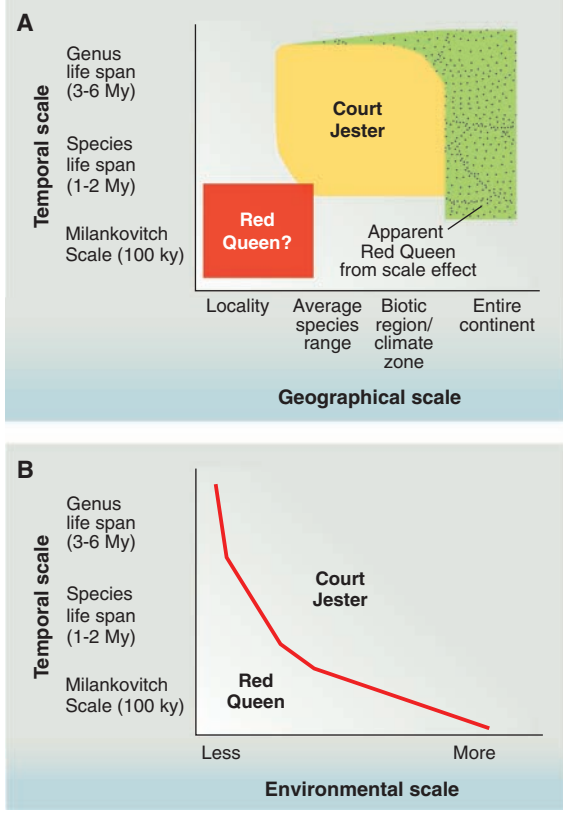


Fig. 1. Operation of Red Queen (biotic causation) and Court Jester (abiotic causation) models at different geographic and temporal scales (**A**). The Red Queen may prevail at organismic and species level on short time scales, whereas the Court Jester holds his own on larger scales. The stippled green shape shows an area where Red Queen effects might be identified erroneously, but these are likely the result of spatial averaging of regional responses to climate change and other complex physical perturbations that may be in opposite directions, and so cancel each other, suggesting no controlling effect of the physical environment on evolution. Physical-environmental disruptions may elicit biotic responses along the red line separating Red Queen and Court Jester outcomes (**B**). The usage here is the microevolutionary Red Queen, as opposed to the macroevolutionary Red Queen that posits constant extinction risk, a view that has been largely rejected (*31*). Illustration based on (*2*).

Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK. E-mail: mike.benton@bristol.ac.uk

marine ecosystem became saturated. Sepkoski's coupled logistic model (*5*) identified three equilibria, in the Cambrian, most of the Paleozoic, and perhaps a third, beginning in the Pliocene and continuing to the present (Fig. 2A). These three equilibrium levels correspond to three sets of phyla, the Cambrian, Paleozoic, and Modern, that interacted and successively replaced each other through the Cambrian-Ordovician and Permo-Triassic intervals, reaching higher equilibrium levels after each long-term replacement event. The second model (*6*, *7*) identifies a single equilibrium level from the early Paleozoic, perhaps 400 Ma, to the present (Fig. 2B). In both models, the equilibria correspond to biodiversity saturation in which new taxa could become established only by driving others to extinction. Key evidence is that both origination and extinction rates appear to have been density-dependent (*5*–*7*), limiting rises in diversity and promoting rapid recovery after extinction events.

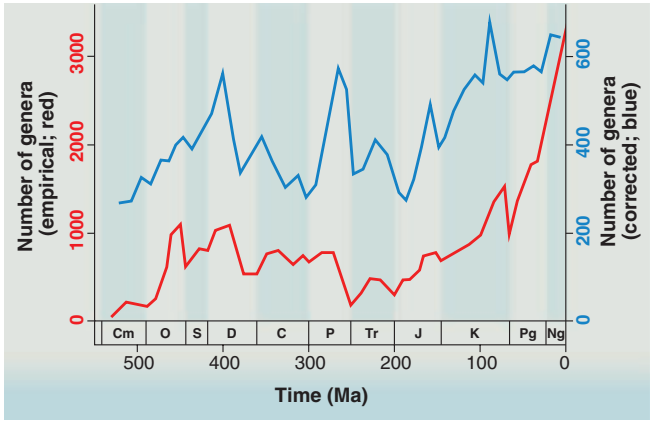


Fig. 2. Patterns of marine animal genus diversification through the past 530 My, the Phanerozoic. The two lines compare current estimates from the empirical (uncorrected) Sepkoski database (red line) and sampling-standardized (corrected) analysis of the Paleobiology Database (blue line). The empirical curve (red line) suggests that global marine diversity reached a possible plateau through the Paleozoic (450 to 250 Ma) and has risen, apparently exponentially, ever since. The sampling-standardized curve (blue line) suggests that global marine diversity reached near-modern levels some 400 Ma and there has been only modest increase since then. Cm, Cambrian; C, Carboniferous; D, Devonian; J, Jurassic; K, Cretaceous; Ng, Neogene; O, Ordovician; P, Permian; Pg, Paleogene; S, Silurian; Tr, Triassic. Based on (*6*).

Alternative models for global diversification are expansionist, allowing global species diversity to rise, with damping, but without a predictable limit (*8*–*11*). Density dependence of origination and extinction rates does not preclude expansionist models because they may be dampened by limiting factors such as shortage of food or space, or active predation, as well as by climate and other physical factors. Further, it seems that the coupled logistic model may be partly an artifact of taxonomic scale (Fig. 2, red curve); it was worked out at ordinal and familial levels but does not work convincingly at generic or specific levels (*10*, *11*), there are problems with key numerical assumptions (*11*, *12*), and the background assumption of a global carrying capacity is doubtful (*8*, *10*, *11*). Further, it has proved hard to export the logistic model to the much more speciose terrestrial realm, whether one considers plants, insects, or vertebrates, because these groups seem to have radiated explosively, without diversity plateaus, particularly in the past 100 million years (My) (*10*).

Resolution between the equilibrium models, and between these and expansionist models, might seem straightforward, but the solution depends on adequate assessment of the quality of the fossil record. The long-term saturation model for global diversification (Fig. 2, blue curve) arises from extensive attempts to correct data sets for sampling error (*6*, *7*), whereas the multiple-equilibria and expansion models originally used raw data, without correction (*5*, *8*–*11*), although more recent analyses return a somewhat dampened but congruent signal when corrections are imposed. Correction for sampling is clearly essential (*6*, *7*), and future investigation must determine appropriate independent proxies for preservation and human error; some current proxies (such as number of fossiliferous localities) are themselves dependent on diversity, and other correction regimes may be so complex as to produce data in which geologic and biologic signals are not obviously separated.

Life on land today may be as much as 25 times as diverse as life in the sea, so it may be wrong to generalize from marine paleontological studies to all life. Perhaps land and sea show similar patterns of exponential increase in species numbers (*8*, *9*, *11*), or perhaps they

Table 1. Macroevolutionary phenomena and their support for either the Red Queen (biotic, intrinsic) or Court Jester (physical, extrinsic) models. Many could fit either worldview, and so are noted as "multilevel mixed."

Red Queen	Court Jester	Multilevel mixed
Interspecific competition	Waxing and waning of clades in association with tectonic and oceanographic events (*2*, *17*)	Vicariance and dispersal in major phylogenetic splits (*17*)
Character displacement	Mass extinctions and smaller extinction events triggered by extrinsic causes such as eruptions, climate change, anoxia, impact (*10*, *11*)	Latitudinal diversity gradient (*22*–*24*)
Evolutionary arms races (*1*)	Coordinated turnovers, originations, and extinctions in response to physical perturbations– termed "coordinated stasis" or "turnover pulse" hypothesis (*2*, *29*, *30*)	Occupation of new ecospace (*25*)
Constancy of ecological guilds through time (*25*)	Nonconstant probability of extinction (*3*, *11*)	Subdivision of niches/specialization (*10*, *25*)
Incumbency advantage (*3*, *24*)	Lack of evidence for a global carrying capacity and equilibrium levels (*8*, *10*)	Declining global extinction rates through time (*1*, *5*)
	Lack of cohesiveness of the great "evolutionary faunas" (*12*)	Onshore-offshore patterns and disturbance (*3*)
	Species richness–energy relationship (*18*, *19*)	Resource use: stenotopes are more speciose than eurytopes (*29*, *30*)
	Inverse relationship between global temperature and biodiversity (*21*)	
	Lack of clear correlation of species richness with body size or other biotic factors (*16*)	

differ in their key rules (*13*, *14*), with the sea acting as a giant Gaussian petri dish, where species diversity is equilibrial and density-dependent, and the land witnessing continuing (dampened) exponential rise in diversity as ever new sectors of ecospace are conquered (*9*, *14*). Any model for global diversification must encompass the independent evidence for increasing complexity of organisms, increases in the occupation of novel ecospace, explosive evolution within particular clades, and addition of novel clades without the loss of precursors (*9*, *11*, *15*), all of which have happened many times in the past 500 My.

Large-Scale Controls on Species Diversity

Taxic paleobiological studies have provided a great deal of evidence about controls, mainly abiotic, on species diversity. Biotic factors, such as body size, diet, colonizing ability or ecological specialization, appear to have little effect on the diversity of modern organisms, although abundance and r-selected life-history characteristics (short gestation period, large litter size, and short interbirth intervals) sometimes correlate with high species richness (*16*).

Geographic and tectonic history has generated patterns of species diversity through time. The slow dance of the continents as Pangaea broke up during the past 200 My has affected modern distribution patterns. Unique terrestrial faunas and floras, notably those of Australia and South America, arose because those continents were islands for much of the past 100 My. Further, major geologic events such as the formation of the Isthmus of Panama have permitted the dispersal of terrestrial organisms and have split the distributions of marine organisms. A classic example of vicariance is the fundamental division of placental mammals into three clades, Edentata in South America, Afrotheria in Africa, and Boreoeutheria in the northern hemisphere, presumably triggered by the split of those continents 100 Ma (*17*). Other splits in species trees may relate to dispersal events, or there may be no geographic component at all.

Species richness through time may correlate with energy. The species richness–energy relationship (*18*) posits correlations with evapotranspiration, temperature, or productivity, and studies of terrestrial and marine ecosystems have shown that these factors may explain as much as 90% of current diversity, although relationships between species diversity and productivity change with spatial scale (*19*). Over long time spans, there are strong correlations between plankton morphology and diversity and water temperature: Cooling sea temperatures through the past 70 My, and consequent increasing ocean stratification, drove a major radiation of Foraminifera, associated with increasing body size (*20*). More widely, there is close tracking between temperature and biodiversity on the global scale for both marine and terrestrial organisms (*21*), where generic and familial richness were relatively low during warm "greenhouse" phases of Earth history, coinciding with relatively high origination and extinction rates.

A much-studied manifestation of energy and temperature gradients is the latitudinal diversity gradient (LDG), namely the greater diversity of life in the tropics than in temperate or polar regions, both on land and in the sea. There are two explanations (*22*): (i) the time and area hypothesis, that the tropical belt is older and larger than temperate and polar zones, and so tropical clades have had longer to speciate, or (ii) the diversification rate hypothesis, that there are higher rates of speciation and lower rates of extinction in the tropics than elsewhere. There is geological and paleontological evidence for a mixture of both hypotheses (*23*, *24*).

Species diversity may increase by the occupation of new ecospace. The number of occupied guilds, that is, broad ecological groupings of organisms with shared habits, has increased in several steps through time, from 20 in the early Paleozoic to 62 in post-Paleozoic marine faunas (*25*). Further, marine animals have shown several step increases in tiering, the ability to occupy and exploit different levels in the habitat: At times, burrowers have burrowed deeper, and reef-builders have built taller and more complex reefs. Analogous, if even more dramatic, expansions of ecospace have occurred on land, with numerous stepwise additions of new habitats, from the water-margin plants and arthropods of the early Paleozoic to the forests and upland habitats of the later Paleozoic when land animals first burrowed, climbed, and flew, through the introduction of herbivory, giant size, endothermy, and intelligence among vertebrates, and the great blossoming of flowering plants (with associated vast expansions in diversity of plant-eating and social insects and modern vertebrates) during the Cretaceous Terrestrial Revolution 100 Ma (*26*).

The other mode of species increase globally or regionally is by niche subdivision, or increasing specialization. This is hard to document because of the number of other factors that vary between ecosystems through time. However, mean species number in communities (alpha diversity) has increased through time in both marine (*15*, *25*) and terrestrial (*10*) systems, even though niche subdivision may be less important than occupation of new ecospace in increasing biodiversity. Further, morphological complexity may be quantified, and a comparative study of crustaceans shows, for example, that complexity has increased many times in parallel in separate lineages (*27*).

Phylogenetic Studies of Clade Histories

Species are not randomly distributed; they have an evolutionary history, and so occur as twigs on a great phylogenetic tree. Studying species as members of clades is a fruitful approach to understanding the drivers and controls on speciation. Key questions include (i) Do species diversify early in a clade's history? (ii) How do diversity and disparity (variance in characters or morphology) covary? (iii) Do major lineages within a clade follow similar, or different, patterns, and if different, why? (iv) Do evolutionary radiations follow the acquisition of new characters or emptying of ecospace? (v) How do major clades of apparent competitors interact over long spans of geologic time? and (vi) How do sister clades vary in species diversity and why?

For such analyses, the ideal is a complete species tree, a phylogenetic tree that contains all species living and extinct, plotted accurately against geologic time (*4*). Simple to say; hard to achieve. More commonly, incomplete trees have been used, with the risk of error in calculations of evolutionary rates or comparisons of subclades. In paleontology, it has proven much easier to work with higher taxa such as genera or families because species fossil records are less complete than those of higher taxa, and yet it is not clear how higher-level patterns relate to those at species level. Many key questions can be tackled by comparing a real tree to a hypothetical tree that follows an equal-rate Markov (ERM) model, equivalent to tree growth after a random walk, where equal chances of speciation and of extinction are shared by all species (*4*).

Major biotic replacements, where one clade replaces another, have been a focus of debate about the roles of competition and progress in macroevolution, and dinosaurs provide a classic example. The standard view was that dinosaurs originated in the Late Triassic, some 230 Ma, by a process of competition in which they prevailed over their precursors, the crocodile-like crurotarsans and others, because of superior adaptations. A comparative phylogenetic study (*28*) shows, however (Fig. 3), that the Dinosauria expanded in two steps, one after an extinction event 225 Ma that removed dominant herbivores, and the second following the end-Triassic extinction 200 Ma that removed most of the crurotarsans. Dinosauria remained at moderate diversity and low disparity, and at lower disparity than the crurotarsans they supposedly out competed, during the 25 My between the events, suggesting that there was no insistent competition driving other groups to extinction but rather that the dinosaurs occupied new ecospace opportunistically, after it had been vacated.

A further study on Dinosauria explored the subsequent evolution of the clade (*26*). Classic views that the dinosaurs arose with a flourish, and then finally gave way in the Cretaceous to the superior mammals, or that they dwindled to extinction because of "racial senility," had long been abandoned. The dinosaurs seemed to be radiating actively in the Cretaceous, with many new clades appearing through their last 55 My, and especially in their final 15 My. The new study (*26*) shows that most diversification shifts (departures from ERM assumptions) fall in the

first one-third of the history of the clade and that their continuing diversification in the Late Jurassic and Cretaceous was mainly indistinguishable from a random walk. In particular, dinosaurs did not participate in the Cretaceous Terrestrial Revolution, some 130 to 100 Ma, when flowering plants, leaf-eating insects, social insects, squamates, and many other modern groups radiated substantially.

There is no geometric reason that diversification shifts should mainly occur low in a clade's history: Clade shapes vary from bottom-heavy to top-heavy, and diversification shifts may be concentrated low (dinosaurs and bats) or high (insects and ants) in a clade (*26*).

In the future, the identification of diversification shifts across numerous taxa may provide evidence for the relative importance of the Red Queen and Court Jester worldviews. If the majority of diversification shifts are coordinated, and associated with particular climatic, tectonic, and geographic drivers, then the Court Jester model of macroevolution would prevail. This would link most increases in species diversity to particular large-scale radiation events, such as the Cretaceous Terrestrial Revolution (*26*), or recoveries after mass extinctions. If, on the other hand, the majority of diversification shifts are unique to particular clades, and not coordinated temporally with others, then the Red Queen worldview might be considered.

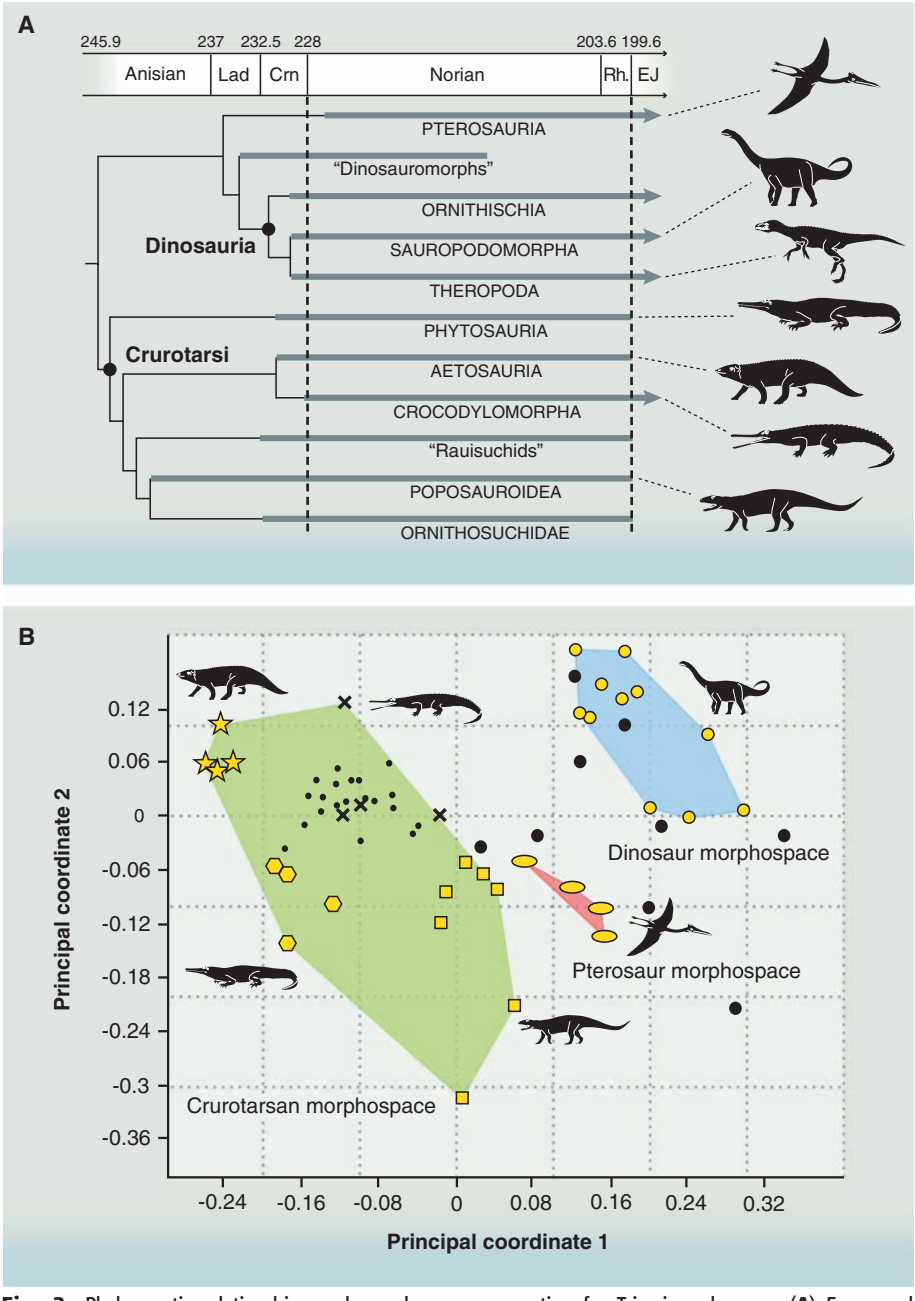


Fig. 3. Phylogenetic relationships and morphospace occupation for Triassic archosaurs. (**A**) Framework phylogeny for Triassic crurotarsans scaled to the Triassic time scale. Numbers at top refer to millions of years before the present; gray bars represent the observed durations of major lineages; vertical dashed lines denote two extinction events, at the Carnian-Norian and Triassic-Jurassic boundaries; arrowheads indicate lineages that survived the latter event. Lad, Ladinian; Crn, Carnian; Rh, Rhaetian; EJ, Early Jurassic. (**B**) Empirical morphospace for Late Triassic archosaurs, based on the first two principal coordinates. Large circles, dinosaurs; ovals, pterosaurs; squares, poposauroids; hexagons, phytosaurs; stars, aetosaurs; crosses, crocodylomorphs; smaller black dots, "rauisuchids"; larger black dots, nondinosaurian dinosauromorphs, *Scleromochlus*. Based on (*28*).

Comparing Sister Taxa

A powerful element of the comparative phylogenetic approach to species diversity through time is the opportunity to compare sister taxa. Sisters arose from a single ancestor, and so their trajectories occupy the same amount of time, and they started with the same genotype and phenotype. Any similarities in their subsequent evolution probably reflect this phylogenetic signal of a common origin, but differences reflect independent aspects of their separate histories.

Comparisons of sister taxa have allowed tests of the resource-use hypothesis (*29*), that generalists are less speciose and have longer species durations than specialists. Specialists divide the physical environment into small patches, each occupied by a species, and each probably more subject to environmental crises than their generalist relatives. Classic examples in support of the resource-use hypothesis come from studies of Neogene mammals (*29*). For example, two antelope subgroups, the tribes Alcelaphini and Aepycerotini, diverged 6 to 8 Ma. The former is now highly speciose, with some 7 living and 25 extinct species, and the latter is represented by two species, only one, the impala *Aepyceros*, surviving. The slowly evolving Aepycerotini consists of few species at any time, and each of those is long lived, whereas the speciose Alcelaphini consists of many short-lived species. The ecological habits of both clades differ: The impala has a broad, generalist diet, whereas the speciose alcelaphines show more dietary specialization. In wider studies of many clades of Neogene African and South American mammals (*30*), the resource-use hypothesis was supported, and some subsidiary predictions confirmed: Specialists are more common than generalists, carnivores include more generalists than herbivores, and there are more specialists in habitats that underwent recent environmental change (tropical rain forests and deserts). The resource-use model then stresses the role of climate and tectonic movements in determining species diversity rather than biological controls such as competition and predation.

Outlook

Paleontologists and evolutionary ecologists have debated species diversity largely independently.

The realization that the Red Queen and Court Jester models may be scale-dependent, and that evolution may be pluralistic (*3*), opens opportunities for dialog. Taxic studies in paleontology continue to have great value in highlighting correlations between species richness and other factors, but comparative phylogenetic methods will illuminate questions about clade dynamics, species richness, and the origin of novelties. Further, methods are shared by paleontologists and neontologists, and this allows direct communication on the patterns and processes of macroevolution. Viewed close up, evolution is all about biotic interactions in ecosystems (Red Queen model), but from further away, the large patterns of biodiversity are driven by the physical environment (Court Jester model).

References and Notes

1. L. M. Van Valen, *Evol. Theory* **1**, 1 (1973).
2. A. D. Barnosky, *J. Vert. Paleont.* **21**, 172 (2001).
3. D. Jablonski, *Evolution* **62**, 715 (2008).
4. A. Purvis, *Annu. Rev. Ecol. Syst.* **39**, 301 (2008).
5. J. J. Sepkoski Jr., *Paleobiology* **10**, 246 (1984).
6. J. Alroy *et al.*, *Science* **321**, 97 (2008).
7. J. Alroy, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11536 (2008).
8. T. D. Walker, J. W. Valentine, *Am. Nat.* **124**, 887 (1984).
9. M. J. Benton, *Science* **268**, 52 (1995).
10. M. J. Benton, B. C. Emerson, *Palaeontology* **50**, 23 (2007).
11. S. M. Stanley, *Paleobiology* **33**, S1 (2007).
12. J. Alroy, *Evol. Ecol. Res.* **6**, 1 (2004).
13. G. J. Eble, *Geobios* **32**, 223 (1999).
14. M. J. Benton, *Geol. J.* **36**, 211 (2001).
15. R. K. Bambach, A. H. Knoll, J. J. Sepkoski Jr., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6854 (2002).
16. N. J. B. Isaac, K. E. Jones, J. L. Gittleman, A. Purvis, *Am. Nat.* **165**, 600 (2005).
17. D. E. Wildman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14395 (2007).
18. G. Hunt, T. M. Cronin, K. Roy, *Ecol. Lett.* **8**, 739 (2005).
19. G. G. Mittelbach *et al.*, *Ecology* **82**, 2381 (2001).
20. D. N. Schmidt, H. R. Thierstein, J. Bollmann, R. Schiebel, *Science* **303**, 207 (2004).
21. P. J. Mayhew, G. B. Jenkins, T. G. Benton, *Proc. R. Soc. London B. Biol. Sci.* **275**, 47 (2008).
22. G. G. Mittelbach, *Ecol. Lett.* **10**, 315 (2007).
23. D. Jablonski, K. Roy, J. W. Valentine, *Science* **314**, 102 (2006).
24. J. W. Valentine, D. Jablonski, A. Z. Krug, K. Roy, *Paleobiology* **34**, 169 (2008).
25. R. K. Bambach, A. M. Bush, D. H. Erwin, *Palaeontology* **50**, 1 (2007).
26. G. T. Lloyd *et al.*, *Proc. R. Soc. London B. Biol. Sci.* **275**, 2483 (2008).
27. S. J. Adamowicz, A. Purvis, M. A. Wills, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4786 (2008).
28. S. L. Brusatte, M. J. Benton, M. Ruta, G. T. Lloyd, *Science* **321**, 1485 (2008).
29. E. S. Vrba, *Am. J. Sci.* **293A**, 418 (1993).
30. A. M. Bofarull *et al.*, *BMC Evol. Biol.* **8**, 97 (2008).
31. S. Finnegan, J. L. Payne, S. C. Wang, *Paleobiology* **34**, 318 (2008).

32. Thanks to S. J. Braddy, P. C. J. Donoghue, D. Jablonski, A. Purvis, M. Ruta, D. N. Schmidt, and anonymous referees for comments and to S. Powell for drafting the figures. Supported by the Natural Environment Research Council and the Royal Society.

10.1126/science.1157719

REVIEW

Adaptive Radiation: Contrasting Theory with Data

Sergey Gavrilets[1]* and Jonathan B. Losos[2]

Biologists have long been fascinated by the exceptionally high diversity displayed by some evolutionary groups. Adaptive radiation in such clades is not only spectacular, but is also an extremely complex process influenced by a variety of ecological, genetic, and developmental factors and strongly dependent on historical contingencies. Using modeling approaches, we identify 10 general patterns concerning the temporal, spatial, and genetic/morphological properties of adaptive radiation. Some of these are strongly supported by empirical work, whereas for others, empirical support is more tentative. In almost all cases, more data are needed. Future progress in our understanding of adaptive radiation will be most successful if theoretical and empirical approaches are integrated, as has happened in other areas of evolutionary biology.

The spectacular diversity of life on Earth that Darwin sought to explain in *On the Origin of Species* emerged through a variety of intricate biological processes. One of these is adaptive radiation, which some consider of foremost importance and potentially responsible for much of the ecological and phenotypic diversity of life (*1*, *2*). "Adaptive radiation" refers to those evolutionary groups that have exhibited an exceptional extent of adaptive diversification into a variety of ecological niches (*2*–*4*), with such divergence often occurring extremely rapidly (*5*). Classic examples of adaptive radiation include Darwin's finches on the Galápagos islands, *Anolis* lizards on Caribbean islands (Fig. 1), Hawaiian silverswords, and cichlids of the East African Great Lakes (Fig. 2), among many others (*1*, *2*, *6*).

Adaptive radiation has two components: the production of new species (speciation) and the adaptation of constituent species to a diversity of ecological niches. Although many classic adaptive radiations are both species rich and adaptively disparate (*7*), this correlation is far from perfect: Some adaptive radiations have relatively low species richness (e.g., Darwin's finches, Australian pygopodid lizards); alternatively, some speciose clades contain little adaptive disparity and thus would not qualify as adaptive radiations (*3*, *8*).

The classic view of adaptive radiation focuses on ecological opportunity, in which an ancestral species finds itself in an environment in which resources are abundant and underutilized. Such resource availability may result from colonization of an underpopulated area (e. g., island or lake), extinction of previously ecologically dominant groups, or evolution of a character—sometimes termed a "key innovation"—that allows the lineage to interact with the environment in novel ways (*1*, *2*). Different evolutionary factors allowing the populations to take advantage of new ecological opportunity have been emphasized, including genetic drift in small founder or peripheral populations (*9*), relaxed (*9*) or strong selection (*2*, *10*), and hybridization (*11*, *12*).

Empirical Approaches for Studying Adaptive Radiation

Four main empirical approaches have been used:

Fossils. Methods based on fossil data allow one to infer the history of the clade through time and to use information from extinct taxa. The disadvantages of this approach are incompleteness of the fossil record, difficulty in assessing the adaptive significance of phenotypic variation among taxa, and the absence of ecological, behavioral, physiological, and other types of data.

Phylogenetic comparative methods. Phylogenetic approaches take advantage of increasingly complete phylogenies for many important groups and have the ability to integrate studies of the evolution of organismal function and ecology. The main disadvantage of these methods is that extinct taxa are often not represented so that there is no way, for example, to detect whether a clade has been more species-rich in the past. Moreover, phylogenetic inferences about character states in the past can be unreliable (*13*, *14*).

Microevolutionary studies of extant taxa. Studies focusing on traits of and processes affecting extant taxa—e.g., phenotypic characters, ecological niches, spatial structure, genetics, local adaptation, competition, and sexual selection—can elucidate much about the processes driving

[1]Departments of Ecology and Evolutionary Biology and Mathematics, National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, TN 37996, USA. [2]Museum of Comparative Zoology and Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: sergey@tiem.utk.edu

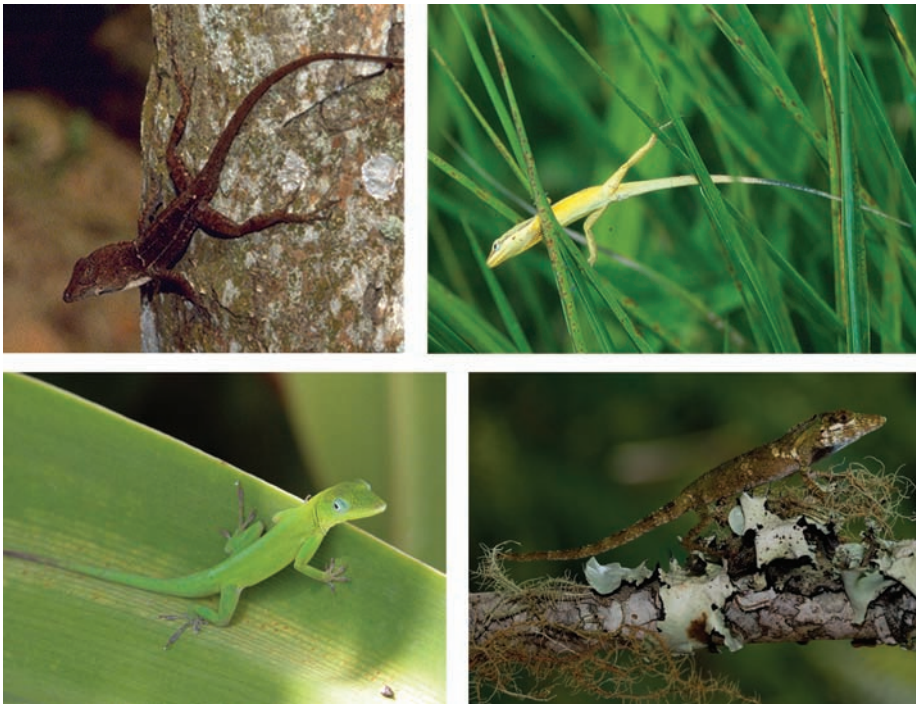


Fig. 1. Adaptive diversification in Caribbean *Anolis* lizards. On each island in the Greater Antilles, anoles have diversified to produce the same set of habitat specialists. Shown here (clockwise from top left) are *A. cybotes*, trunk-ground specialist, Hispaniola (photo: J. Losos); *A. pulchellus*, grass-bush, Puerto Rico (photo: J. Losos); *A. chlorocyanus*, trunk-crown, Hispaniola (photo: M. Losos); *A. insolitus*, twig, Hispaniola (photo: L. Mahler).

adaptive radiation and sometimes even include manipulative experiments. However, extrapolation from processes operating today to what happened early in the history of a radiation is problematic; in the past, different processes may have operated or the outcome of these processes may have been different. These differences might result because conditions may have been different early in a radiation's history, including greater resource abundance, less ecological specialization, and less genetic canalization. Alternatively, one can study taxa that may be in early stages of radiation (e.g., stickleback fishes), but such ecologically and morphologically nondisparate groups are not necessarily good models for the early stages of adaptive radiation.

Adaptive radiation in the laboratory. The advantage of studies of adaptive radiation in microbial microcosms [e.g., (*15*, *16*)] and digital cyberworlds (*17*) is experimental flexibility. However, it is not clear whether these systems are good analogs to adaptive radiations in nature.

Mathematical Modeling

Adaptive radiation can also be studied by using mathematical methods. The latter have traditionally played a major role both in evolutionary biology and in ecology. For example, the modern synthesis of the 1930s and 1940s was a direct result of the development of theoretical population genetics by Fisher, Wright, and Haldane (*18*). A quantitative theory of speciation, which has emerged over the past 40 years [reviewed in (*19*–*21*)], has clarified many questions and topics hotly debated by generations of biologists. The main advantages of mathematical modeling are its generality, the ability to identify crucial parameters, factors, and relevant temporal and spatial scales, as well as to point to the gaps in biological knowledge and intuition. However, the biological realism of models and their underlying assumptions can always be questioned.

Our goal here is threefold. First, we summarize recent theoretical findings on the dynamics of adaptive radiation. Second, we attempt to test theoretical predictions against existing data obtained using a variety of empirical approaches. Third, we identify important areas of theoretical and empirical research where notable advances can soon occur and can contribute dramatically to improving our understanding of adaptive radiation and the generation of biodiversity.

Dynamic Models of Adaptive Radiation—Extensions on Speciation Theory

Adaptive radiation can be viewed as the processes of speciation and adaptation extended to larger spatial and temporal scales. In classifying mechanisms of speciation (and, by extension, adaptive radiation), different approaches are possible. Most commonly, mechanisms of speciation are discussed according to the level of migration between the diverging (sub)populations (*9*, *22*). In this classification, the basic modes of speciation are allopatric, parapatric, and sympatric, corresponding to zero, intermediate, and maximum migration, respectively (*19*–*21*). Alternatively, mechanisms of speciation can be classified according to biological mechanisms driving divergence and speciation, e.g., random drift, ecological selection, or sexual selection (*23*, *24*).

However, sometimes very different biological mechanisms have underlying dynamic commonalities and therefore can be described by very similar mathematical models. Therefore, classifying mechanisms of speciation and adaptive radiation on the basis of similarity of the corresponding models can be more general and insightful than on the basis of particular biological factors.

Five partially overlapping sets of models can be identified. In "spontaneous clusterization" models, an initially random mating population accumulates a substantial amount of genetic variation by mutation, recombination, and random drift and then splits into partially or completely reproductively isolated groups [reviewed in (*20*)]. Spontaneous clusterization models describe the accumulation of Dobzhansky-Muller genetic incompatibilities, speciation by hybridization, divergence in mating preferences, or allochronic speciation (i.e., speciation via divergence in the timing of life-cycle events related to reproduction). In "invasion of empty niches" models [e.g., (*25*, *26*)], a few founders enter a new environment with a number of novel discrete ecological niches. As selection acts on the new genetic variation supplied by mutation, different lineages become adapted to and simultaneously develop genetic preferences for different ecological niches. Ecological and phenotypic diversification is accompanied by the growth in the densities of emerging species (Fig. 3). The process of local adaptation is opposed by deleterious gene flow from subpopulations adapting to alternative niches. Under these conditions, a reinforcement-like process can result in the evolution of premating reproductive isolation between different locally adapted groups. "Selection gradient" models [e.g., (*27*, *28*)] are similar to "invasion of empty niches" models except that environmental conditions vary in a continuous gradient-like fashion and selection for local adaptation acts on a single quantitative trait, the optimum value of which changes linearly across space. In "sympatric diversification" models [e.g., (*29*, *30*)], a spatial component is not present, so that these models describe sympatric speciation. In these models, emerging species specialize on relatively narrow parts of a continuous unimodal distribution of abundances of a particular set of resources. In the fifth set of models, diversification is driven by coevolutionary interactions [e.g., (*31*)].

Ten Patterns of Adaptive Radiation

In mathematical modeling, details do matter and generalizations are often difficult to make. But some general patterns do emerge in models of speciation and adaptive radiation. Here we re-

view and evaluate with empirical data 10 patterns concerning temporal, spatial, and genetic/morphological properties of adaptive radiation. In some cases, mathematical models and generalizations from data have suggested the same patterns independently.

1) Early burst of evolutionary divergence: Typically, there is a burst of speciation and morphological diversification soon after the beginning of the radiation rather than similar rates through time.

Mathematical models of adaptive radiation that explicitly describe microevolutionary processes predict a burst of speciation soon after colonization (*25*, *26*). This happens because the rate of speciation declines as fewer niches remain empty (reduced ecological opportunity) and species become more specialized (increased genetic constraints). In addition, phenomenological null models of clade diversification (*32*, *33*) predict that morphological disparity grows more rapidly than species diversity early in the clade history due to the structure of multidimensional phenotype space and contrasting effects of speciation and extinction on disparity and diversity.

Both paleontological and phylogenetic comparative data indicate that ecologically important morphological variation often arises early in a clade's history. The radiation of placental mammals is a classic example, with most modern orders appearing in the fossil record within a short period (*34*). Many fossil groups experience a peak in morphological disparity relatively early in their history and subsequently decline (*35*). Phylogenetic studies also indicate that major ecological differences often evolve early in clade history, such as tree versus ground Darwin's finches or the "ecomorph" habitat specialists of *Anolis* (*12*, *36*).

An increasingly common theme in molecular studies of speciation is the finding of a burst of species diversification early in a clade's history (*37*, *38*). This pattern is also seen in the fossil records for some groups. This finding is often discussed in the context of adaptive radiation as evidence for the early burst model of diversification. However, implicit in this discussion is the assumption that species diversity is correlated with adaptive diversity, an assumption that probably is often, but not always, correct. Coupling of studies of species diversification and the evolution of adaptive disparity is an important new direction (*39*).

2) Overshooting: An early increase in species diversity is followed by a decline that plateaus (or substantially decelerates).

Overshooting can result if speciation rate is decreasing and/or extinction rate is increasing in time. The rate of speciation can decrease because fewer niches remain empty (reduced ecological opportunity) and species become more specialized (increased genetic constraints). Mathematical models provide only partial support because overshooting is predicted only under some conditions (*25*).

There are several ways to examine these ideas. The most direct is to follow a radiation through history. Some clades do, in fact, exhibit an early peak in species richness, followed by a decline and then leveling-off of diversity (*40*); however, alternative interpretations exist to explain these patterns (*41*).

An alternative approach is to compare related clades of different ages, with the assumption that all else is equal [but see (*42*)]. Taking this approach, overshooting has also been demonstrated in a comparison of the *Tetragnatha* spiders of the Hawaiian islands, in which the relatively young (but not youngest) island of Maui has the most species; the high diversity is attained by having several niches occupied by sets of allopatric species. Gillespie (*43*) hypothesized that in time, the ranges of such species expand and come into contact, at which point competitive exclusion occurs, leading to a decrease in species richness on older islands. Among African Rift lakes, the dependence of cichlid species diversity on time, once corrected for lake surface area, appears to be L-shaped [figure 2d in (*38*)], as expected under overshooting.

3) Stages of radiation: All else being equal, the following sequence of the diversification events is expected: (i) divergence with respect to macrohabitat; (ii) evolution of microhabitat choice and divergence with respect to microhabitat; (iii) divergence with respect to traits that simultaneously control the degree of local adaptation and nonrandom mating; and (iv) divergence with respect to other traits controlling survival and reproduction.

This hypothesis is mostly based on generalizations from the mathematical theory of speci-

Fig. 2. Adaptive diversification in cichlids. Several examples of cichlids that use different habitats. (Upper left) *Gnathochromis permaxillaris* occurs in the deeper (>35 m) intermediate (rock-sand interface) habitat with sediment-rich bottoms in Lake Tanganyika. (Upper right) *Lethrinops furcifer* is found in sandy environments near beaches in Lake Malawi. (Center) *Hemitilapia oxyrhynchus*, which is found in shallow vegetated habitats in Lake Malawi. (Lower left) *Cyprichromis* sp. "Leptosoma Jumbo" (Nkondwe) occurs in open water in Lake Tanganyika. (Lower right) *Petrotilapia nigra* occurs in the upper part (>10 m) of the sediment-free rocky habitat of Lake Malawi. [Photographs by A. Konings]

ation (*20*). Only a limited amount of theoretical work has so far been done with models describing the evolution of various types of traits for a sufficiently long time. But in these cases, simulations did support the above hypothesis (*25*, *26*). However, many factors can interfere with the order in which niche axes are partitioned during species divergence, such as strength of selection, available genetic variation, extent of environmental heterogeneity, or historical contingencies such as the ancestral starting condition of a clade (*2*, *20*).

Stages hypotheses have been proposed for mountain New Guinea birds (divergence in habitat elevation first, followed by segregation in diet, foraging techniques, and size), *Phylloscopus* leaf warblers (order of divergence: body size, foraging morphology and behavior, habitat), cichlids and parrotfish (habitat, diet, sexually selected traits), and *Anolis* lizards (body size, structural microhabitat, microclimate) [reviewed in (*44, 45*)]. However, these hypotheses have been proposed on the basis of phylogenetic analyses that may be incapable of accurately identifying character states during early stages of a radiation for traits that are evolutionarily labile (*45*).

4) Area effects: Speciation driven by ecological factors is promoted by larger geographic areas (e.g., of islands or lakes).

Both mathematical models (*20*, *25*) and verbal theory show that several factors might contribute to increasing speciation rates with area. First, larger areas imply that larger population sizes can be achieved, leading to increased selection efficiency and more advantageous mutations for selection to act upon. Second, the environment may be more heterogeneous in larger areas, thus containing more ecological niches. For example, a metric of niche availability better explains the extent of species diversity in Galápagos snails than does island area (*46*). Third, larger areas may allow for more opportunity for allopatric or parapatric speciation (*25*, *47*). In all three cases in which this hypothesis has been examined (Caribbean *Anolis* lizards, Galápagos snails, and African cichlids), the area effect results in part because a threshold island/lake area exists below which in situ cladogenesis does not occur, even though some islands below the threshold are environmentally quite heterogeneous (*38*, *46*, *48*).

5) Nonallopatric diversification: Speciation during adaptive radiation can occur in the absence of allopatry.

In the traditional view [e.g., (*9*, *49*)], divergent adaptive evolution occurred in allopatric populations that adapted to different ecological niches and thus incidentally evolved reproductive isolation. Subsequent sympatry could lead to reinforcement of incomplete reproductive isolation and enhancement of ecological differences (character displacement).

Recent years have seen a resurgence of interest in nonallopatric modes of speciation driven by divergent natural selection pressures (*19*, *20*, *50*). In line with these theories, some see adaptive radiation as the expected outcome of sympatric or parapatric speciation in the context of abundant ecological opportunity. Mathematical models have identified a number of conditions promoting nonallopatric speciation, including strong disruptive selection, strong nonrandom mating, high levels of genetic variation, close correlation of traits experiencing disruptive selection with those controlling nonrandom mating, and absence of costs of being choosy in mating (*19*, *50*). How common these conditions are in nature is controversial (*20*, *21*).

The strongest evidence for nonallopatric adaptive radiation comes from examples of diversification in areas in which it is hard to envision an allopatric phase. The classic case is the clade of 11 cichlid species found in a small and environmentally homogeneous crater lake in Cameroon (*51*). At the same time, the lack of speciation in some groups on relatively large and environmentally heterogeneous, but isolated, islands suggests that geographic isolation may be needed for speciation to occur, as discussed above.

6) Selection gradient effect: Selection gradients of intermediate slopes promote speciation.

Models of parapatric speciation in which spatially heterogeneous selection for local adaptation acts on an additive trait (*27*, *28*) show that if selection gradients are too shallow, differences in selection experienced in different parts of the species range are not strong enough to drive divergence and speciation. If selection gradients are too steep, selection is so strong that it prevents the population's spread into unoccupied areas with different selection. As a result, intermediate slopes of selection gradients are most conducive for speciation. This prediction is supported by a recent study of divergence in male coloration and female preferences in cichlids (*52*).

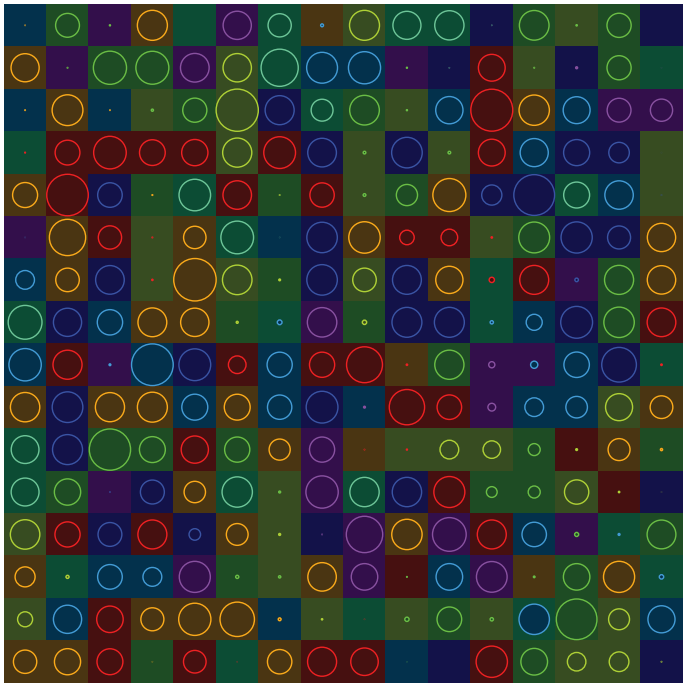


Fig. 3. An illustration of an adaptive radiation in a mathematical model (*25*, *26*). Each square represents a unit spatial area (patch). The color of the square defines the ecological niche assigned to the patch. Each local population is represented by a circle. The radius of the circle is proportional to the population size. The color of the circle defines the niche preferred by most individuals. Matching of the color of the corresponding square and circle (observed in most cases) means that most individuals in the patch prefer the ecological conditions they experience. In the case shown, there are eight different groups of species utilizing eight different ecological niches. Each of these groups is composed of two or three different species differentiated by mating preferences.

7) Spatial dimensionality effect: All else being equal, geographic areas that can approximately be viewed as one-dimensional (such as rivers or shores of lakes and oceans) promote more speciation and tend to maintain higher species richness and phenotypic and genetic diversity per unit area than geographic areas that are two-dimensional (such as lakes, oceans, and continental areas).

This prediction follows both from classical results on patterns of geographic variation in neutral loci induced by isolation by distance [e.g., (*53*)] and from speciation models (*20*). This spatial dimensionality effect may have contributed to the extraordinary divergence of cichlids in the great lakes of Africa, most species of which inhabit the relatively narrow band along the shoreline and have extremely restricted dispersal abilities. However, it does not account for the low diversity of cichlids in African rivers (*54*). In addition, radiation of other lacustrine taxa is not limited to diversification in shoreline habitats (*55*). Moreover, in the oceans, some types of speciation and evolutionary diversification occur more readily in open, two-dimensional habitats than along continental shelves (*56*).

8) Least action effect: Speciation occurring after the initial burst usually involves a minimum phenotypic change.

This pattern is explained by the joint action of genetic constraints and selection for local adaptation: As each lineage gets more specialized for a particular ecological niche, it becomes more and more difficult to establish adaptations for alternative niches (*25*, *26*). The widely observed pattern of phylogenetic niche conservatism, in which closely related species are ecologically and phenotypically similar, is consistent with this pattern (*57*, *58*).

9) Effect of the number of loci: Rapid and extensive diversification is most likely if the number of underlying loci is small.

This is one of the most general patterns emerging from a variety of models (*20*, *25*, *26*). A smaller number of loci implies that their individual contributions to the traits are larger, the effect of selection on each locus is stronger, the number of genetic changes necessary for divergence is smaller, and recombination and migration are less efficient in destroying the coadapted combinations of genes that emerge during speciation. This prediction is supported by data on cichlids, finches, and monkeyflowers, all of which indicate the existence of a small number of loci underlying traits involved in ecological interactions and reproduction (*59*–*61*).

10) Porous genome effect: Species can stably maintain their divergence in a large number of selected loci for very long periods despite substantial hybridization and gene flow that decreases or removes differentiation in neutral markers.

This effect has been observed in many different models (*25*, *26*). A related prediction is that during earlier stages of diversification, when reproductive isolation is still weak, new traits (ecological or involved in mating) that emerge in some lineages can spread to other lineages as a result of hybridization.

Botanists have long been aware that sympatric and ecologically differentiated taxa can maintain their distinctiveness in the face of ongoing hybridization. Recent years, however, have shown that a similar pattern also occurs in many species of animals (*62*). Indeed, recent work has suggested that such interspecific gene flow can be an important source of genetic variation that enables adaptive diversification in early stages of radiation (*11*, *12*).

On the other hand, the homogenizing effect of gene flow has classically been considered an impediment to evolutionary divergence (*9*), a role that it probably plays in many cases (*63*). Moreover, in some cases in which reproductive isolation is based on ecological differences among species, environmental perturbations can negate these isolating mechanisms, leading previously genetically distinct taxa to interbreed and even "de-speciate" (*64*).

General Conclusions

This brief survey suggests the following:

1) More empirical studies are needed specifically aimed at assessing the predictions discussed above. At present, we are left with a haphazard set of studies that happen to be relevant.

2) More generally, evolutionary biology is an inductive science in which we establish generalities by the accumulation of case studies. The number of adaptive radiations that have been extensively studied from the many different perspectives relevant to our discussions is surprisingly small. More detailed studies, integrating across a variety of approaches and disciplines, is needed to build a reservoir of case studies from which generalizations can be drawn.

3) We need studies of general models aiming to capture the most widespread patterns of adaptive radiation. At the same time, we need models that are more closely tailored to the biology of particular taxa. For example, some models (*25*, *26*) assume that mating occurs in the ecological niche that a species exploits; this assumption is true for some taxa, such as some host-specific insect herbivores, but not for other classic adaptive radiations. The extent to which relaxing these assumptions or tailoring them to other biological situations (e.g., different species concepts) would change the predictions needs to be explored. The complexity of the processes of adaptive radiation is reflected in the complexity of corresponding mathematical models. Using the emerging tools of high-performance computing will be crucial for better understanding of models (and nature).

Darwin was confronted with the fruits of adaptive radiation throughout his 5-year journey around the world, and their evolutionary exuberance made an impression on him. Speaking of the Galápagos finches that now bear his name, he wrote in *The Voyage of the Beagle*:

"Seeing this gradation and diversity of structure in one small, intimately related group of birds, one might really fancy that from an original paucity of birds in this archipelago, one species has been taken and modified for different ends."

In the 150 years since publication of the *Origin*, adaptive radiations have continued to astonish and inspire scientists and the public alike. But how exactly radiation occurs, and how it differs among taxa and in different settings, as well as why some lineages radiate and others do not, are still unclear. Most likely this is because there is no single answer: Lineages vary in manifold ways, various evolutionary factors act simultaneously, similar evolutionary outcomes can be achieved via alternative paths, and the contingencies of place and time play a large role in guiding the evolutionary process.

References and Notes

1. G. G. Simpson, *The Major Features of Evolution* (Columbia Univ. Press, New York, 1953).
2. D. Schluter, *The Ecology of Adaptive Radiation* (Oxford Univ. Press, Oxford, 2000).
3. J. B. Losos, D. B. Miles, *Am. Nat.* **160**, 147 (2002).
4. T. J. Givnish, in *Molecular Evolution and Adaptive Radiation*, T. J. Givnish, K. J. Sytsma, Eds. (Cambridge Univ. Press, New York, 1997).
5. Some workers add to the definition that divergence occurs unusually rapidly (*2*) or is unusually great (*3*) to emphasize the difference between adaptive radiation and standard processes of evolutionary diversification. In addition, some workers consider any ecologically diverse evolutionary group to represent an adaptive radiation, but we believe that such an interpretation makes "adaptive radiation" nothing more than the product of evolutionary diversification, as opposed to highlighting those groups that have experienced unusually great evolutionary divergence, the cause of which requires explanation. This begs the question of how to identify those groups that constitute adaptive radiations, a topic that has only recently been broached (*3*). Note also that some workers consider the rapidity in which diversification occurred as part of the definition, whereas others consider it an ancillary hypothesis to be tested (*4*).
6. T. Givnish, K. Sytsma, *Molecular Evolution and Adaptive Radiation* (Cambridge Univ. Press, New York, 1997).
7. "Disparity" is a measure of phenotypic—usually morphological—variety, whereas the term "diversity" is reserved for quantification of species richness.
8. D. H. Erwin, *Hist. Biol.* **6**, 133 (1992).
9. E. Mayr, *Animal Species and Evolution* (Belknap, Cambridge, MA, 1963).
10. N. Eldredge *et al.*, *Paleobiology* **31**, 133 (2005).
11. O. Seehausen, *Trends Ecol. Evol.* **19**, 198 (2004).
12. P. R. Grant, B. R. Grant, *How and Why Species Multiply: The Radiation of Darwin's Finches* (Princeton Univ. Press, Princeton, NJ, 2008).
13. D. Schluter, T. D. Price, A. Ø. Mooers, D. Ludwig, *Evolution* **51**, 1699 (1997).
14. C. W. Cunningham, *Syst. Biol.* **48**, 665 (1999).
15. T. Fukami, H. Beaumont, X. Zhang, P. Rainey, *Nature* **446**, 436 (2007).
16. J. Meyer, R. Kassen, *Nature* **446**, 432 (2007).
17. S. S. Chow, C. O. Wilke, C. Ofria, R. Lenski, C. Adami, *Science* **305**, 84 (2004).
18. W. B. Provine, *Stud. Hist. Biol.* **2**, 167 (1978).
19. S. Gavrilets, *Evolution* **57**, 2197 (2003).
20. S. Gavrilets, *Fitness Landscapes and the Origin of Species* (Princeton Univ. Press, Princeton, NJ, 2004).
21. J. Coyne, H. A. Orr, *Speciation* (Sinauer, Sunderland, MA, 2004).
22. J. A. Endler, *Geographic Variation, Speciation, and Clines* (Princeton Univ. Press, Princeton, NJ, 1977).
23. A. R. Templeton, *Annu. Rev. Ecol. Syst.* **12**, 23 (1981).
24. M. Kirkpatrick, V. Ravigné, *Am. Nat.* **159**, S22 (2002).
25. S. Gavrilets, A. Vose, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18040 (2005).
26. S. Gavrilets, A. Vose, in *Speciation and Patterns of Diversity,* R. Butlin, J. Bridle, D. Schluter Eds. (Cambridge Univ. Press, Cambridge, 2009).
27. M. Kawata, A. Shoji, S. Kawamura, O. Seehausen, *BMC Evol. Biol.* **7**, 99 (2007).
28. M. Doebeli, U. Dieckmann, *Nature* **421**, 259 (2003).
29. D. Bolnick, *J. Theor. Biol.* **241**, 734 (2006).
30. H. C. Ito, U. Dieckmann, *Am. Nat.* **170**, E96 (2007).
31. K. Christensen, S. A. de Collobiano, M. Hall, H. J. Jensen, *J. Theor. Biol.* **216**, 73 (2002).
32. S. Gavrilets, *Proc. R. Soc. London B. Biol. Sci.* **266**, 817 (1999).
33. M. R. Pie, J. S. Weitz, *Am. Nat.* **166**, E1 (2005).
34. M. Foote, J. P. Hunter, C. M. Janis, J. J. Sepkoski Jr., *Science* **283**, 1310 (1999).
35. M. Foote, *Annu. Rev. Ecol. Syst.* **28**, 129 (1997).
36. J. B. Losos, *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles* (Univ. of California Press, Berkeley, CA, 2009).
37. D. L. Rabosky, I. J. Lovette, *Proc. R. Soc. London B. Biol. Sci.* **275**, 2363 (2008).
38. O. Seehausen, *Proc. R. Soc. London B. Biol. Sci.* **273**, 1987 (2006).
39. L. J. Harmon, J. A. Schulte II, A. Larson, J. B. Losos, *Science* **301**, 961 (2003).
40. J. J. Sepkoski Jr., M. L. Hulver, in *Phanerozoic Diversity Patterns: Profiles in Macroevolution*, J. W. Valentine, Ed. (Princeton Univ. Press, Princeton, NJ, 1985).

41. D. Jablonski, *Evolution* **62**, 715 (2008).
42. R. J. Whittaker, K. A. Triantis, R. J. Ladle, *J. Biogeogr.* **35**, 977 (2008).
43. R. Gillespie, *Science* **303**, 356 (2004).
44. J. Todd Streelman, P. D. Danley, *Trends Ecol. Evol.* **18**, 126 (2003).
45. D. Ackerly, D. Schwilk, C. Webb, *Ecology* **87**, S50 (2006).
46. C. E. Parent, B. J. Crespi, *Evolution* **60**, 2311 (2006).
47. J. A. Coyne, T. D. Price, *Evolution* **54**, 2166 (2000).
48. J. B. Losos, D. Schluter, *Nature* **408**, 847 (2000).
49. D. Lack, *Darwin's Finches* (Cambridge Univ. Press, Cambridge, 1947).
50. U. Dieckman, M. Doebeli, J. A. J. Metz, D. Tautz, *Adaptive Speciation* (Cambridge Univ. Press, Cambridge, 2004).
51. U. K. Schliewen, D. Tautz, S. Pääbo, *Nature* **368**, 629 (1994).
52. O. Seehausen *et al.*, *Nature* **455**, 620 (2008).
53. R. Lande, *Genetics* **128**, 443 (1991).
54. D. A. Joyce *et al.*, *Nature* **435**, 90 (2005).
55. G. Fryer, *Environ. Biol. Fishes* **45**, 109 (1996).
56. J. W. Valentine, D. Jablonski, in *Evolution, Time and Space*, R. W. Sims, J. H. Price, P. E. S. Whalley, Eds. (Academic Press, London, 1983), pp. 201–226.
57. C. O. Webb, D. D. Ackerly, M. A. McPeek, M. J. Donoghue, *Annu. Rev. Ecol. Syst.* **33**, 475 (2002).
58. J. J. Wiens, C. H. Graham, *Annu. Rev. Ecol. Syst.* **36**, 519 (2005).
59. S. Skulason, T. B. Smith, *Trends Ecol. Evol.* **10**, 366 (1995).
60. H. D. Bradshaw, D. W. Schemske, *Nature* **426**, 176 (2003).
61. R. C. Albertson, J. T. Streelman, T. D. Kocheer, *J. Hered.* **94**, 291 (2003).
62. K. Schwenk, N. Brede, B. Streit, *Philos. Trans. R. Soc. London Ser. B* **363**, 2805 (2008).
63. D. Garant, S. E. Forde, A. P. Hendry, *Funct. Ecol.* **21**, 434 (2007).
64. O. Seehausen, *Curr. Biol.* **16**, R334 (2006).
65. C. Darwin, *Voyage of the Boogle* (Collier, New York, 1909), p. 383.
66. We thank A. Birand, R. Glor, A. Harrison, L. Harmon, D Jablonski, L. Mahler, C. Marshall, and the anonymous reviewers for useful comments. Supported by the NSF (grant DEB-0519777) and the NIH (grant GM56693).

10.1126/science.1157966

REVIEW

Evidence for Ecological Speciation and Its Alternative

Dolph Schluter

Natural selection commonly drives the origin of species, as Darwin initially claimed. Mechanisms of speciation by selection fall into two broad categories: ecological and mutation-order. Under ecological speciation, divergence is driven by divergent natural selection between environments, whereas under mutation-order speciation, divergence occurs when different mutations arise and are fixed in separate populations adapting to similar selection pressures. Tests of parallel evolution of reproductive isolation, trait-based assortative mating, and reproductive isolation by active selection have demonstrated that ecological speciation is a common means by which new species arise. Evidence for mutation-order speciation by natural selection is more limited and has been best documented by instances of reproductive isolation resulting from intragenomic conflict. However, we still have not identified all aspects of selection, and identifying the underlying genes for reproductive isolation remains challenging.

It took evolutionary biologists nearly 150 years, but at last we can agree with Darwin that the origin of species, "that mystery of mysteries" (*1*), really does occur by means of natural selection (*2*–*5*). Not all species appear to evolve by selection, but the evidence suggests that most of them do. The effort leading up to this conclusion involved many experimental and conceptual advances, including a revision of the notion of speciation itself, 80 years after publication of *On the Origin of the Species*, to a definition based on reproductive isolation instead of morphological differences (*6*, *7*).

The main question today is how does selection lead to speciation? What are the mechanisms of natural selection, what genes are affected, and how do changes at these genes yield the habitat, behavioral, mechanical, chemical, physiological, and other incompatibilities that are the reproductive barriers between new species? As a start, the many ways by which new species might arise by selection can be grouped into two broad categories: ecological speciation and mutation-order speciation. Ecological speciation refers to the evolution of reproductive isolation between populations or subsets of a single population by adaptation to different environments or ecological niches (*2*, *8*, *9*). Natural selection is divergent, acting in contrasting directions between environments, which drives the fixation of different alleles, each advantageous in one environment but not in the other. Following G. S. Mani and B. C. Clarke (*10*), I define mutation-order speciation as the evolution of reproductive isolation by the chance occurrence and fixation of different alleles between populations adapting to similar selection pressures. Reproductive isolation evolves because populations fix distinct mutations that would nevertheless be advantageous in both of their environments. The relative importance of these two categories of mechanism for the origin of species in nature is unknown.

In this review, I summarize progress in understanding the general features of speciation by selection. I do not differentiate speciation by sexual selection here because natural selection drives the divergence of mate preferences, by either ecological or mutation-order mechanisms, in most theories of the process (*8*, *11*). I leave out discussion of sympatric and allopatric speciation but instead identify the likelihood of ecological and mutation-order speciation when there is gene flow. I ignore reinforcement, a special type of natural selection thought to favor stronger premating reproductive isolation once postzygotic isolation has evolved. I also ignore speciation by polyploidy, even though selection might be crucial in the early stages.

Speciation and Adaptation from Darwin to Dobzhansky

Appreciation of the connection between adaptation and speciation began with Darwin when a morphological concept of species largely prevailed. In *On the Origin of Species*, Darwin wrote that "I look at the term species, as one arbitrarily given for the sake of convenience to a set of individuals closely resembling each other..." and "The amount of difference is one very important criterion in settling whether two forms should be ranked as species or varieties" (*1*). Under this view, speciation is defined as the accumulation of sufficiently many differences between populations to warrant their classification as separate taxonomic species. Darwin understood the importance of reproductive barriers between species (*1*), but the study of speciation after the publication of this work focused mainly on the evolution of species differences, particularly of morphological traits but also of behavioral and other phenotypic traits.

Under this Darwinian perspective, linking speciation with adaptation was relatively straightforward, requiring only a test of whether phenotypic differences between species were caused by natural selection. For example, at the American Association for the Advancement of Science 1939 speciation symposium [the last major symposium on speciation before the biological species concept (*7*)], an extensive comparative and biogeographic study showcased instances in which derived morphological and life history forms of fishes had arisen over and over again from the same ancestral type (*12*). The repeated, parallel origin of nonparasitic lamprey in streams from the same migratory, parasitic ancestor showed that "Again and

Biodiversity Research Centre and Zoology Department, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. E-mail: schluter@zoology.ubc.ca

again the parasitic lampreys have evolved into nonparasitic forms...correlated with life in small streams, where a suitable food supply in the way of large fish is scarce or seasonal" (*12*). When correlated with environmental factors, such repetition is unlikely to result from chance; environmental selection pressures must therefore be the cause of speciation. "As a result of our recent studies on fishes...weight is constantly being added to the theory that speciation is...under the rigid control of the environment" (*12*). However, this case is only referring to the origin of morphological species.

The turning point for speciation studies came with the modern concept of speciation "Species separation is defined as a stage of the evolutionary process at which physiological isolating mechanisms become developed" (*6*) (here, "physiological" is interpreted to mean evolved reproductive isolation between populations, as distinct from geographical barriers to interbreeding). Subsequently, species were defined as "groups of interbreeding natural populations that are reproductively isolated from other such groups" (*7*). From this point on, the study of speciation was the study of the evolution of reproductive isolation (*3*). Progress up to then in understanding the link between morphological speciation and adaptation was largely forgotten, its contributions uncertain under the new concept.

The biological species concept must surely have made it more difficult to investigate any link between speciation and natural selection. T. Dobzhansky (*13*) suggested that the genes underlying differences between populations in ordinary phenotypic traits were unlikely to be the basis of reproductive isolation. He later changed his mind, but at the time this viewpoint, and the generally greater difficulty of studying reproductive isolation than morphology, must have discouraged many from pursuing the connection. Virtually no research effort followed that tested the role of adaptation in speciation.

Models of Speciation by Selection

The topic of natural selection in speciation is once again receiving attention. The two most general hypotheses involving selection are ecological and mutation-order speciation. Ecological speciation is defined as the evolution of reproductive isolation between populations by divergent natural selection arising from differences between ecological environments (*2*, *8*, *9*, *14*). It predicts that reproductive isolation should evolve between populations adapting to contrasting environments but not between populations adapting to similar environments. The basic idea has been around for a while (*7*), although it was tested only recently. The agents of divergent selection are extrinsic and can include abiotic and biotic factors such as food resources, climate, habitat, and interspecies interactions such as disease, competition, and behavioral interference. Ecological speciation can lead to the evolution of any type of reproductive isolation, including premating isolation, hybrid sterility, and intrinsic hybrid inviability as well as extrinsic, ecologically based pre- and postzygotic isolation. Speciation by sexual selection is ecological speciation if ecologically based divergent selection drives divergence of mating preferences, for example by sensory drive (*15*).

In accordance with (*10*), mutation-order speciation is defined as the evolution of reproductive isolation by the fixation of different advantageous mutations in separate populations experiencing similar selection pressures. Whereas different alleles are favored between populations under ecological speciation, the same alleles would be favored in different populations under mutation-order speciation. Divergence occurs anyway because, by chance, the populations do not acquire the same mutations or fix them in the same order. Divergence is therefore stochastic but the process is distinct from genetic drift. It can occur in both small and large (though not infinite) populations. Selection can be ecologically based under mutation-order speciation, but ecology does not favor divergence as such. It can lead to the evolution of any type of reproductive isolation, with the exception of ecologically based pre- and postzygotic isolation.

Speciation resulting from intragenomic conflict such as meiotic drive or cytoplasmic male sterility (Fig. 1B) is likely to be mutation-order speciation because, by chance, the initial mutations causing drive and those countering it are unlikely to be the same in separate populations. Speciation by sexual selection is mutation-order speciation if divergence of mate preferences or gamete recognition occurs by the fixation of alternative advantageous mutations in different populations, as by sexual conflict (*16*). Divergence in song and other learned components of behavior under purely social selection, not molded by selection for efficient signal transmission (*5*), is the cultural equivalent of the mutation-order process. Additional scenarios are elaborated in (*5*).

Both models of speciation, ecological and mutation-order, are theoretically plausible, and only data can determine their relative importance in nature. The key is to figure out by which mechanism reproductive isolation first evolved (*3*). Once the earliest genetic differences have accumulated between populations by either process, subsequent mutations might be favored in one population and not the other because of epistatic interactions with genetic background (*10*). Hence, epistasis, including that producing Dobzhansky-Muller incompatibilities in hybrids between species (*3*), can result from either ecological or mutation-order speciation.

Speciation can be rapid under both speciation models, because alleles are driven to fixation by natural selection in both cases. However, under the mutation-order process, the same alleles, if present, would be favored in every population, at least in the early stages of divergence. For this reason, mutation-order speciation is difficult when

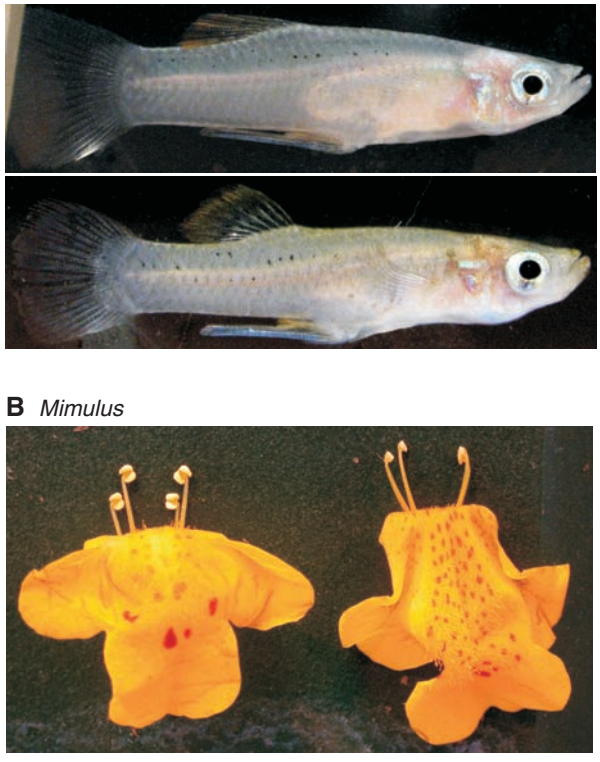


Fig. 1. (**A**) Example of ecological speciation. Repeatedly and independently, the mosquito fish, *Gambusia hubbsi*, inhabiting blue holes in the Bahamas has evolved a larger caudal region and smaller head in the presence of predators (top) than in their absence (bottom) (*29*). In laboratory trials, the probability of two individuals mating was higher when they were from different populations having the same predation environment (and similar body shape) than when they were from opposite predation environments. [Photo credit: Brian Langerhans (*29*)]. (**B**) Example of reproductive isolation evolving under the mutation-order mechanism. Male-fertile (left) and male-sterile (right) flowers of F2 hybrids between an Oregon population of monkey flowers (*M. guttatus*) having a cytoplasmic male sterility element and nuclear restorer and a closely related species (*M. nasutus*) having neither (*46*, *47*). Both flowers shown have *M. guttatus* cytoplasm. The flower on the left also has the nuclear restorer, whereas the one on the right, with undeveloped anthers, lacks the restorer. [Photo credit: Andrea Case (*47*)]

there is gene flow, because gene flow increases the possibility that favorable mutations occurring in one population will spread to other populations, preventing divergence (*17*, *18*). Any process resulting in low levels of gene flow, including selection, facilitates subsequent divergence by the mutation-order process (*19*). In contrast, ecological speciation can proceed with or without gene flow, although it is easiest when gene flow is absent.

Experiments with laboratory populations of *Drosophila* and yeast demonstrate the plausibility of ecological speciation. In those instances when measurable pre- and postmating reproductive isolation evolved, it was greater between lines subjected to different environments than between lines raised under homogeneous conditions (*20*, *21*). Laboratory experiments on various microbes maintained under homogeneous conditions for many generations have detected genetic divergence consistent with the mutation-order process (*22*), but effects on reproductive isolation have not been explored.

Two approaches investigate the mechanisms of speciation by natural selection in nature. The bottom-up approach involves (i) genetic mapping of reproductive isolation between closely related species, (ii) testing whether discovered genes exhibit a genomic signature of positive selection, and (iii) identifying the phenotype and source of fitness effects of alternative alleles at selected loci. The approach has been hugely successful in identifying major genes implicated in hybrid inviability (*Hmr*, *Lhr*, *Nup96*), sterility (*Odsh*, *JYAlpha*), and sexual isolation (*ds2*) between *Drosophila* species. Most of these genes show molecular signatures of positive selection, proving natural selection's role (*3*), provided that fixation occurred before complete reproductive isolation rather than afterward. The top-down approach involves identifying (i) the phenotypic traits under divergent selection, (ii) those traits associated with reproductive isolation, and (iii) the genes underlying traits and reproductive isolation. Step (iii) has been challenging under both approaches but is needed to understand how selection has led to reproductive isolation.

Ecological Speciation

Evidence for ecological speciation has accumulated from top-down studies of adaptation and reproductive isolation [reviewed in (*2*, *8*, *9*)]. We now know of many real species that have, at least in part, evolved by divergent natural selection between environments. The connections between selection on ordinary phenotypic traits and reproductive isolation are often strong and straightforward. It follows that much of the genetic basis of reproductive isolation should involve ordinary genes that underlie differences in phenotypic traits. But we still know little about the genetics of ecological speciation.

One line of evidence comes from tests of parallel speciation, whereby greater reproductive isolation repeatedly evolves between independent populations adapting to contrasting environments than between independent populations adapting to similar environments (*20*, *23*). A major challenge in applying the test to natural populations is to eliminate the possibility that each ecotype has originated just once and has spread to multiple locales. This is difficult because gene flow of neutral markers between closely related but nearby populations can result in the false appearance of multiple independent origins of these populations when evaluated by phylogenies (*3*, *24*). However, multiple origins are supported in several examples of parallel speciation, including the sympatric benthic-limnetic species pairs of threespine stickleback in young lakes of British Columbia (*25*, *26*), the repeated origin of divergent marine and stream populations of threespine stickleback around the Northern Hemisphere (*27*), ecotypes of *Timema* walking stick insects living on different host plants (*28*), *Littorina* marine snail ecotypes inhabiting different zones of the intertidal (*24*), and mosquito fish inhabiting blue holes with and without fish predators in the Bahamas (*29*) (Fig. 1A). In these studies, it was shown that males and females are more likely to mate if they are of the same ecotype, regardless of relatedness as indicated by phylogenetic affinity.

Ecological speciation is also supported by examples of premating reproductive isolation in which individuals choose or preferentially encounter mates on the basis of phenotypic traits that are under ecologically based divergent selection. Examples include assortative mating by host choice in insects, body size and coloration in fish, beak size in birds, pollinator preferences for specific phenotypic floral traits, and variation in flowering time—traits inferred to be under divergent selection between environments [see examples in (*8*, *30*, *31*)].

Ecologically based divergent selection has also been directly measured, as shown by reduced fitness of each ecotype in the environment of the other [immigrant inviability; reviewed in (*31*, *32*)] and by reduced fitness of hybrids in the parental environments [extrinsic postzygotic isolation (*33*)]. For example, each of the coastal perennial and inland annual races of the monkey flower (*Mimulus guttatus*) along the west coast of North America has low fitness when transplanted to the habitat of the other (*31*). This is an example of active selection on phenotypic differences, and it also constitutes direct reproductive isolation because it is an evolved barrier to gene flow between parental populations. Multiple traits are probably involved, including flowering time and tolerance of salt and drought. This type of reproductive isolation is context-dependent and is weakened in intermediate environments. On the other hand, active selection favors the evolution of ever-greater differences between populations, which may strengthen the barrier to gene flow (*20*).

It is unclear how much reproductive isolation typically evolves by ecologically based divergent selection in nature. We can approximate an answer from estimates of the combined contribution of active selection on traits and trait-based assortative mating, as compared with other forms of reproductive isolation (Fig. 2 and table S1). These estimates are incomplete because individual studies may lack data on components of reproductive isolation, separate components may not be independent, and the strength of barriers between species may not be symmetric (*34*). Nevertheless, compilation of the data shows that the amount of reproductive isolation attributable to active selection and trait-based assortative mating is at least as strong, on average, as the amount from components of reproductive isolation lacking identifiable causes (Fig. 2). The unidentified component of speciation, if built by selection and not genetic

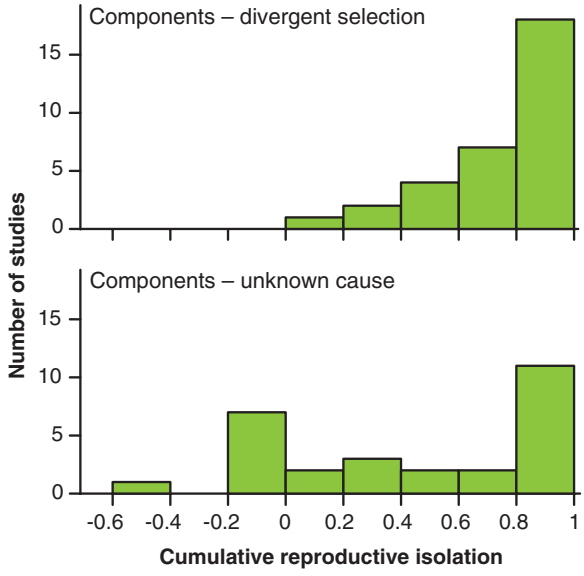


Fig. 2. Estimates of the magnitude of reproductive isolation resulting from divergent selection components (**top**), compared with other components lacking identifiable causes (**bottom**). Divergent selection components include those attributable to active selection on traits (immigrant inviability and extrinsic postzygotic isolation) and to trait-based assortative mating (habitat preference, floral isolation, and breeding time). The unattributed components include intrinsic hybrid inviability, sexual selection against hybrids, pollen competition, and reduced hybrid fecundity. Data were taken from (*32*, *31*) (table S1). A negative value indicates that hybrids had higher fitness than the parental species for at least one component of postzygotic isolation. One data value of −2.66 was left out of the bottom panel.

drift, could be the result of either ecological or mutation-order mechanisms.

These examples indicate a growing knowledge of the mechanisms of selection and its consequences for reproductive isolation. At this point, the most glaring deficiency is our knowledge of the impact of selection on genes. Optimistically, progress is being made with genetic mapping to identify quantitative trait loci (QTLs) and genes or regulatory control regions that affect individual phenotypic traits on which components of reproductive isolation depend. Examples include the *yup* QTL, which affects flower color differences between the monkey flowers, *Mimulus cardinalis* and *M. lewisii* (*35*). Swapping alleles of this QTL between the species with repeated backcrossing resulted in shifts in pollinator preference and, hence, indirectly affected premating isolation. Survival and salt tolerance of second-generation hybrids between the sunflowers *Helianthus annuus* and *H. petiolaris* transplanted to the salt marsh habitat of their hybrid descendent species (*H. paradoxus*) mapped strongly to a QTL identified as the salt tolerance gene *CDPK3* (*36*).

Another form of investigation involves the analysis of genome scans of ecologically different populations and species. These scans compare allelic variation within and between populations at many marker loci spaced throughout the genome (*37*). Markers that show excessive differentiation between populations (outliers) may indicate selection on nearby genes. The method is particularly informative when applied to populations with ongoing hybridization, because outlier markers may identify points in the genome that resist the homogenizing influence of gene flow, perhaps indicating genomic regions under divergent selection. However, sets of genes that diverged under a mutation-order process can produce the same pattern (*17*, *18*), which makes analysis of such studies more difficult. Clues to whether ecologically based divergent selection is involved are gained if outliers at the same genomic locations turn up repeatedly in scans between populations that inhabit contrasting environments (*38*) and by identifying phenotypic traits under divergent selection that map to those locations in the genome (*36*, *39*, *40*). As genomic resources increase for more species, it will be possible to measure natural selection directly on genomic regions of interest by transplanting otherwise relatively homogenous experimental populations containing alternative alleles into the environments of the parent species (*35*).

Mutation-Order Speciation

Mounting evidence for divergent selection in speciation does not diminish the potential role of mutation-order divergence. It may be that the mutation-order process is more difficult to detect, or perhaps we have not looked hard enough at species with only small ecological differences (*5*). We still do not know much about the selective factors causing mutation-order speciation.

Evidence for mutation-order speciation comes from instances in which reproductive isolation apparently evolved as a by-product of conflict resolution between genetic elements within individuals (intragenomic conflict), such as cytoplasmic male sterility in hermaphroditic plants (Fig. 1B), and genetic elements conferring meiotic drive. Under both mechanisms, a mutation arises that can distort representation in gametes and spreads in a selfish manner, even though such an element reduces overall fitness of the organism that bears it. This, in turn, places selection on mutations in other genes that counter the selfish element's effects and restore more equal genetic representation in gametes. Distorter and restorer mutations are unlikely to be the same in different populations regardless of environment; thus the process leads to divergence. The mismatch between the distorter in one population and the restorer in the other can result in hybrid sterility or inviability and, thus, reproductive isolation (*3*, *41*). Numerous examples of selfish elements, such as those observed in cytoplasmic male sterility of plants, support these hypotheses (*42*, *43*). In addition, partial reproductive isolation generated by meiotic drive has been identified in *Drosophila* [reviewed in (*3*, *41*)]. Sexual conflict is also expected to lead to mutation-order speciation, but there are few compelling examples (*3*). The contribution by these mechanisms to speciation is still uncertain, however. The alleles responsible for meiotic drive and cytoplasmic male sterility may be prevented from spreading to fixation because selection on such elements is frequency-dependent (*43*) and because restorer alleles arise and weaken selection on the distorter elements (*44*). Second, if divergent populations come into secondary contact, the alleles within each population causing cytoplasmic male sterility or meiotic drive (and the corresponding restorer alleles) will spread between the populations by gene flow, eliminating that component of reproductive isolation (*43*). Hence, for these mechanisms to contribute to speciation, the fitness of hybrids must be reduced to very low levels, or other incompatibilities must arise that interact with these genes to prevent their spread after secondary contact.

Conclusions

Our understanding of the role of natural selection in speciation has come a long way since Darwin's time. If he were here to witness, he would most likely be staggered by the discoveries of genes and molecular evolution and astonished at the prospect that evolutionary conflict between genes could generate reproductive isolation (*45*). Mostly, I expect that he would be chuffed by mounting evidence for the role of natural selection on phenotypic traits in the origin of species. This is really what *On the Origin of Species* was all about. Between 1859 and the present, the general acceptance of the biological species concept altered the focus of speciation studies. Yet, the discovery that reproductive isolation can be brought about by ecological adaptation in ordinary phenotypic traits bridges Darwin's science of speciation and our own.

The most obvious shortcoming of our current understanding of speciation is that the threads connecting genes and selection are still few. We have many cases of ecological selection generating reproductive isolation with little knowledge of the genetic changes that allow it. We have strong signatures of positive selection at genes for reproductive isolation without enough knowledge of the mechanisms of selection behind them. But we hardly have time to complain. So many new model systems for speciation are being developed that the filling of major gaps is imminent. By the time we reach the bicentennial of the greatest book ever written, I expect that we will have that much more to celebrate.

References and Notes

1. C. Darwin, *On the Origin of Species by Means of Natural Selection* (J. Murray, London, 1859).
2. D. Schluter, *The Ecology of Adaptive Radiation* (Oxford Univ. Press, Oxford, 2000).
3. J. A. Coyne, H. A. Orr, *Speciation* (Sinauer Sunderland, MA, 2004).
4. P. R. Grant, B. R. Grant, *How and Why Species Multiply: The Radiation of Darwin's Finches* (Princeton Univ. Press, Princeton, NJ, 2008).
5. T. Price, *Speciation in Birds* (Roberts & Company, Greenwood Village, CO, 2008).
6. T. G. Dobzhansky, *Genetics and the Origin of Species* (Columbia Univ. Press, New York, 1937).
7. E. Mayr, *Systematics and the Origin of Species* (Columbia Univ. Press, New York, 1942).
8. D. Schluter, *Trends Ecol. Evol.* **16**, 372 (2001).
9. H. D. Rundle, P. Nosil, *Ecol. Lett.* **8**, 336 (2005).
10. G. S. Mani, B. C. Clarke, *Proc. R. Soc. London Ser. B Biol. Sci.* **240**, 29 (1990).
11. M. Turelli, N. H. Barton, J. A. Coyne, *Trends Ecol. Evol.* **16**, 330 (2001).
12. C. L. Hubbs, *Am. Nat.* **74**, 198 (1940).
13. T. Dobzhansky, *Am. Nat.* **74**, 312 (1940).
14. Additional references are available as supporting material on *Science* Online.
15. J. A. Endler, *Am. Nat.* **139**, S125 (1992).
16. W. R. Rice *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6527 (2005).
17. N. Barton, B. O. Bengtsson, *Heredity* **57**, 357 (1986).
18. A. S. Kondrashov, *Evolution* **57**, 151 (2003).
19. L. H. Rieseberg, personal communication (2008).
20. W. R. Rice, E. E. Hostert, *Evolution* **47**, 1637 (1993).
21. J. R. Dettman, C. Sirjusingh, L. M. Kohn, J. B. Anderson, *Nature* **447**, 585 (2007).
22. Z. D. Blount, C. Z. Borland, R. E. Lenski, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7899 (2008).
23. D. Schluter, L. M. Nagel, *Am. Nat.* **146**, 292 (1995).
24. R. K. Butlin, J. Galindo, J. W. Grahame, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **363**, 2997 (2008).
25. H. D. Rundle, L. Nagel, J. W. Boughman, D. Schluter, *Science* **287**, 306 (2000).
26. E. B. Taylor, J. D. McPhail, *Proc. R. Soc. London Ser. B Biol. Sci.* **267**, 2375 (2000).
27. J. S. McKinnon *et al.*, *Nature* **429**, 294 (2004).
28. P. Nosil, S. P. Egan, D. J. Funk, *Evolution* **62**, 316 (2008).
29. R. B. Langerhans, M. E. Gifford, E. O. Joseph, *Evolution* **61**, 2056 (2007).
30. S. Via, *Trends Ecol. Evol.* **16**, 381 (2001).
31. D. B. Lowry, J. L. Modliszewski, K. M. Wright, C. A. Wu, J. H. Willis, *Philos. Trans. Royal Soc. London Ser. B Biol. Sci.* **363**, 3009 (2008).

32. P. Nosil, T. H. Vines, D. J. Funk, *Evolution* **59**, 705 (2005).
33. H. D. Rundle, M. C. Whitlock, *Evolution* **55**, 198 (2001).
34. N. H. Martin, J. H. Willis, *Evolution* **61**, 68 (2007).
35. H. D. Bradshaw, D. W. Schemske, *Nature* **426**, 176 (2003).
36. C. Lexer, Z. Lai, L. H. Rieseberg, *New Phytol.* **161**, 225 (2004).
37. M. A. Beaumont, *Trends Ecol. Evol.* **20**, 435 (2005).
38. P. Nosil, D. J. Funk, D. Ortiz-Barrientos, *Mol. Ecol.* **18**, 375 (2009).
39. S. M. Rogers, L. Bernatchez, *Mol. Biol. Evol.* **24**, 1423 (2007).
40. S. Via, J. West, *Mol. Ecol.* **17**, 4334 (2008).
41. H. A. Orr, J. P. Masly, N. Phadnis, *J. Hered.* **98**, 103 (2007).
42. D. A. Levin, *Syst. Bot.* **28**, 5 (2003).
43. L. H. Rieseberg, J. H. Willis, *Science* **317**, 910 (2007).
44. D. C. Presgraves, *Trends Genet.* **24**, 336 (2008).
45. D. C. Presgraves, *Curr. Biol.* **17**, R125 (2007).
46. L. Fishman, J. H. Willis, *Evolution* **60**, 1372 (2006).
47. A. L. Case, J. H. Willis, *Evolution* **62**, 1026 (2008).
48. We thank M. Arnegard, R. Barrett, A. Case, H. Hoekstra, M. Noor, P. Nosil, S. Otto, T. Price, L. Rieseberg, S. Rogers, A. Schluter, S. Via, M. Whitlock, J. Willis, and a reviewer for assistance and comments. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Canada Foundation for Innovation.

Supporting Online Material
www.sciencemag.org/cgi/content/full/323/5915/737/DC1
Tables S1 to S3
References

10.1126/science.1160006

REVIEW

The Bacterial Species Challenge: Making Sense of Genetic and Ecological Diversity

Christophe Fraser,[1]* Eric J. Alm,[2,3,4] Martin F. Polz,[2] Brian G. Spratt,[1] William P. Hanage[1]

The Bacteria and Archaea are the most genetically diverse superkingdoms of life, and techniques for exploring that diversity are only just becoming widespread. Taxonomists classify these organisms into species in much the same way as they classify eukaryotes, but differences in their biology—including horizontal gene transfer between distantly related taxa and variable rates of homologous recombination—mean that we still do not understand what a bacterial species is. This is not merely a semantic question; evolutionary theory should be able to explain why species exist at all levels of the tree of life, and we need to be able to define species for practical applications in industry, agriculture, and medicine. Recent studies have emphasized the need to combine genetic diversity and distinct ecology in an attempt to define species in a coherent and convincing fashion. The resulting data may help to discriminate among the many theories of prokaryotic species that have been produced to date.

The species debate in microbiology is not only about a human desire to catalog bacterial diversity in a consistent manner, but is also a fundamental argument because of what it reveals about our ignorance of how evolutionary forces form, shape, and extinguish bacterial genetic lineages, of the mechanisms of differentiation between subpopulations sharing common descent, and of the process of adaptation to new niches and changing environments. Animal species are defined by their morphological and behavioral traits and by their ability or inability to interbreed, but such categories cannot easily be applied to the Bacteria or Archaea (or indeed to many eukaryotic microbes). Instead, taxonomists have been forced to rely on biochemical tests and limited morphological characteristics for this purpose. Naturally, biochemical characters have been selected for the convenience of taxonomists; they reflect only a tiny subset of those characters that allow bacteria to use different resources in the environment, and only capture a small fraction of the true diversity in this superkingdom of life. More recently, molecular methods [particularly DNA-DNA hybridization and ribosomal RNA (rRNA) sequencing] have helped to define species, but these methods have serious limitations and cannot reliably assign a large collection of similar strains to species (e.g., rRNA sequences are too conserved to resolve similar species). rRNA sequence surveys have, however, revealed the extraordinary variety of microbial life, much of it uncultured (*1*). Beyond this, taxa too similar to be distinguished and circumscribed by rRNA sequences have revealed further diversity through multilocus sequence analysis (MLSA) (*2*) and metagenomic studies (*1*), and this diversity needs to be explained by theory. Thus, practical difficulties, lack of theory, and observations of vast amounts of as yet unclassified microbial diversity have all fueled the controversy of how one defines a bacterial species (*3*–*8*).

Genetic Clustering

Darwin commented that "all true classification is genealogical" [(*9*), p. 404]. Taxonomists have thus used sequence relatedness to define cutoff values that place two bacterial isolates into the same or different species. The overall genetic relatedness of isolates may be measured by the extent of DNA hybridization between them, and those that show 70% or more DNA hybridization are defined as the same species (*2*, *10*). Such cutoffs imply that sequences that cluster together with a certain amount of similarity must be from the same species, and moreover that this cutoff value is applicable to all groups of bacteria or archaea. Recent MLSA studies, which use the concatenated sequences of multiple housekeeping genes to discern clustering patterns among populations of closely related taxa, suggest that species defined by taxonomists in many cases correspond to well-resolved sequence clusters. However, these studies also show that there is no universal cutoff or descriptor of clusters that characterizes a species. Furthermore, inspection of the clusters does not always clearly reveal which level in the hierarchy is more fundamental than any other (Fig. 1) (*7*).

As an example, Fig. 1A shows the relationships among multiple isolates of three closely related streptococcal species. *Streptococcus pneumoniae* is a major human pathogen, *S. mitis* is a commensal bacteria with a history of taxonomic uncertainty (*11*), and *S. pseudopneumoniae* is a recently described organism of uncertain status that nonetheless corresponds to a distinct cluster in these data (*12*). There are striking differences in the amount of sequence diversity observed within homologous housekeeping genes in these named species, ranging from 1.2% for *S. pneumoniae* to 3.0% for *S. pseudopneumoniae* and up to 5.0% for *S. mitis*. The distance between two randomly selected *S. mitis* genotypes is similar to the average distance between *S. pneumoniae* and *S. pseudopneumoniae* genotypes (5.1%) (*2*). This implies that the use of a fixed level of sequence divergence for differentiating species would tend to either rejoin *S. pneumoniae* and *S. pseudopneumoniae*, or break up *S. mitis* so that nearly every isolate was a species of its own. This is clearly unsatisfactory.

Habitats and Ecological Differentiation

A clear natural criterion to identify clusters of evolutionary importance, which we might want to call species, is to find ecological features that distinguish them from close relatives. Among pathogens, the ability to cause a distinctive disease has historically been used to define species,

[1]Department of Infectious Disease Epidemiology, Imperial College London, London W2 1PG, UK. [2]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Broad Institute of MIT and Harvard University, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: c.fraser@imperial.ac.uk

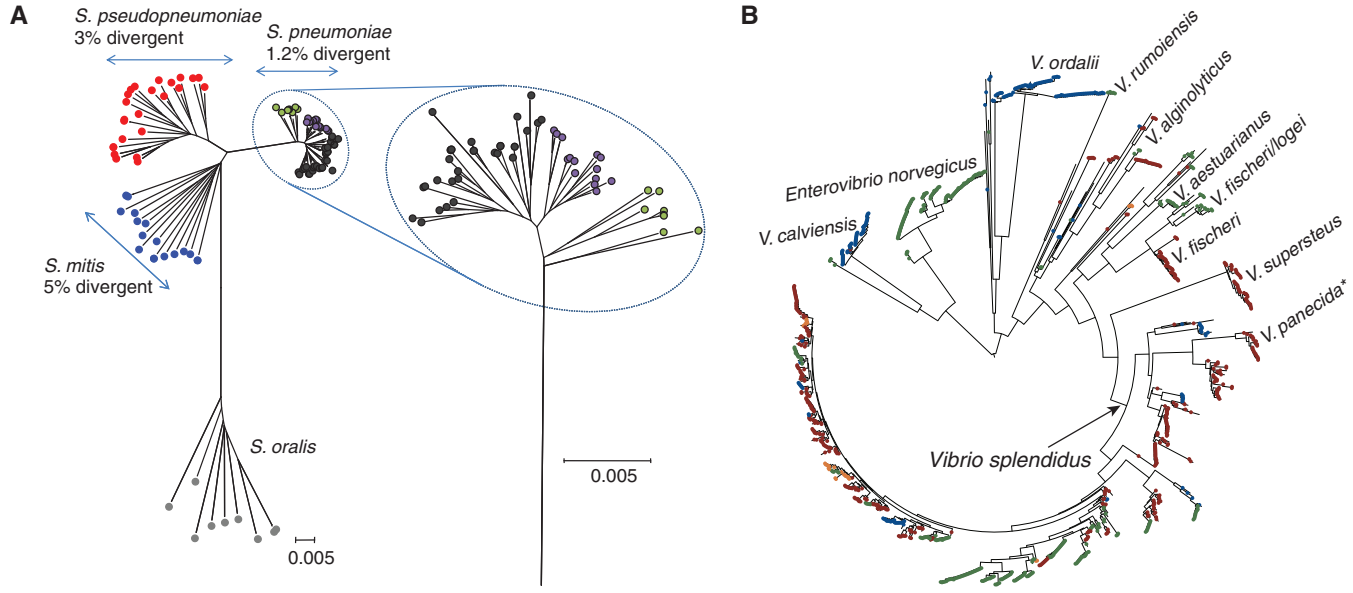


Fig. 1. Multilocus sequence analysis of closely related species. (**A**) Radial minimum evolution tree constructed using MEGA4, showing clusters among 97 isolates of four *Streptococcus* species identified as indicated. The tree was built using concatenates of six housekeeping loci, resulting in a total of 2751 positions in the final data set (*2*). Distances were calculated as the percentage of variant nucleotide sites. The mean distance within the clusters, calculated by MEGA4, is shown. To the right, the pneumococcal cluster is shown at larger scale, and putative subclusters are indicated in dark gray, purple, and green. (**B**) Ecological associations of Vibrionaceae sequence clusters (*13*). Habitats (colored dots) were estimated as differential distributions of groups of closely related strains among samples (size fractions enriched in different environmental resources). Clusters associated with named species are evident, and in most cases species show a clear predilection for one of the habitats. The exception is *V. splendidus*, which breaks up into many closely related ecological populations. Asterisk denotes that trees based on additional loci indicate that the placement of *V. panecida* within *V. splendidus* may be an artifact of horizontal gene transfer at the Hsp60 locus.

but pathogens constitute only a minute fraction of overall bacterial diversity. Mapping of bacterial diversity onto environmental resources indicates that closely related groups of bacteria can be ecologically divergent. For example, fine-scale resource partitioning has been observed among coastal *Vibrio* populations coexisting in the water column (*13*). Partitioning was discovered because strains were collected from distinct, ecologically informative samples, and the phylogenetic structure of the ecologically differentiated populations was superimposed on their habitats. Habitats were defined using an empirical modeling approach. This analysis revealed high levels of specialization for some populations (e.g., *V. ordalii* is only found as single free-swimming cells), whereas others are more generalist (Fig. 1B) and can colonize a wide variety of surfaces, including organic particles and zooplankton in the water column (*13*). Most of the predicted *Vibrio* populations are deeply divergent from each other, and in many cases are congruent with named species; however, *V. splendidus* is a notable exception and splits into numerous closely related groups with distinct ecological preferences, presumably indicating recent ecological radiation from a sympatric ancestral population (*13*). Thus, genetic clusters that correlate with ecology can be discerned.

What do the genetic data tell us about mechanisms of population differentiation and the evolutionary history of the microbes in question? That bacteria are organized into genetic clusters is not, per se, a very interesting observation; many or most models of a population reproducing with a small amount of mutation will eventually produce populations consisting of clusters of related organisms, irrespective of the details of the evolutionary forces or ecological differentiation. A more substantial observation is that there is very little neutral diversity in many populations of microbes, from which we may infer some features of the selective landscape. Neutral diversity is the amount of polymorphism that is evident in noncoding regions or results in synonymous substitutions. One common measure of neutral diversity is the effective population size N_e, defined as the size of a population evolving in the absence of selection that would generate as much neutral diversity as is actually observed. Estimates of N_e for bacteria range from 10^5 to 10^9 (*14*–*18*). To put this into context, the numbers of *Vibrio* cells per cubic meter of seawater in temperate coastal regions range from 10^8 to 10^9 (*19*), which suggests vast census population sizes ($>10^{20}$). This observation—a mismatch of many orders of magnitude between effective population size and census population size (true of most bacteria studied to date)—was originally used to counter claims of neutrality and instead argue that all genetic variation was adaptive (*20*, *21*). However, there are several different mechanisms that can explain this mismatch (Fig. 2).

Whatever mechanisms are driving the differentiation of bacteria into clusters, they must restrict the accumulation of neutral diversity. The first proposed mechanism was based on artificial selection experiments with bacteria grown for extended periods under stable conditions in chemostats, which showed repeated selective sweeps in which the whole genome was thought to hitchhike to fixation along with an advantageous mutation (periodic selection) (*22*). Selective sweeps can purge almost all genetic diversity in the population and thus constitute a candidate mechanism for reducing neutral variation (*23*).

Niches and Ecotypes

To extend this model, one can consider multiple ecological niches characterized by the selective advantages they confer to specific genes. This is the ecotype model, where genes adapted to specific niches cause selective sweeps within those niches but not in other niches. In this way the population will undergo adaptation and differentiation while maintaining relatively low levels of neutral diversity, as selective sweeps confined to each ecotype regularly purge the population of any diversity that might have accumulated (Fig. 2A). Crucially, what neutral diversity we do observe is predicted to be associated with adaptive traits. The ability of such selective sweeps to limit the effective population size has been recognized for some time (*17*, *23*), and this model has been substantially developed by Cohan and colleagues (*4*, *16*, *24*). Because it links patterns of genetic differentiation with adaptation, and makes

reference to the unifying biological principles of selection and niche partitioning, the ecotype has rightly become popular as a framework within which to discuss bacterial evolution, speciation, and ecology.

The ecotype model (*4*, *16*, *24*) predicts that common ancestry will be preserved among bacterial populations within niches (which should be monophyletic), and thus predicts that ecotypes are coherent self-contained gene pools. As a result, it has been suggested that ecotypes should be considered as putative or actual species, depending on the level of genetic differentiation from the ancestral population. This model therefore has the advantage of providing a mechanistic understanding of the evolutionary processes, as well as an organizing principle for classifying species, that is based on experimental observations of bacterial populations.

However, these observations of repeated selective sweeps were made in chemostats, whereas natural environments are markedly unstable and diverse. How would one detect the presence of selective sweeps in natural bacterial populations? The most conclusive examples come not from bacteria but from RNA viruses, which mutate at much higher rates than DNA-based life forms. It has been established from sequences collected over many years that the population structure of the human influenza virus is predominantly driven by repeated selective sweeps (*25*) and that the resulting effective population size N_e (<100) is very much smaller than observed for bacteria. The use of longitudinal ecological and genetic data to distinguish between competing models of evolution has a long pedigree in eukaryotic biology (*26*). On the basis of these analogies, any inference of a population structure driven by selective sweeps would require good longitudinal data from natural bacterial populations, as well as observations of episodic crashes in diversity causally associated with genetic changes and not associated with changes in ecological covariates.

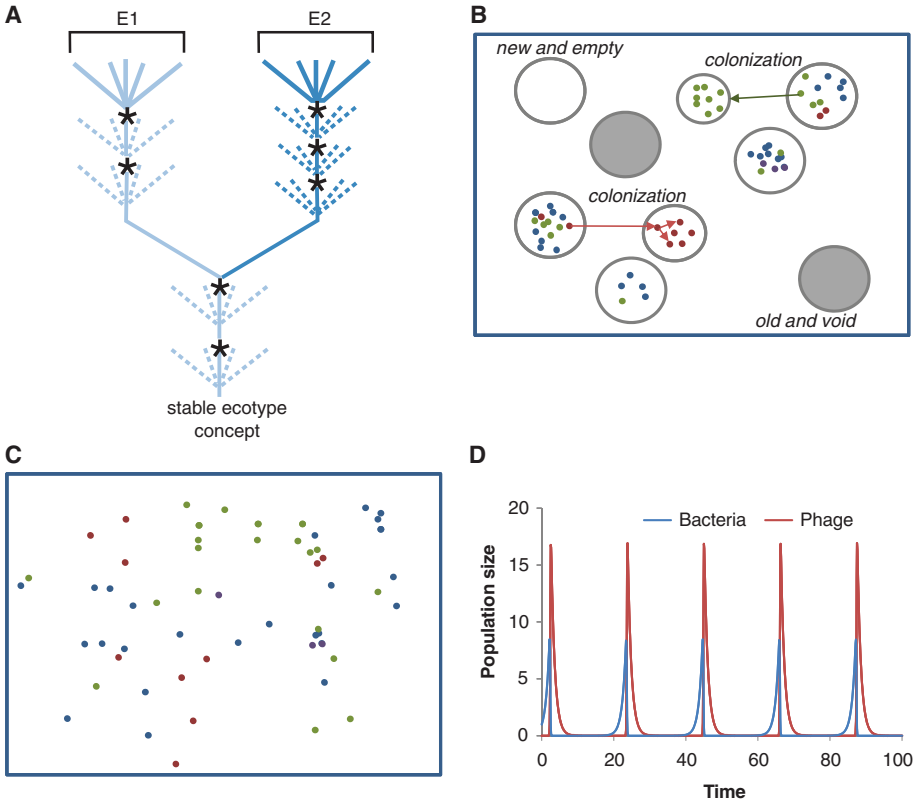


Fig. 2. Different models of microbial evolution that lead to low values of N_e. (**A**) The ecotype model of bacterial population differentiation. The tree shows a single bacterial lineage that differentiates into two sublineages (E1 and E2) that differ in some aspect of their ecology. Periodic selection (a selective sweep) occurs at the points marked by asterisks and eliminates almost all of the diversity that has arisen since the last episode of periodic selection, which is shown by the dashed branches (diversity purged by periodic selection) or solid branches (existing diversity) on the tree. As the two populations are ecologically distinct (i.e., ecotypes), periodic selection in one sublineage does not influence diversity in the other sublineage and vice versa. Each ecotype can therefore diverge to become separate species. Reproduced from (*24*) with permission. (**B**) A metapopulation. Patches of varying size (gray circles) are vacant (empty) or may be colonized by a single genotype randomly acquired from another patch. Strains may diversify within a patch (as shown by different colors representing distinct genotypes), which may colonize empty patches as described above. A characteristic of this sort of metapopulation is patch turnover, in which patches occasionally become unable to support colonization and their inhabitants are removed (solid gray circles). (**C**) A neutral model with small population size. Different genotypes (different colors) arise by mutation or recombination and increase or decrease in the population by random drift. For some purposes, this simple model is an adequate effective description of the more complex processes represented in (A), (B), and (D), and of other more complex evolutionary models not described in this review. (**D**) Predator-prey dynamics and population bottlenecks. Regular population bottlenecks can drastically shrink the effective population size. In this case, bacteria-phage predator-prey dynamics are simulated with a classical Lotka-Volterra model, which can generate oscillations in population size of any amplitude. Population sizes and time axes are in arbitrary units for illustrative purposes only.

Bottlenecks, Metapopulations, and Local Extinctions

The essential element of the ecotype model with respect to limiting neutral diversity is not niche adaptation per se, but rather the effective bottleneck caused by the replacement of the whole population by descendants from a single individual and the resulting extinction of all other lineages (Fig. 2A). Other mechanisms that induce or involve regular population bottlenecks will also restrict neutral diversity. Metapopulation structure, in which the population is divided into patches and where individuals disperse between patches, can generate very low effective population sizes if patches turn over (i.e., if patches are only intermittently able to support bacterial growth, and if a small number of bacteria are dispersed to colonize empty patches) (Fig. 2B) (*27*). This structure well describes the situation for parasites, which can colonize a host but are then forced to move on because the host develops immunity or dies (*17*). It also describes any situation where bacteria use a limited resource intensively for short bursts, followed by dispersal to new resource patches (e.g., colonization of organic particles in seawater by *Vibrio* populations). This metapopulation model is fundamentally different from the ecotype model because it does not predict an association between neutral diversity and adaptive traits.

The relevance of the metapopulation model to the species question is that, although highly idealized and simplified, it may capture some of the effects of complexity and instability of actual ecosystems on population structure. Selective sweeps are predicted to be inevitable in simple, stable environments but not in complex metapopulations [a point partly addressed in (*28*)].

A metapopulation may evolve, differentiate, and adapt without global selective sweeps. Diversity lost by a local selective sweep in one patch may be rescued and reintroduced from other patches. The ecotype model, with its predicted monophyletic relationship between niche and genotype, may therefore not be an appropriate model of speciation in complex ecosystems.

Choosing Between Models

It has proven difficult to discriminate between models of population differentiation that focus on ecotypes or metapopulations. For example, the ecotypic structure of a soil *Bacillus* has been modeled to predict a priori which sequence clusters were ecotypes, and hence which ones should be associated with specific ecological properties (*16*). Some clusters are associated with certain phenotypic traits, such as a propensity to grow on shady north-facing slopes or sunny south-facing slopes. However, this model fitted no better (and in fact slightly worse) than a version of the model with several subpopulations and diversity generated only by neutral drift. This version of the model was dismissed because of its association with a very low estimate of population size (*14*). However, estimates of effective population size N_e are often grossly disconnected from census population sizes. It has proven very challenging to find models that successfully explain low estimated values of N_e while providing better predictions than models based on simple neutral drift. The analysis of *Bacillus* partly did this by predicting more ecotypes in the model than were observed using established ecological criteria, a hypothesis that can be tested.

This problem of low power to detect selection (or, more accurately, to reject neutrality) is a very general problem in population genetics that does not negate the importance of adaptation in evolution, but rather suggests that more work is needed if we want model-based methods to discriminate among different biologically plausible explanations of genetic data. In Table 1 we propose a scheme for performing analyses that could be used to test, develop, and validate different competing models more systematically.

Homologous Recombination

One specific challenge to models that invoke ecotypic structure involves a feature of bacterial evolution—homologous recombination—that we have not yet discussed. Bacterial reproduction does not involve the obligate reassortment of genetic material observed in most higher organisms. However, recombination does occur in bacteria and archaea (*29*) and typically involves the replacement of a short piece of DNA with the homologous segment from another strain. Recombination becomes less probable with increasing sequence divergence between the donor and the recipient (*30*, *31*), which reduces but does not eliminate recombination between closely related species. Because of such interspecies recombination, any given isolate within a species is almost certain to contain at least some genetic material that is characteristic of other closely related species. Hence, whereas it was once thought that bacteria do not form species in the eukaryotic sense because they do not recombine at all (*32*), one current view is that they do not form species because they recombine too much (*5*).

In asexual clonal organisms, even in the absence of any selective pressure, clusters will spontaneously split into multiple lineages or "daughter" clusters (*15*). However, under certain circumstances recombination can prevent this, and we can hence divide the bacteria into "sexual" and "nonsexual" species. This effect, described at greater length elsewhere (*15*), is summarized in Fig. 3, which shows the rate at which two clusters diverge over time—that is, the increase in the mean genetic distance between them. If this becomes negative, then the two clusters will stop diverging and instead converge. The three examples shown in Fig. 3 differ only in the rate of homologous recombination between the clusters, all other parameters being held constant. As recombination increases, we see a distinction between a "clonal" organism in which clusters are predicted to diverge (the green line) and a "sexual" organism (the blue line) in which they are predicted to be held together by recombination. For "sexual" species, the divergence of clusters requires a process that reduces the rate of recombination between them—for example, a period of allopatry or ecological differentiation. The speciation point is the amount of divergence between clusters that needs to accumulate to prevent them from returning to a single cluster if the barriers to recombination are removed. A recent study hypothesized that two related *Campylobacter* species are currently undergoing this process of merging into a single species as a result of changes in their environment (*33*).

The above insights were reached using models based on the assumption that genetic variation is neutral. Although this is obviously not always an appropriate assumption, it is plausible that the number of loci explicitly involved in adaptive ecological differentiation will be small, and thus that in an unstable landscape, genomic barriers to recombination will depend more on the accumulation of differences at neutral loci than at adaptive loci. The models also assumed a homogeneous distribution of polymorphisms across

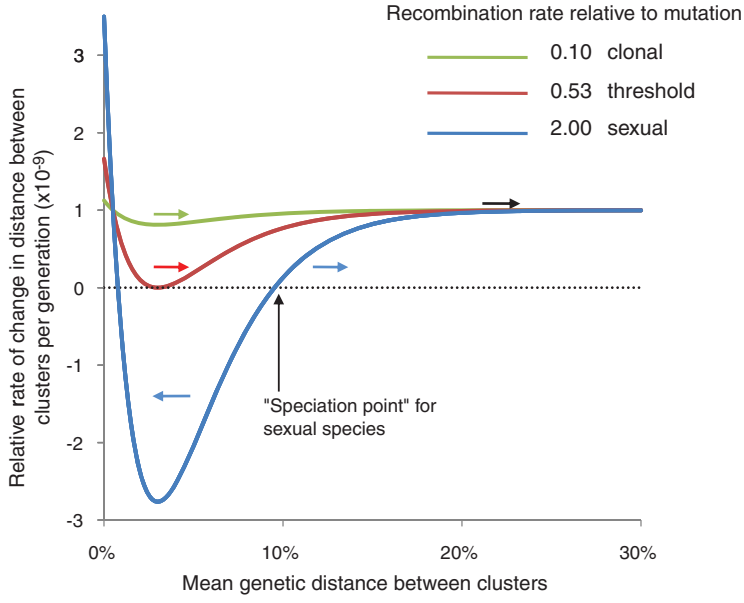


Fig. 3. The dynamics of cluster divergence. The figure summarizes some key results from (*15*) in a phase-space plot of the genetic dynamics of two populations, with recombination occurring between them at a rate that is varied for the three different simulations. The *y* axis shows the rate of change of genetic distance between the clusters as a function of the genetic distance itself (*x* axis). When the rate of change is positive, the populations will diverge genetically; when negative, they converge. The direction of change for each scenario is shown by arrows color-coded to each scenario. For low recombination rates, the populations are effectively clonal and always diverge (green line). As the recombination rate increases, the cohesive effects of recombination slow the rate of divergence, until a threshold is passed (red line) and the populations become effectively sexual in the sense that the populations no longer diverge. For recombination rates above this level, the fate of the two populations will depend on how genetically distinct they are at the outset. If they are within the "speciation point," then recombination will cause them to merge. If they are farther away than this "speciation point," they will continue to diverge from each other. These curves are derived using the model described in (*15*).

the genome, and violation of this may alter the tempo and mode of these processes (*34*, *35*).

Illegitimate Recombination and Gene Content Variation

Illegitimate recombination or gene acquisition is another unusual feature of bacteria. In this case, genes or clusters of genes are acquired that typically have no homolog(s) in the recipient strain. The importance of this phenomenon is evident in the clear and ubiquitous signature of such events in the growing body of genomic data. These are identified by differences in the characteristics of the acquired DNA and that of the host strain, for example, in base composition or codon usage; in most cases, the donor of the DNA in question is unknown. Gene acquisition leads to genomes being punctuated by stretches of foreign DNA. The largest of these (which may be many kilobases in length) were initially termed "pathogenicity islands," because the new functions encoded by the imports were often involved in virulence, but a better term is "genomic islands" as the phenomenon is far from limited to pathogens (*36*, *37*). Although it is hard to quantify the selective impact of importing any given gene(s) into a new background, the occasional ability to gain a new adaptation in this fashion—such as a new metabolic capability or a new mode of transmission for a pathogen—may be of enormous importance in terms of speciation.

Perhaps even more striking is the amount of variation in gene content revealed by multiple genomes from the same species, which implies that gene acquisition occurs at a surprisingly high frequency. It is now commonplace to speak of the "core" genome, which encodes fundamental functions shared by all members of a species (and, it should go without saying, other related species), onto which is bolted the "auxiliary" or "accessory" genome, composed of genes and operons that may or may not be present in all isolates. It seems likely that such auxiliary genes help to determine the specific ecological properties of the organism. For example, a group of related *Leptospirillum* has recently been hypothesized to adapt to different areas of an acid mine drainage system by shuffling of chromosome segments enriched in noncore genes (*38*, *39*). We should, however, be aware that changes in core genes may also lead to ecological differentiation, a phenomenon well documented in experimental studies of bacteria growing in structured environments (*40*).

Estimates vary, depending on the genomes that are available, but as little as 40% of genes may be present in all sequenced genomes of a named species (*41*). We may consider genes within a named species as being characteristic of different levels of ecological specificity, ranging from highly conserved core functions that are essential for growth in all environments to loci that are involved with adaptation to a specific habitat. Some narrow niche-specific genes may be distributed across species, being transferred between them by mobile elements. The evolutionary fate of such genes may hence be only loosely coupled with that of any particular species or strain in which they are found, and they are maintained through selection by the habitat to which each host strain is adapted. In the case of very mobile elements—for example, plasmids encoding resistance to antibiotics or heavy metals—the ecological specificity determined by these accessory loci may have no link to the sequence clusters we observe using housekeeping genes (Fig. 4).

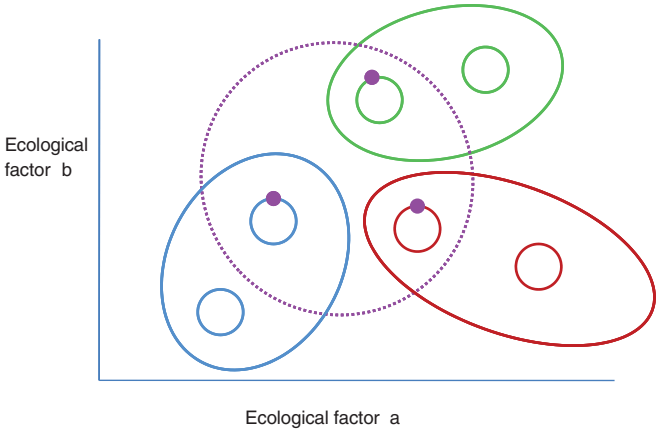


Fig. 4. Differences between core and auxiliary genes. This schematic illustrates the relationships between three species in "ecotype space," shown here in two dimensions, and a mobile gene common to all three. The areas occupied by the species are shown as solid lines in red, blue, and green. The part of the ecological space where the shared mobile gene is selected in each species is shown by a dashed purple line and overlaps all three species ranges. Examples of (circular) genomes from each species with and without the purple mobile element are also illustrated. Note that for each species, the locus is not selected for all isolates, and its evolutionary fate is uncoupled from that of each host species, because if one undergoes a selective sweep or goes extinct, the mobile gene may be reintroduced from one of the other species. Examples of such distributed loci include drug resistance determinants in pathogens (e.g., β-lactamase genes) and heavy metal resistance in environmental organisms. These genes may be transferred among strains and species by conjugative plasmids or other mobile elements (including transducing phage).

Identifying Mechanisms and Delineating Species

What do we want from bacterial species? Do we need theoretical consistency even at the expense of taxonomic practicality, incorporating both "clonal" and "sexual" populations into a single theoretical framework? One unifying theoretical concept is to consider species as the arena within which individuals are similar enough, or interbreed enough, that individual variant genes compete directly for reproductive success. Practical advances building on this or other theoretical concepts will only come when these are developed into explicit models and model-based algorithms that are tested and refined on a wide range of data. Alternatively, it may be sensible to suggest an ad hoc application of principles to different genera on the basis of their specific characteristics, including the extent of variation in gene content and recombination. In any case, no biologist would deny the importance of ecology to what we observe, but it may not be easy to incorporate it in a fashion that is convenient for taxonomists. Nonetheless, population geneticists may have little choice but to tackle the question of defining bacterial species or, at the very least, populations. Whether we are estimating effective population size from neutral diversity or choosing an appropriate set of strains to test for positive selection at a locus of interest, species definitions are implicit in much of the analytical toolkit of population genetics.

Distinguishing among mechanisms of population differentiation in bacteria ultimately comes down to testing the ability of different models to explain highly variable patterns within and between genetic-ecological clusters (Fig. 1). It is still unclear whether these patterns are maintained by gene flow or selection, and what the effect of population structure is. The joint distribution of genetic and ecological data can be used, as described above for *Vibrio* species (*13*), to define populations without making a strong theoretical commitment to either of these alternatives. One clear result from all of the studies discussed here is that the underlying theoretical questions concerning species will not be answered in the absence of more detailed genetic-environmental mapping. Moreover, some guidelines for the types of ecological studies that will be most informative are emerging. Most important, the ecological data collected must be relevant to the niche boundaries of the populations studied. And if genetic groups do not map exclusively onto sampling categories (as is likely to be the case), more complex statistical models will be needed to identify and describe the underlying niche structure. Longitudinal studies that measure the dynamics of ecological associations over time will also be helpful to determine how transient natural habitats are, and thus how likely bottlenecks are to result. Finally, whole-genome sequences from entire populations of environ-

Table 1. A proposed strategy for developing and validating models of bacterial evolution that might eventually be used to classify genetic diversity data and provide a firm foundation for a bacterial species concept.

1. Collect samples according to systematic ecological stratification. Focus on longitudinal studies, geographical studies, and measurement of physical and chemical gradients affecting bacterial growth. Consider biotic factors such as the presence of other competing bacteria or parasitic phage.
2. For each isolate, sequence as much as possible and affordable (16*S* rRNA, MLSA, auxiliary genes, full genomes, etc.).
3. Use empirical classification algorithms that use genetic and ecological data to jointly map isolates.
4. To guide model formulation, use population genetic tests on observed clusters, focusing on tests for selection, population structure, and gene flow.
5. Generate evolutionary models and simulate populations.
6. Test, then reject or adapt, evolutionary models according to agreement between simulations and real populations; if necessary, return to step 1.
7. For successful models, develop model-based methods for interpreting pure genetic data (without ecological covariates) and test on new data.
8. If one or more validated models emerge, use these to classify genetic data and to develop bacterial species concepts.

mental bacteria will be useful in dissecting the roles of the auxiliary and core genome in ecological differentiation. If after this process it emerges that some model or models are consistently validated for different study systems, these would inevitably form a good basis for identifying fundamental levels of clustering, or species.

In the foregoing we have emphasized ecotype and metapopulation models, but there are others that deserve consideration—notably the epidemic clonal model (*42*) and the impact of phage epidemics causing classic Lotka-Volterra boom-bust dynamics (*43*) illustrated in Fig. 2D—and it is possible, even likely, that more than one of these mechanisms may be relevant to any given problem in speciation and cluster formation. Distinguishing among these mechanisms is the bacterial species challenge (Table 1), described in 1991 by John Maynard Smith as follows: "Ecotypic structure, hitch-hiking, and localized recombination can explain the observed patterns of variation. The difficulty, of course, is that the model is sufficiently flexible to explain almost anything. To test the hypothesis of ecotypic structure, we need to know the distribution of electrophoretic types [i.e., genotypes] in different habitats" (*17*).

Much research on bacterial species to date has come from studies on pathogens, where the correct identification of species is crucial for accurate clinical diagnoses. However, for pathogens the identification of the multiple ecological niches within (for example) the nasopharynx or gut is difficult, and studies of the relationships between bacterial populations and ecology may be more fruitful for some environmental species where the categorization of niches is a more tractable enterprise. Hopefully, we will soon obtain richer data sets that map bacterial diversity onto ecology and provide a way to distinguish among various models of population differentiation and speciation, including those based on ecotypes or metapopulations.

References and Notes

1. J. Handelsman, *Microbiol. Mol. Biol. Rev.* **68**, 669 (2004).
2. W. P. Hanage, C. Fraser, B. G. Spratt, *Philos. Trans. R. Soc. London Ser. B* **361**, 1917 (2006).
3. M. Achtman, M. Wagner, *Nat. Rev. Microbiol.* **6**, 431 (2008).
4. F. M. Cohan, *Annu. Rev. Microbiol.* **56**, 457 (2002).
5. W. F. Doolittle, R. T. Papke, *Genome Biol.* **7**, 116 (2006).
6. D. Gevers *et al.*, *Nat. Rev. Microbiol.* **3**, 733 (2005).
7. M. F. Polz, D. E. Hunt, S. P. Preheim, D. M. Weinreich, *Philos. Trans. R. Soc. London Ser. B* **361**, 2009 (2006).
8. D. B. Rusch *et al.*, *PLoS Biol.* **5**, e77 (2007).
9. C. Darwin, *The Origin of Species* (Penguin Classics, London, 1985).
10. E. Stackebrandt *et al.*, *Int. J. Syst. Evol. Microbiol.* **52**, 1043 (2002).
11. R. Facklam, *Clin. Microbiol. Rev.* **15**, 613 (2002).
12. J. C. Arbique *et al.*, *J. Clin. Microbiol.* **42**, 4686 (2004).
13. D. E. Hunt *et al.*, *Science* **320**, 1081 (2008).
14. See supplementary information in (*16*) for a statistical comparison of the ecotype model with an effective neutral model and an implicit estimate of N_e.
15. C. Fraser, W. P. Hanage, B. G. Spratt, *Science* **315**, 476 (2007).
16. A. Koeppel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2504 (2008).
17. J. Maynard Smith, *Proc. R. Soc. London Ser. B* **245**, 37 (1991).
18. H. Ochman, A. C. Wilson, in Escherichia coli *and* Salmonella typhimurium*: Cellular and Molecular Biology*, F. C. Neidhart, Ed. (ASM Press, Washington, DC, 1987), pp. 1649–1654.
19. J. R. Thompson *et al.*, *Appl. Environ. Microbiol.* **70**, 4103 (2004).
20. M. Kimura, *Trends Biochem. Sci.* **1**, N152 (1976).
21. R. Milkman, *Trends Biochem. Sci.* **1**, N152 (1976).
22. K. C. Atwood, L. K. Schneider, F. J. Ryan, *Proc. Natl. Acad. Sci. U.S.A.* **37**, 146 (1951).
23. B. R. Levin, *Genetics* **99**, 1 (1981).
24. F. M. Cohan, E. B. Perry, *Curr. Biol.* **17**, R373 (2007).
25. A. Rambaut *et al.*, *Nature* **453**, 615 (2008).
26. R. A. Fisher, E. B. Ford, *Heredity* **1**, 143 (1947).
27. M. Slatkin, *Theor. Popul. Biol.* **12**, 253 (1977).
28. J. Majewski, F. M. Cohan, *Genetics* **152**, 1459 (1999).
29. E. J. Feil, B. G. Spratt, *Annu. Rev. Microbiol.* **55**, 561 (2001).
30. J. Majewski, F. M. Cohan, *Genetics* **153**, 1525 (1999).
31. J. Majewski, P. Zawadzki, P. Pickerill, F. M. Cohan, C. G. Dowson, *J. Bacteriol.* **182**, 1016 (2000).
32. S. T. Cowan, in *Microbial Classification, 12th Symposium of the Society for General Microbiology*, G. C. Ainsworth, P. H. A. Sneath, Eds. (Cambridge Univ. Press, Cambridge, 1962), pp. 433–455.
33. S. K. Sheppard, N. D. McCarthy, D. Falush, M. C. J. Maiden, *Science* **320**, 237 (2008).
34. A. C. Retchless, J. G. Lawrence, *Science* **317**, 1093 (2007).
35. K. Vetsigian, N. Goldenfeld, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7332 (2005).
36. A. Tuanyok *et al.*, *BMC Genomics* **9**, 566 (2008).
37. U. Dobrindt, B. Hochhut, U. Hentschel, J. Hacker, *Nat. Rev. Microbiol.* **2**, 414 (2004).
38. P. Wilmes, S. L. Simmons, V. J. Denef, J. F. Banfield, *FEMS Microbiol. Rev.* **33**, 109 (2009).
39. V. J. Denef *et al.*, *Environ. Microbiol.* **11**, 313 (2008).
40. E. Bantinaki *et al.*, *Genetics* **176**, 441 (2007).
41. R. A. Welch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 17020 (2002).
42. J. Maynard Smith, N. H. Smith, M. O'Rourke, B. G. Spratt, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4384 (1993).
43. K. H. Hoffmann *et al.*, *FEMS Microbiol. Lett.* **273**, 224 (2007).

44. We thank T. Connor and S. Deeny for useful discussions. Supported by University Research Fellowships from the Royal Society (C.F. and W.P.H.), a program grant from the Wellcome Trust (B.G.S.), grants from the U.S. Department of Energy Genomes to Life program (M.F.P. and E.J.A.), and the NSF/National Institute of Environmental Health Sciences Woods Hole Centre for Oceans and Human Health, the NSF Biological Oceanography Program, and the Moore Foundation (M.F.P.).

10.1126/science.1159388

REVIEW

Is Genetic Evolution Predictable?

David L. Stern[1]* and Virginie Orgogozo[2]*

Ever since the integration of Mendelian genetics into evolutionary biology in the early 20th century, evolutionary geneticists have for the most part treated genes and mutations as generic entities. However, recent observations indicate that all genes are not equal in the eyes of evolution. Evolutionarily relevant mutations tend to accumulate in hotspot genes and at specific positions within genes. Genetic evolution is constrained by gene function, the structure of genetic networks, and population biology. The genetic basis of evolution may be predictable to some extent, and further understanding of this predictability requires incorporation of the specific functions and characteristics of genes into evolutionary theory.

One hundred and fifty years ago, Charles Darwin and Alfred Russell Wallace proposed that biological diversity results from natural selection acting on heritable variation in populations. Both Darwin and Wallace recognized the importance of heritable variation to evolutionary theory, but neither man knew the true cause of inheritance. Early in the 20th cen-

tury, the rediscovery of Mendel's studies allowed for a formal mathematical treatment of alleles in populations, generating the field of population genetics. Population geneticists treated genes and alleles as generic entities, particles that were inherited and somehow caused variation in the appearance, behavior, and physiology of organisms—what we call collectively the phenotype. This level of abstraction was appropriate given that a molecular understanding of gene function lay many decades in the future. Even with this rudimentary view of gene function, however, population genetics greatly clarified how real populations evolve, and this theoretical understanding spurred the New Synthesis, combining population genetics with ecology, systematics, and biogeography to explain and explore many questions in evolution.

In the past 40 years, molecular biologists have elucidated how genes regulate biological processes, but only the most basic mechanistic observations have been integrated into evolutionary biology. For example, evolutionary theory has effectively absorbed the distinction between coding (nonsynonymous) and silent (synonymous) substitutions in protein-coding regions, but other aspects of molecular biology currently contribute little to evolutionary thought. The time has now come to integrate the specifics of molecular and developmental biology into evolutionary biology. Over the past 15 years, many examples of the genes and mutations causing evolutionary change have been identified (*1*). Patterns in these data suggest that a synthesis of molecular developmental biology with evolutionary theory will reveal new general principles of genetic evolution.

Nonrandom Distribution of Evolutionarily Relevant Mutations

Recent studies suggest that the mutations contributing to phenotypic variation [evolutionarily relevant mutations (*2*)] are not distributed randomly across all genetic regions. The most compelling evidence comes from cases of parallel genetic evolution: the independent evolution of similar phenotypic changes in different species due to changes in homologous genes or sometimes in the same amino acid position of homologous genes.

Many cases of parallel evolution have been discovered across all of the kingdoms. At least 20 separate populations of the plant *Arabidopsis thaliana* have evolved null coding mutations (mutations that completely eliminate protein function) in the *Frigida* gene that cause early flowering (*3*). Resistance to DDT and pyrethroids has evolved in 11 insect species by mutations in either amino acid Leu^{1014} or Thr^{929} of the voltage-gated sodium channel gene *para* (*4*). Two virus populations independently subjected to experimental evolution in a novel host accumulated many of the same amino acid mutations (*5*). In total, about 350 evolutionarily relevant mutations have been found in plants and animals, and more than half of these represent cases of parallel genetic evolution (*1*).

One explanation for parallel genetic evolution is that most genes play specialized roles during development, and only some genes can evolve to generate particular phenotypic variants. For example, mutations in *rhodopsin* can alter light-wavelength sensitivity (*6*), and mutations in *lysozyme* may enhance enzyme activity at the particular pH of a fermenting gut (*7*). But the reverse would not be true. Mutations in *rhodopsin* are unlikely to enhance fermentation, and mutations in a digestive enzyme will not aid detection of a particular wavelength of light, even if each protein was expressed in the reciprocal organ.

Gene function explains part but not all of the observed pattern of parallel genetic evolution. In several cases, parallelism has been observed even though mutations in a large number of genes can produce similar phenotypic changes. For example, although more than 80 genes regulate flowering time (*8*), changes in only a subset of these genes have produced evolutionary changes in flowering time (*3*). Hundreds of genes regulate the pattern of fine epidermal projections, called trichomes, on *Drosophila melanogaster* larvae. But only one gene, called *shavenbaby*, has evolved to alter larval trichome patterns between *Drosophila* species, and this gene has accumulated multiple evolutionarily relevant mutations (*9*). What is special about these hotspot genes?

Developmental biology illuminates why hotspot genes such as *shavenbaby* exist. During development, multiple cell-signaling pathways and transcription factors act together to progressively divide the embryo into a virtual map that specifies when and where organs will form. The interactions between the genes encoding these signaling molecules and transcription factors can be represented as a genetic network. Gene interactions are modulated in large part by the cis-regulatory regions of patterning genes. (All genes are composed of two fundamentally different regions: a region encoding the gene product—a protein or an RNA—and adjacent *cis*-regulatory DNA that encodes the instructions governing when and where the gene product will be produced.) Transcription factors bind to *cis*-regulatory regions of target genes, and the summed effect of many such interactions at a target gene determines whether the gene is expressed or not. Patterning genes act within complex genetic networks, and usually each patterning gene contributes to the development of multiple cell types. For example, most patterning genes that are active during embryonic development of the epidermis contribute to the development of muscle-attachment sites, sensory organs, tracheal pits, trichomes, or other cell types.

The importance of regulatory networks in determining which genes may be evolutionary hotspots can be illustrated with the genetic network that governs larval trichome development in *D. melanogaster* (Fig. 1). In this network, developmental patterning genes first collaborate to divide the embryonic epidermis into domains expressing distinct transcription factors. These patterning genes then regulate the expression of the *shavenbaby* gene, a so-called input-output gene (*10*). Input-output genes integrate complex spatiotemporal information (the input) and trigger development of an entire program of cell differentiation (the output). The Shavenbaby protein activates expression of a battery of target genes that transform an epidermal cell into a trichome cell. Each target gene triggers a specific aspect of cell differentiation, and production of a differentiated trichome requires coordinated expression of all target genes. The pattern of trichomes over the body is thus determined by the distribution of Shavenbaby protein in the epidermis, which is controlled by the *cis*-regulatory region of the *shavenbaby* gene. The *shavenbaby* gene serves as a nexus for patterning information flowing in and for cell-fate information flowing out.

In the entire regulatory network governing development of the *Drosophila* embryo, only *shavenbaby*, with its specialized function to rally the entire module of trichome morphogenesis, can accumulate mutations that alter trichome patterns without disrupting other developmental processes. Genetic changes in upstream developmental genes will alter trichome production, but these mutations also disrupt other organs. Changes in any one of the downstream genes are not sufficient to create or eliminate a trichome; concerted changes in multiple downstream genes are required to build a trichome (*11*). Furthermore, all of the evolutionarily relevant mutations in *shavenbaby* that have been identified so far alter the *cis*-regulatory region and not the protein-coding region. Mutations in the protein-coding region would alter *shavenbaby* function in every cell that accumulates Shavenbaby protein, and this would alter every trichome produced in larvae and adults. Thus, a developmental perspective clarifies why *shavenbaby* is a hotspot for evolutionarily relevant mutations and why these mutations occur in the *cis*-regulatory region of the gene. We predict that the *cis*-regulatory regions of other input-output genes may be hotspots for other phenotypic characteristics.

The *shavenbaby* gene provides one example of a more general principle: that mutations af-

[1]Department of Ecology and Evolutionary Biology, Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544, USA. [2]CNRS, Université Pierre et Marie Curie, Bâtiment A, 5ème Étage, Case 29, 7 Quai Saint Bernard, 75005 Paris, France.

*To whom correspondence should be addressed. E-mail: dstern@princeton.edu (D.L.S.); virginie.orgogozo@normalesup.org (V.O.)

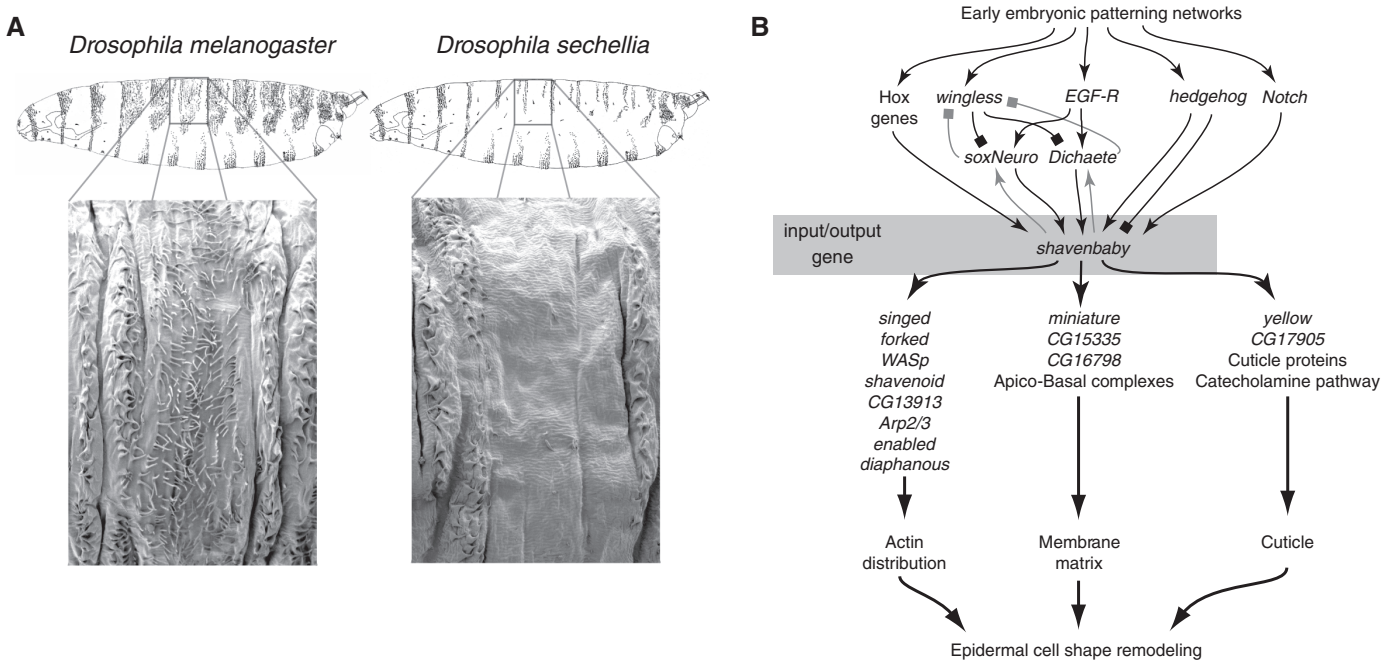


Fig. 1. Morphological divergence between species has been caused by repeated evolution at an input-output gene. (**A**) *D. melanogaster* and *D. sechellia* differ in the pattern of fine trichomes decorating the dorsal and lateral surfaces of the larvae. This difference is caused entirely by evolution of the *cis*-regulatory region of the *shavenbaby* gene (*9*). (**B**) The *cis*-regulatory region of the *shavenbaby* gene integrates extensive information from developmental patterning genes to generate a pattern of Shavenbaby protein expression that prefigures the pattern of trichomes on the first-instar larva. Cells accumulating Shavenbaby will differentiate a trichome because Shavenbaby protein regulates a large battery of genes that act together to transform an epithelial cell into a trichome (*11*).

fecting multiple phenotypic traits, so-called pleiotropic mutations, are unlikely to contribute to adaptive evolution. As we discuss next, pleiotropy and other genetic and population-genetic parameters seem to influence the distribution of evolutionarily relevant mutations.

The Factors Influencing the Distribution of Evolutionarily Relevant Mutations

Pleiotropy. Some mutations generate specific phenotypic changes, whereas pleiotropic mutations alter several seemingly unrelated traits. Two mutations that cause evolutionary increases in the number of thoracic bristles in *Drosophila* illustrate the difference between mutations with specific and pleiotropic effects (Fig. 2). A *cis*-regulatory change in the *scute* gene affects the number of sensory organs only on the thorax (*12*), whereas a coding mutation in the *poils au dos* gene increases the number of sensory organs on both the thorax and the wing (*13*). The *poils au dos* mutation is more pleiotropic than the *scute* mutation. *Scute*, like *shavenbaby*, is an input-output gene, whereas *poils au dos* is a patterning gene that, together with other pattering genes, regulates *scute* expression (Fig. 2). Mutations with pleiotropic effects will rarely change all phenotypic traits in a favorable way, and experimental evidence indicates that pleiotropic effects tend to reduce fitness (*14*). Selection may favor extra bristles on the thorax, but not extrasensory organs on the wing. Even if one effect of a pleiotropic mutation provides a major improvement in fitness, the other effects may be deleterious and will reduce the likelihood that the mutation will become established in the population (*15*).

Epistasis. When examined in a single genetic background, a mutation may have a specific or a pleiotropic effect. But in another genetic background, the same mutation may produce a different phenotypic effect because of nonadditive interactions of alleles: so-called epistasis. For example, one allele in *A. thaliana* increases growth in one genetic background but reduces growth in a different genetic background (*16*). The second genetic background is not simply deleterious in general because a variant allele at a second locus causes higher growth in this background. Thus, the effects of one mutation can depend on the genetic variation present at other loci.

Epistasis is extremely common in natural populations and it may sometimes reduce the rate of evolution (*17*). Epistasis increases the phenotypic variance associated with a particular mutation, causing a mutation to have a fluctuating fitness effect dependent on the genetic background. Thus, in an *Arabidopsis* population containing multiple genetic backgrounds, we expect that selection for increased size will tend to favor nonepistatic alleles that increase growth in all backgrounds rather than epistatic alleles that increase growth in only one genetic background.

Plasticity. Populations exposed to repeated environmental changes may evolve genetic mechanisms that produce different phenotypes suited to different environmental conditions: so-called phenotypic plasticity. For example, aphids can produce multiple phenotypic forms in response to environmental conditions, including asexual forms that reproduce quickly and sexual forms that lay overwintering eggs. Mutations that eliminate sexual forms—that reduce plasticity—may provide a lineage with a short-term advantage, a much faster reproductive rate. But in the long term, aphid lineages that do not produce sexual forms tend to go extinct, perhaps because they fail to adapt to changing environmental conditions.

Similarly, in *A. thaliana* the *Frigida* gene controls plasticity for flowering time. *Frigida* responds to cold temperatures to induce flowering. In regions with warm winters, null *Frigida* mutations may provide a short-term advantage by consistently triggering flowering, even in the absence of a cold winter. But these mutations eliminate plasticity for flowering time, possibly preventing these plants from adapting to colder temperatures or from recolonizing areas in colder climates. Thus, the abundance of null *Frigida* mutations in *Arabidopsis* populations must result from factors that override the negative consequences of reduced plasticity.

Strength of selection. When an environmental change favors a phenotype that is vastly different

from the mean phenotype in a population, mutations causing large phenotypic changes toward the new optimum will be favored, at least initially (*18*). For example, recently domesticated races have probably experienced strong selection by farmers, and many recent domestication traits result from mutations that cause large phenotypic effects, including pleiotropic deleterious effects. As an example, six different null-coding mutations in the *myostatin* gene cause muscle hypertrophy in different breeds of cattle (*19*). *Myostatin* is a member of the *transforming growth factor*–β superfamily of growth factors and acts as a negative regulator of muscle development. Although null mutations of *myostatin* generate cattle with more and leaner meat, these cattle experience difficulties in calving and have reduced stress tolerance. Strong selection during domestication can obviously overcome the negative pleiotropic effects of null *myostatin* mutations.

Population history. The past and current sizes of a population also influence genetic evolution. Small population size increases the effects of random sampling of alleles, so-called genetic drift. In small populations, genetic drift will allow deleterious alleles to occasionally increase in frequency. For example, a small inbred population of Bedouins in Israel has evolved a high frequency of a recessive allele that causes deafness (*20*). With stronger genetic drift in small populations, natural selection will fail to promote the spread of adaptive mutations of small effect. Instead, in comparison with large populations, adaptive mutations of relatively large effect will tend to evolve by natural selection in small populations.

Small populations also have another critical effect on evolution: They limit the total number of new mutations introduced into the population each generation. Thus, small populations may end up selecting far-from-ideal mutations (those with pleiotropic consequences and epistatic effects) simply because potentially superior mutations occur at a lower rate.

The abundance of null *Frigida* mutations in populations of *A. thaliana* highlights the importance of population history in genetic evolution. Null *Frigida* mutations have the negative consequence of reducing

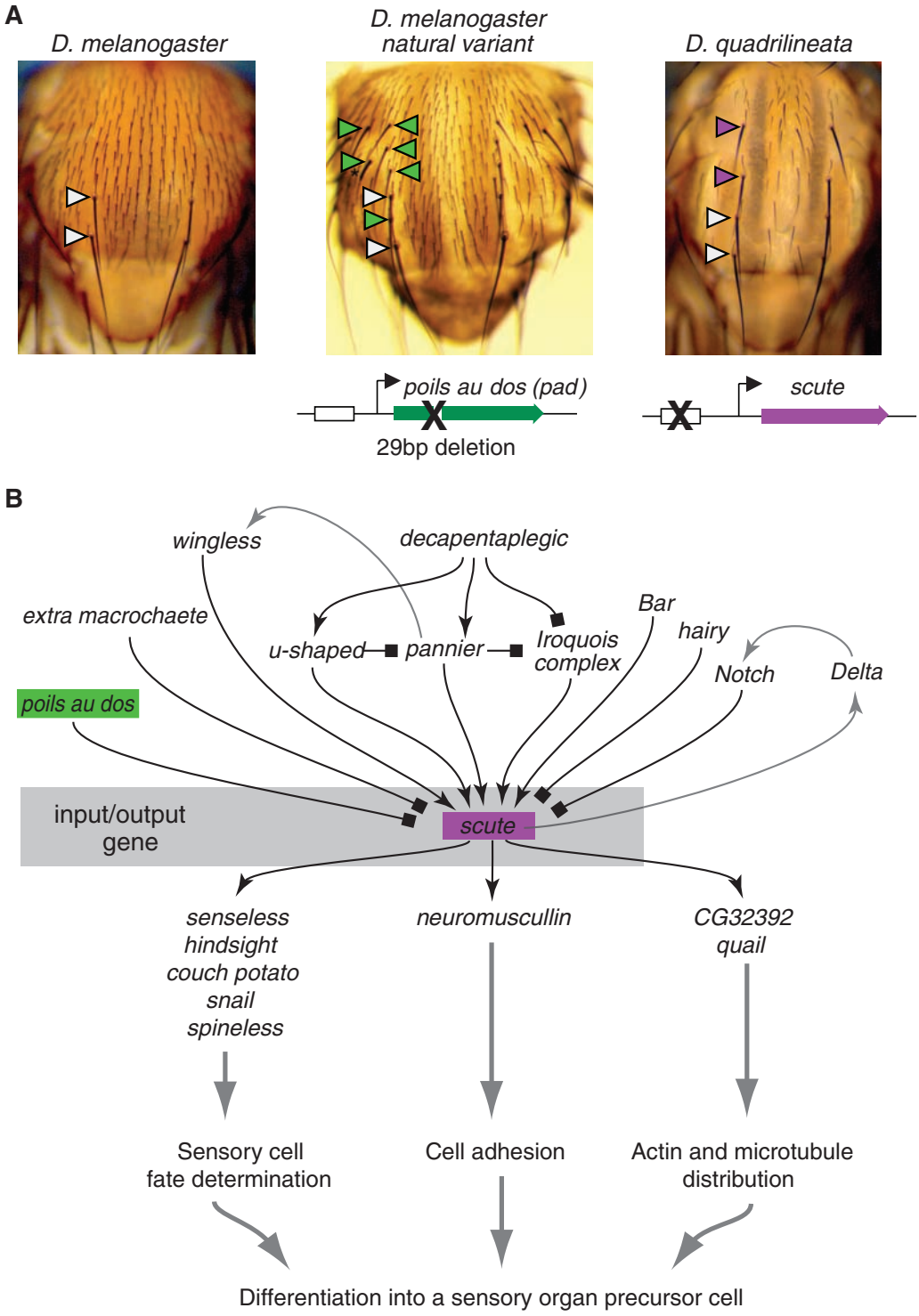


Fig. 2. Bristle patterns on the dorsal thorax of *Drosophila* species have evolved within species and between species because of different kinds of mutations. (**A**) A mutation generating a null allele of the *poils au dos* gene within a population of *D. melanogaster* increases the number of large bristles on the dorsal thorax (white triangles indicate normal bristles and green triangles indicate extra bristles) (*13*). In contrast, the increased number of bristles in *D. quadrilineata* results at least in part from changes in the *cis*-regulatory region of the *scute* gene (*12*). The extra bristles caused by the *poils au dos* mutation are not as precisely positioned as the extra bristles caused by the *scute* mutation (indicated by purple triangles). (**B**) The two evolving genes, *poils au dos* and *scute*, occupy different locations in the genetic network that generates the pattern of bristles. The *scute* gene is an input-output gene, whereas the *poils au dos* gene is a developmental patterning gene. The null mutation in *poils au dos* increases sensory organ numbers not only in the thorax but also in the wing.

plasticity for flowering time. These mutations also have pleiotropic effects [they reduce fruit production (*21*)] and display epistasis with respect to other genes that control flowering time (*22*). These observations suggest that null *Frigida* mutations are not ideal alleles for controlling flowering time. In fact, null *Frigida* mutations must only rarely, if ever, be involved in phenotypic divergence between species because homologs of the *Frigida* gene exist in diverse plant species. But natural selection has overcome the deleterious effects of null *Frigida* mutations to promote the spread of these mutations in small populations. *A. thaliana* has migrated from Scandinavia around the world in the footsteps of agriculture. Subpopulations have adapted to local conditions, including the relatively warm and short winters of more temperate regions. *A. thaliana* plants are self-fertile, so even a single plant can give rise to a new population. These small subpopulations provide fewer opportunities for beneficial mutations of specific effect to appear, and strong selection for rapid flowering has favored whatever mutations of strong effect arose in the population, such as null *Frigida* mutations. The abundance of null *Frigida* mutations probably reflects the fact that these mutations occur at a higher rate than mutations without associated deleterious consequences.

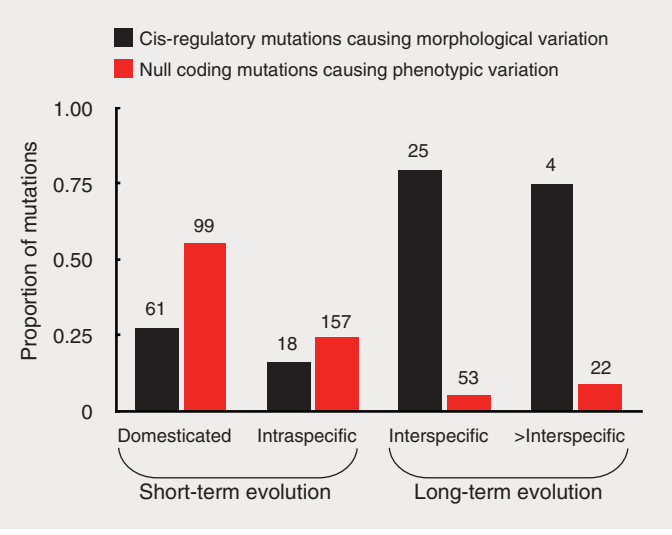


Fig. 3. Different kinds of mutations occur with different frequency during short-term and long-term evolution. Among all mutations causing morphological variation identified to date, the proportion of *cis*-regulatory mutations (black bars) is higher for long-term evolution than for short-term evolution. For all mutations that have been reported to cause phenotypic variation in either morphology or physiology, the proportion caused by null coding mutations (red bars) is higher for short-term evolution than for long-term evolution. The numbers above the bars refer to the total number of examples in each category. The number of cases of morphological evolution (black bars) is a subset of the number of cases of phenotypic evolution (red bars). Data are from (*1*).

The Genetic Basis of Short-Term and Long-Term Evolution

The *Frigida* example is not unique. In many plants and animals, evolution over long periods (variation between species) appears to differ in several ways from evolution over shorter periods (variation between domesticated races and between individuals within a species) (*1*). Here are three general ways in which long-term and short-term genetic evolution differ.

First, epistasis is commonly found for the mutations that contribute to phenotypic variation within species, whereas it is rarely observed for the mutations that cause differences between species. Within *D. melanogaster*, variation in bristle number is caused by multiple loci of relatively small effect, and these loci have epistatic effects of the same order of magnitude as the additive effects (*23*, *24*). In contrast, morphological differences between *Drosophila* species result from multiple loci of intermediate-to-small effect that only rarely show epistasis (*25*, *26*). Studies of body size variation in chickens show a similar pattern, with alleles segregating within species showing more epistasis than alleles differentiating species (*27*, *28*).

Second, null mutations, which arise frequently and often cause pleiotropic and epistatic effects, seem to contribute more to phenotypic variation within species than to phenotypic differences between species. About 55% of the 99 mutations known to cause domestication traits are null-coding mutations, whereas only 7% of the 75 mutations known to cause interspecific differences are null-coding mutations (Fig. 3). For example, although domesticated cattle stocks have evolved multiple null mutations of the *myostatin* gene, all mammal species investigated so far possess a functional *myostatin* gene.

Third, the frequency of *cis*-regulatory mutations causing morphological variation differs between taxonomic levels. Morphological changes may occur either through coding changes or through *cis*-regulatory changes (Fig. 2). Because mutations in *cis*-regulatory regions often have fewer pleiotropic effects than mutations in coding regions, morphological changes are expected to involve mainly *cis*-regulatory mutations (*1*, *2*, *29*). Within species, most mutations that cause morphological variation have been found in protein-coding regions (Fig. 3). In contrast, between species most mutations that cause morphological differences have been found in *cis*-regulatory regions. Presumably, many of the coding mutations found within species fail to spread through populations, perhaps because of pleiotropic deleterious effects.

These striking and unexpected differences between short-term and long-term genetic evolution have emerged only recently with the accumulation of a sufficient number of case studies. These patterns are consistent with theoretical expectations of how the five parameters discussed earlier (pleiotropy, epistasis, plasticity, strength of selection, and population structure) should influence genetic evolution. Evolution over long periods, reflected in the differences between species, should result from mutations relatively devoid of pleiotropic and epistatic effects. In contrast, evolution over shorter periods, reflected in the differences between domesticated races and in the variation segregating within species, may often result from mutations that disrupt plasticity or that have pleiotropic and epistatic effects. In summary, differences between species are caused by a biased subset of the mutations that have appeared within natural populations (*1*).

Conclusions

Although mutations are thought to occur randomly in the genome, the distribution of mutations that cause biological diversity appears to be highly nonrandom. Gene function, gene structure, and the roles of genes and gene products in genetic networks all influence whether particular mutations will contribute to phenotypic evolution. Thus, for some phenotypic changes, evolutionarily relevant mutations are expected to accumulate in a few hot-spot genes and even in particular regions within single genes. In addition, population biology and ecology influence the spectrum of evolutionarily relevant mutations. Over short periods, adaptive mutations with deleterious pleiotropic effects may be selected because mutations without deleterious effects have not yet appeared. In contrast, over long periods adaptive mutations without pleiotropic deleterious effects have more opportunity to arise and be selected.

The genetic basis of phenotypic evolution thus appears to be somewhat predictable. These emerging patterns in the distribution of mutations causing phenotypic diversity derive, however, from a limited set of data culled from the published literature. It is possible that these patterns reflect biases in the way scientists have searched for evolutionarily relevant mutations (*1*). For

example, many researchers focus on candidate genes, which precludes the discovery of previously unknown genes. In the future, we expect that widespread adoption of unbiased experimental approaches, for example genetic mapping, will provide data for robust tests of the predictability of genetic evolution. Genetic mapping can be performed within species and, in rare cases, between closely related species. Alternatively, gene-by-gene replacement of all genes from one species into a second species, although experimentally tedious, may allow unbiased surveys for species that cannot be crossed. This approach would allow comparisons of distantly related taxa and provide a direct test of whether distantly related taxa have accumulated different kinds of evolutionarily relevant mutations than have closely related species.

More precise quantitative predictions about the mutations responsible for phenotypic evolution will probably result from further synthesis of molecular biology and population genetics. New theoretical models will encompass multiple population-genetic parameters within a genomic and developmental framework. These models may provide insight into how the distribution of spontaneously arising mutations is translated into the distribution of mutations segregating within populations and how these two distributions impact short-term and long-term evolution.

Finally, the fact that long-term genetic evolution may represent a biased subset of mutations has applied consequences, from the development of more efficient computer algorithms that utilize evolutionary search strategies to the improvement of agricultural crops and animals. Domestication often selects for mutations that have pleiotropic deleterious effects. Long-term evolution, in contrast, selects for mutations with specific phenotypic effects, and this class of mutations might be exploited to engineer domesticated races that possess desirable characteristics without associated unfavorable properties.

References and Notes

1. D. L. Stern, V. Orgogozo, *Evolution* **62**, 2155 (2008).
2. D. L. Stern, *Evolution* **54**, 1079 (2000).
3. C. Shindo *et al.*, *Plant Physiol.* **138**, 1163 (2005).
4. R. H. ffrench-Constant, B. Pittendrigh, A. Vaughan, N. Anthony, *Philos. Trans. R. Soc. B* **353**, 1685 (1998).
5. H. A. Wichman, M. R. Badgett, L. A. Scott, C. M. Boulianne, J. J. Bull, *Science* **285**, 422 (1999).
6. S. Yokoyama, *Annu. Rev. Genet.* **31**, 315 (1997).
7. C.-B. Stewart, A. C. Wilson, *Cold Spring Harb. Symp. Quant. Biol.* **52**, 891 (1987).
8. Y. Y. Levy, C. Dean, *Plant Cell* **10**, 1973 (1998).
9. A. P. McGregor *et al.*, *Nature* **448**, 587 (2007).
10. E. H. Davidson *et al.*, *Science* **295**, 1669 (2002).
11. H. Chanut-Delalande, I. Fernandes, F. Roch, F. Payre, S. Plaza, *PLoS Biol.* **4**, e290 (2006).
12. S. Marcellini, P. Simpson, *PLoS Biol.* **4**, e386 (2006).
13. J. M. Gibert, S. Marcellini, J. R. David, C. Schlotterer, P. Simpson, *Dev. Biol.* **288**, 194 (2005).
14. T. F. Cooper, E. A. Ostrowski, M. Travisano, *Evolution* **61**, 1495 (2007).
15. S. P. Otto, *Proc. Biol. Sci.* **271**, 705 (2004).
16. J. Kroymann, T. Mitchell-Olds, *Nature* **435**, 95 (2005).
17. R. Yukilevich, J. Lachance, F. Aoki, J. R. True, *Evolution* **62**, 2215 (2008).
18. H. A. Orr, *Evolution* **52**, 935 (1998).
19. R. H. Bellinge, D. A. Liberles, S. P. Iaschi, P. A. O'Brien, G. K. Tay, *Anim. Genet.* **36**, 1 (2005).
20. D. A. Scott *et al.*, *Am. J. Hum. Genet.* **57**, 965 (1995).
21. N. Scarcelli, J. M. Cheverud, B. A. Schaal, P. X. Kover, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16986 (2007).
22. T. M. Korves *et al.*, *Am. Nat.* **169**, E141 (2007).
23. A. D. Long *et al.*, *Genetics* **139**, 1273 (1995).
24. M. C. Gurganus, S. V. Nuzhdin, J. W. Leips, T. F. C. Mackay, *Genetics* **152**, 1585 (1999).
25. Z.-B. Zeng *et al.*, *Genetics* **154**, 299 (2000).
26. S. J. Macdonald, D. B. Goldstein, *Genetics* **153**, 1683 (1999).
27. L. Jacobsson *et al.*, *Genet. Res.* **86**, 115 (2005).
28. S. Kerje *et al.*, *Anim. Genet.* **34**, 264 (2003).
29. S. B. Carroll, *Cell* **134**, 25 (2008).

30. We thank two referees for helpful comments, P. Simpson for providing photographs and advice for Fig. 2, and CNRS for research funding for O.V.

10.1126/science.1158997

The Sea-Level Fingerprint of West Antarctic Collapse

Jerry X. Mitrovica,[1] Natalya Gomez,[1] Peter U. Clark[2]

There is widespread concern that the West Antarctic Ice Sheet (WAIS), which is characterized by extensive marine-based sectors (*1*), may be prone to collapse in a warming world. The recent Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) (*2*) estimated that this collapse would lead to a sea-level rise of ~5 m. This estimate is derived by converting the total volume of all grounded portions of WAIS into water, filling in any topographic holes associated with marine-based sectors, and spreading the remaining water uniformly (i.e., eustatically) across the oceans (*3*). Although the appropriate effective eustatic value (EEV) for WAIS collapse is uncertain (e.g., will non–marine-based sectors vanish?), we show that, whatever the value, sea-level changes at some coastal sites will be significantly higher (or, less commonly, lower) than the EEV.

The rapid melting of ice sheets and glaciers leads to a sea-level change that departs dramatically from the assumption of a uniform redistribution of meltwater (*4*). An ice sheet exerts a gravitational attraction on the nearby ocean and thus draws water toward it. If the ice sheet melts, this attraction will be reduced, and water will migrate away from the ice sheet. The net effect, despite the increase in the total volume of the oceans after a melting event, is that sea level will actually fall within ~2000 km of the collapsing ice sheet and progressively increase as one moves further from this region. Each ice reservoir will produce a distinct geometry, or fingerprint, of sea-level change.

Although the physics of fingerprinting has been embraced in studies of past sea-level change, it has been largely ignored in discussions of future projections. A widely neglected exception is an analysis by Clark and Lingle (*5*), who were concerned with sea-level changes after a "uniform thinning" of the WAIS. Their fingerprint calculation had its basis in a standard sea-level theory that accounts for load self-attraction (as described above) and the associated elastic deformation of the solid Earth. Their results, reproduced in Fig. 1A, show a zone of sea-level fall close to the WAIS and a maximum rise in the North Pacific. Coastal sites well away from the WAIS have peak values ~5 to 10% higher than the EEV.

The sea-level theory adopted by Clark and Lingle does not allow for shoreline migration, including the inundation and adjustment of regions vacated by grounded, marine-based ice cover, or any feedback onto sea level of Earth rotation changes. We show (Fig. 1B) a projection based on a sea-level theory (*6*) that overcomes these limitations. These results show a highly accentuated sea-level rise in the oceans bordering North America and in the Indian Ocean. Coastal sites in North America would experience a rise ~30% higher than the EEV.

The difference between this and the standard (Clark and Lingle) calculations is shown in Fig. 1C. The far-field geometry of the differential sea-level signal, which includes a roughly uniform rise in combination with a quadrantial form, is diagnostic of the physical mechanisms responsible for the accentuated signal (*1*). In particular, the uniform rise is due to the expulsion of water from the West Antarctic as flooded, marine-based sectors of this region uplift (elastically) in response to the unloading (fig. S1B). The dominant quadrantial signal arises from a feedback associated with Earth rotation (*7*). In particular, the collapse of the WAIS leads to a displacement of the south rotation pole of ~100 m × EEV toward the West Antarctic; this shift drives a sea-level rise in North America and the Indian Ocean and a fall over South America and Asia relative to the EEV (fig. S1, A and C).

These results reinforce serious concerns about the impact on some coastal communities of a future instability in the WAIS. Consider Washington, DC, and the case where we adopt the conventional value of 5 m for the EEV. We predict a sea-level rise 1.3 m higher than the EEV (or 6.3 m total) at this site, an increase above the EEV that is three times greater than predicted using the standard sea-level theory (Fig. 1A). Any robust assessment of the sea-level hazard associated with the loss of major ice reservoirs must, of course, account for other potential sources of meltwater, namely Greenland, the East Antarctic, and mountain glaciers (*8*). Nevertheless, future projections should avoid simple, eustatic estimates and be based on a suitably complete sea-level theory.

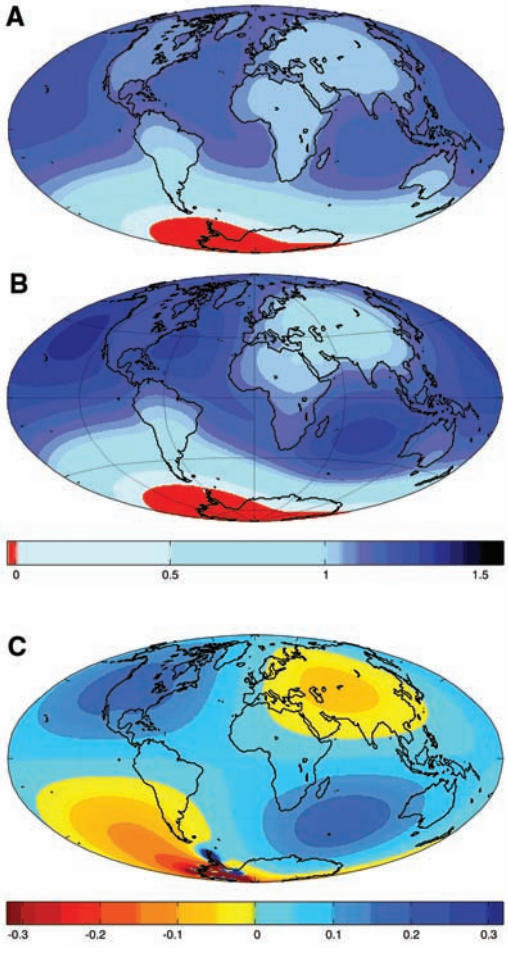


Fig. 1. Sea-level change in response to the collapse of the WAIS computed by using (**A**) a standard sea-level theory (*5*), which assumes a nonrotating Earth, no marine-based ice, and shorelines that remain fixed to the present-day geometry with time, as well as (**B**) a prediction based on a theory (*6*) that overcomes these limitations. Both predictions are normalized by the EEV associated with the ice collapse. In (B), the total volume of the WAIS is used in the calculation, whereas in (A) only an amount of ice with a volume that matches the EEV is removed (because the latter cannot take into account the inundation of marine-based sectors). (**C**) The difference between predictions generated by using the two sea-level theories [(B) minus (A)].

References and Notes

1. Materials and methods are available as supporting material on *Science* Online.
2. IPCC, *Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, Cambridge, 2007).
3. J. H. Mercer, *Nature* **271**, 321 (1978).
4. J. X. Mitrovica, M. E. Tamisiea, J. L. Davis, G. A. Milne, *Nature* **409**, 1026 (2001).
5. J. A. Clark, C. S. Lingle, *Nature* **269**, 206 (1977).
6. R. A. Kendall, J. X. Mitrovica, G. A. Milne, *Geophys. J. Int.* **161**, 679 (2005).
7. G. A. Milne, J. X. Mitrovica, *Geophys. J. Int.* **126**, F13 (1996).
8. The sea-level fingerprints associated with these sources will be purely additive, with a weighting reflecting the relative contributions to the total EEV.

9. We thank the Canadian Institute for Advanced Research, the Natural Sciences and Engineering Research Council of Canada, the John Simon Guggenheim Memorial Foundation, and NSF for support.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5915/753/DC1
Materials and Methods
Fig. S1
References

29 September 2008; accepted 20 November 2008
10.1126/science.1166510

[1]Department of Physics, University of Toronto, 60 St. George Street, Toronto, ON M5S 1A7, Canada. [2]Department of Geosciences, Oregon State University, Corvallis, OR 97331, USA.

The Formation of Massive Star Systems by Accretion

Mark R. Krumholz,[1]* Richard I. Klein,[2,3] Christopher F. McKee,[2,4]
Stella S. R. Offner,[4] Andrew J. Cunningham[3]

Massive stars produce so much light that the radiation pressure they exert on the gas and dust around them is stronger than their gravitational attraction, a condition that has long been expected to prevent them from growing by accretion. We present three-dimensional radiation-hydrodynamic simulations of the collapse of a massive prestellar core and find that radiation pressure does not halt accretion. Instead, gravitational and Rayleigh-Taylor instabilities channel gas onto the star system through nonaxisymmetric disks and filaments that self-shield against radiation while allowing radiation to escape through optically thin bubbles. Gravitational instabilities cause the disk to fragment and form a massive companion to the primary star. Radiation pressure does not limit stellar masses, but the instabilities that allow accretion to continue lead to small multiple systems.

Stars can form with masses up to at least 120 times that of the Sun (*1*, *2*), but the mechanism by which the most massive stars form is a long-standing mystery. Stars greater than ~20 solar masses ($M_\odot$) have Kelvin times (the time required for a star to radiate away its gravitational binding energy) that are shorter than their formation times, and as a result they attain their full luminosities while still accreting from their natal clouds. As the radiation from such an embedded, massive star diffuses outward through the dusty gas in the protostellar envelope, it exerts a force that opposes gravity. Spherically averaged, the ratio of the radiative and gravitational forces is $7.7 \times 10^{-5}\ \kappa_0\ (L/M)_0$, where κ_0 is the specific opacity of the gas in cm^2/g and $(L/M)_0$ is the stellar light-to-mass ratio in units of $L_\odot/M_\odot$ (where $L_\odot$ is the luminosity of the Sun). Because the dusty envelopes of massive protostars have $\kappa_0 \sim 5$, the ratio of radiative to gravitational force exceeds unity for all stars with $(L/M)_0 \geq 2500$. Main-sequence stars reach this value of $(L/M)_0$ at masses of ~20 $M_\odot$. Therefore, radiation is expected to halt spherically symmetric infall (*3*, *4*) near this mass. Two-dimensional simulations of massive star formation, which include rotation and thus an accretion disk that partially shields the gas from radiation (*5*, *6*), find that radiation completely halts accretion once stars reach ~40 $M_\odot$ (*7*); this is inconsistent with the highest known stellar masses.

Here, we report a three-dimensional simulation of the formation of stars with masses greater than 20 $M_\odot$ that includes the effects of radiation pressure. Three-dimensionality is important because instabilities that determine the interaction of gas and radiation, such as Rayleigh-Taylor instabilities, show faster growth rates and higher saturation amplitudes in three dimensions (*8*). Additionally, instabilities in accretion disks (*9*, *10*) and the resulting formation of companion stars (*11*) can only be simulated when the disk is represented nonaxisymmetrically. It is important to consider these effects because most massive stars are members of multiple systems (*12*, *13*).

Our initial conditions consisted of a gas cloud with mass = 100 $M_\odot$, radius = 0.1 pc, and density profile $\rho \propto r^{-1.5}$, consistent with models (*14*, *15*) and observations (*16*) of the initial states of massive prestellar cores. Its initial temperature was 20 K and it was in slow, solid-body rotation at a rate such that the ratio of rotational kinetic energy to gravitational binding energy was 0.02, which is consistent with the rotation rates seen in lower-mass cores (*17*). Previous two-dimensional simulations suggest that varying these parameters within the observed range of massive core properties would not alter the qualitative behavior (*7*). Even though observed massive cores have large turbulent velocities (*16*), we did not include turbulence in our initial conditions so as to focus on the effects of radiation pressure. Simulations that include the effects of turbulence (*10*) do not appear to produce qualitatively different results.

We evolved this initial state by means of our adaptive mesh refinement code ORION (*18*–*22*), which solves the equations of gravito-radiation-hydrodynamics in the gray, flux-limited diffusion approximation. The code dynamically increases resolution as needed down to a minimum cell size of 10 AU; regions that collapse to densities above the Jeans density at the maximum resolution become star particles, each of which produces a luminosity determined by a protostellar evolution model (*23*).

The simulation passed through several distinct phases (Fig. 1, figs. S1 to S3, and movie S1), forming multiple stars (Fig. 2). The cloud began to collapse immediately and a central protostar formed 3600 years afterward. For the next 17,000 years, the protostar accreted smoothly via an axisymmetric disk (Fig. 1A). During this phase, the mass of the star grew to 11 $M_\odot$ and its luminosity remained below ~10^4 $L_\odot$. Because $(L/M)_0 < 1000$, radiation pressure produced no noticeable effects. After ~20,000 years, the disk became gravitationally unstable and developed a pronounced two-armed spiral that transported angular momentum efficiently (Fig. 1B) (*24*). Accretion onto the protostar continued smoothly.

Accretion, unimpeded by radiation pressure, continued until 25,000 to 26,000 years, when the star reached a mass of roughly 17 $M_\odot$ and achieved a sustained $(L/M)_0$ value of ~2500, driven by Kelvin-Helmholtz contraction. This was luminous enough for its radiation pressure force to exceed the gravitational force, and the star began to drive gas outward around the polar axis, inflating radiation-filled bubbles both above and below the accretion disk (Fig. 1C). The density inside the bubbles was very low, and within them the radiation pressure exceeded the gas pressure by orders of magnitude. Almost all gas falling onto the protostar struck the walls of the bubbles, where it was shocked and swept up into the bubble walls. However, this did not slow accretion in our simulation, because the gas that struck the bubble walls eventually traveled along the margin until it reached the disk, at which point it continued to accrete onto the star. During this phase, radiation forces acting on material accreting onto the disk and gravitational forces acting on material in the disk caused it to become increasingly nonaxisymmetric. A series of small secondary stars formed in the disk, most of which advected inward because of dynamical friction with the gas and collided with the central protostar. As a result, the accretion rate onto the central star became variable, but its mean value remained roughly unchanged.

Around 35,000 years, a series of disk-borne stars collided and became massive enough to resist being dragged inward (Fig. 1D). The secondary star was initially smaller than the primary and orbited it, intercepting and accreting much of the inflowing gas (*25*). As a result, the secondary star acquired its own disk and grew faster at first than the primary, and the ratio of the mass of the secondary to that of the primary reached a value of >0.5. Thereafter, the disk continued to fragment but at a much-diminished rate, and accretion became almost evenly divided between the two massive stars. A third small disk-borne star was ejected into a wide orbit in our simulation, but eventually fell back and was captured and accreted. The total accretion rate onto the binary system varied periodically as the stars orbited one another, but its time-averaged value remained about the same as before binary formation. From this point onward, the bubbles showed instability and constantly changed shape while undergoing slow overall expansion. Accretion onto the system continued uninterrupted.

[1]Department of Astronomy, University of California, Santa Cruz, CA 95064, USA. [2]Department of Astronomy, University of California, Berkeley, CA 94720, USA. [3]AX Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. [4]Department of Physics, University of California, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: krumholz@ucolick.org

We halted the simulation at 57,000 years, after a ~20,000-year period when there was no further qualitative change in the evolution (Fig. 1E). At this point the system was a binary with a total mass of 70.7 $M_\odot$ and a time-averaged total luminosity of ~5×10^5 $L_\odot$. The two stars had masses of 41.5 $M_\odot$ and 29.2 $M_\odot$ and were 1590 AU apart. Neglecting the effects of the gas, the semimajor axis of the orbit was 1280 AU (eccentricity 0.25), but because this neglects the gas, it may be an overestimate. Orbits like this are typical of young O stars, at least 40% of which are visual binaries with separations of ~1000 AU (*12*). These are not the final system parameters, because the envelope and the disk still contained 28.3 $M_\odot$ of gas and the accretion rate had not diminished. However, the qualitative nature of the final system was well established.

We compared our result to two-dimensional simulations. The largest star that formed in any two-dimensional simulation with gray radiative transfer had a mass of 22.9 $M_\odot$. If the simulation included a multifrequency treatment of the radiation, which we omitted because of its computational cost (*23*), the maximum mass of the star that formed was 42.9 $M_\odot$ (*7*). In these two-dimensional simulations, the initial phases of collapse, disk formation, and growth of a polar bubble were quite similar to ours, although the disk lacked nonaxisymmetric structure. In both cases there was a "flashlight effect" (*7*, *26*) in which the disk beamed radiation preferentially in the polar direction. In two dimensions, however, as the star's mass grew, radiation halted accretion over an ever larger fraction of the solid angle around the star. This eventually stopped infall onto the disk. Some of the gas remaining in the disk continued to accrete onto the star, but at a diminishing rate, and eventually the disk density became low enough for stellar radiation to blow it away.

This never happened in our simulation. Instead, when the luminosity became large enough that our bubbles no longer delivered mass to the disk efficiently, they became asymmetric and clumpy. In some places radiation blew out sections of the bubble wall, whereas in others dense filaments of gas fell toward the stars (Fig. 3). The structure of dense fingers of heavy, downward-moving fluid alternating with chimneys of outgoing radiation is analogous to that of a classical Rayleigh-Taylor instability, with radiation taking the place of the light fluid. Radiation forces away from the star are stronger than gravity when averaged over 4π sr, producing velocities and net forces that have an outward direction over most of the solid angle. Much of the mass is concentrated into the dense fingers, and because radiation flows around rather than through these structures, within them the velocity and the net force have an inward direction. However, this did not remove the angular momentum of the gas, so it continued to fall onto the disk rather than directly onto the stars. The growth of clumps in the disk that form secondary stars is a natural side effect of this process, but radiation may be just accelerating a process that is caused by gravity. At least 40% of the accreting gas reached the disk through this Rayleigh-Taylor mechanism; gas falling onto the outer disk directly accounted for ~25% of the accretion, and gas reaching the disk by traveling along the bubbles' outer walls contributed the remaining ~35% (*23*).

Continued disk feeding is what made the three-dimensional results different from earlier two-dimensional ones. At 34,000 and 41,700 years (Fig. 1, C and D), bracketing the onset of

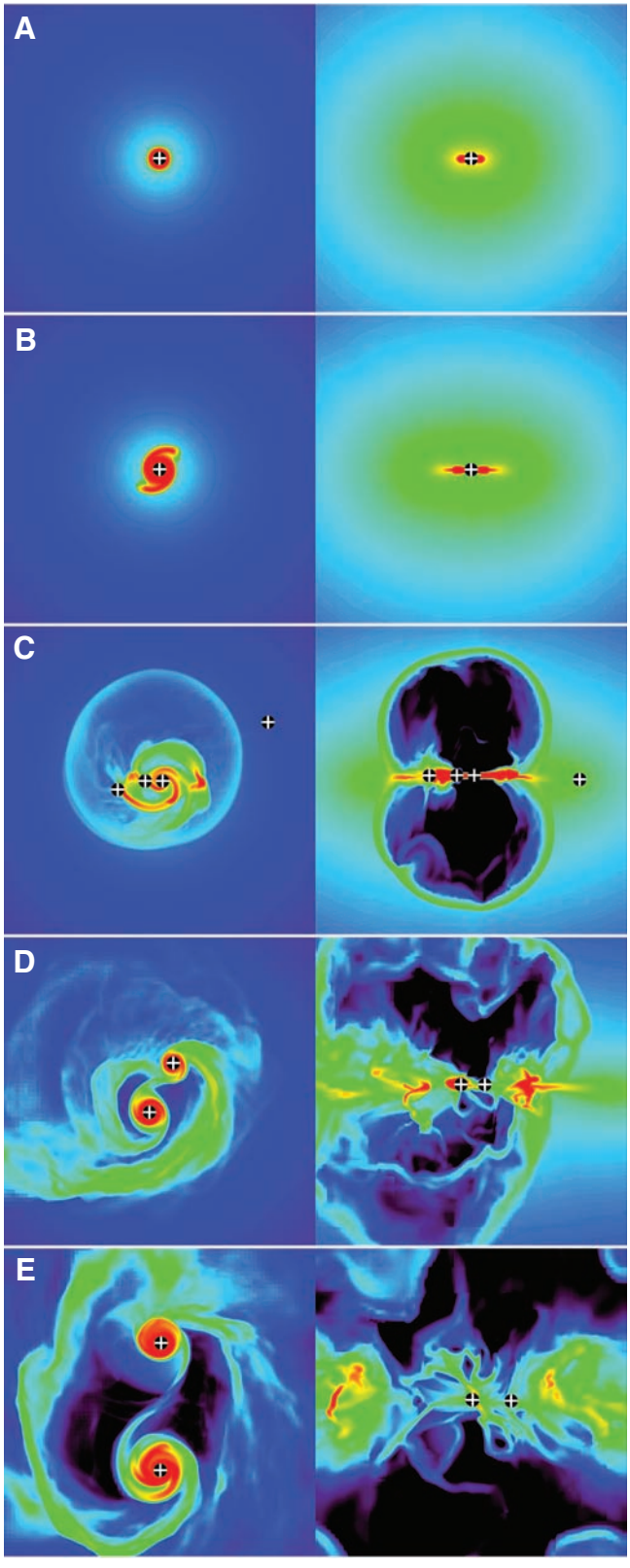


Fig. 1. Snapshots of the simulation at (**A**) 17,500 years, (**B**) 25,000 years, (**C**) 34,000 years, (**D**) 41,700 years, and (**E**) 55,900 years. In each panel, the left image shows column density perpendicular to the rotation axis in a $(3000\ \mathrm{AU})^2$ region; the right image shows volume density in a $(3000\ \mathrm{AU})^2$ slice along the rotation axis. The color scales are logarithmic (black at the minimum, red at the maximum), from 10^0 to $10^{2.5}$ g cm^{-2} on the left and 10^{-18} to 10^{-14} g cm^{-3} on the right. Plus signs indicate the projected positions of stars. See figs. S1 to S3 and movie S1 for additional images.

instability, the total stellar mass was 32.4 $M_\odot$ and 46.9 $M_\odot$, respectively, and the disk mass was 4.0 to 5.7 $M_\odot$ and 4.5 to 9.1 $M_\odot$, respectively. (The range reflects the use of density cutoffs of 10^{-14} to 10^{-16} g cm^{-3} to separate disk from nondisk material.) If accretion of gas onto the disk had halted, then the evolution would likely have been the same as in the two-dimensional simulations. The remnant disk would still have accreted, but its low mass means that the stars would have gained less than 10 $M_\odot$ from it. In our three-dimensional simulation they instead accreted 25 to 40 $M_\odot$, and did so at a constant rather than a declining accretion rate. Accretion of stars rather than gas did not contribute appreciably to this. The star with a final mass of ~40 $M_\odot$ gained only 1.8 $M_\odot$ via collisions, whereas the ~30 $M_\odot$ star gained 1.2 $M_\odot$, excluding the initial collision that created it.

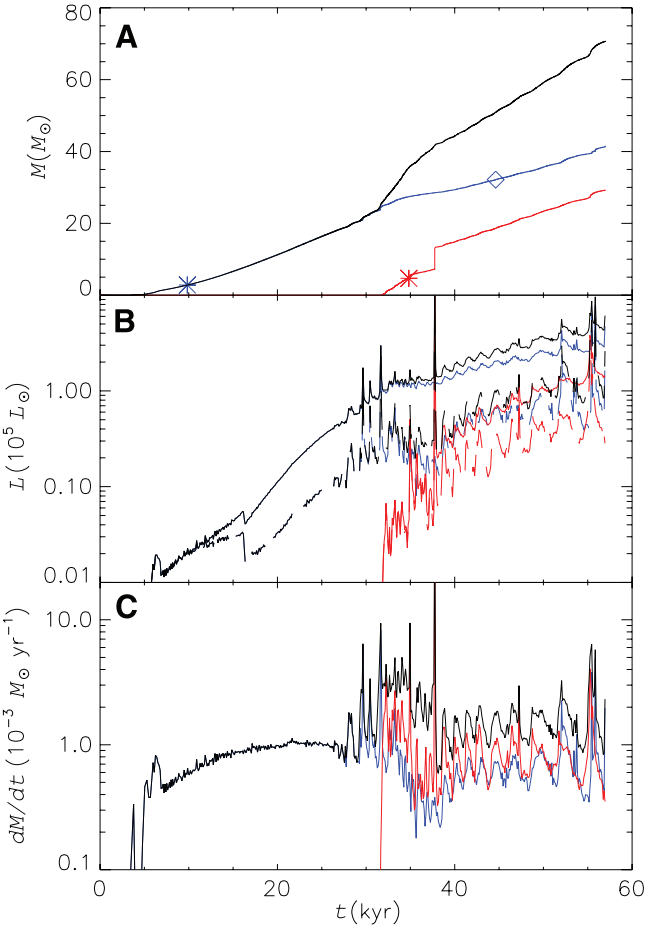


Fig. 2. (**A**) Stellar mass, (**B**) stellar luminosity, and (**C**) accretion rate as a function of time. Black lines show values summed over all stars, blue lines show values for the most massive star, and red lines show values for the second most massive star. In (**A**), asterisks mark the onset of deuterium burning and the diamond marks hydrogen burning. In (**B**), solid lines show luminosities from all sources (accretion, Kelvin-Helmholtz contraction, and nuclear burning); dashed lines show accretion luminosity only. Luminosities and accretion rates are 100-year running averages.

The final star formation efficiency for our simulated core was at least 70%, and a majority of the mass accreted within one mean-density free-fall time of the initial core (52,500 years). Because infall was continuing at a roughly constant rate and no further qualitative changes were occurring at the time we halted the simulation, it is likely that much of the remaining mass would accrete onto the star system. In reality, protostellar outflows, which we have not included in our simulation, would limit the star formation efficiency to ~50% (*27*, *28*). Our result indicates that, relative to outflows, radiation pressure does not affect star formation efficiency or time scale. The cavities generated by outflows would reduce the effects of radiation pressure even further (*29*) and would modify the geometry of the radiation pressure bubbles or would prevent their formation altogether. Photon bubble instabilities, which can occur if the gas is sufficiently magnetized (*30*), might also reduce the effects of radiation pressure and modify the bubble geometry. Our simulation shows that even if these effects are omitted, radiation pressure does not present a barrier to massive star formation.

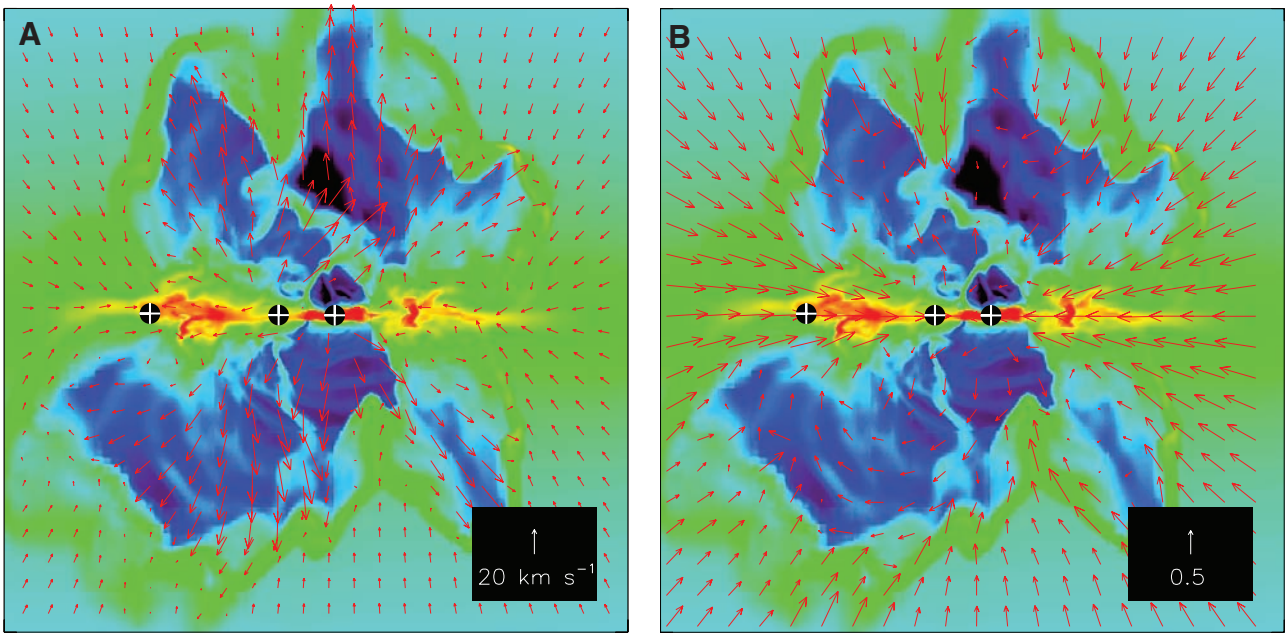


Fig. 3. Snapshot of a (6000 AU)^2 slice along the rotation axis at 51,100 years. Color indicates density from 10^{-20} to 10^{-14} g cm^{-3} on a logarithmic scale as in Fig. 1. Plus signs show projected stellar positions. (**A**) Arrows show gas velocity. (**B**) Arrow directions indicate the direction of the net (radiation plus gravitational) force; lengths are proportional to the magnitude of the net force divided by the magnitude of the gravitational force. Thus, an inward arrow of length 1 represents negligible radiation force.

References and Notes

1. C. Weidner, P. Kroupa, *Mon. Not. R. Astron. Soc.* **348**, 187 (2004).
2. D. F. Figer, *Nature* **434**, 192 (2005).
3. F. D. Kahn, *Astron. Astrophys.* **37**, 149 (1974).
4. M. G. Wolfire, J. P. Cassinelli, *Astrophys. J.* **319**, 850 (1987).
5. T. Nakano, T. Hasegawa, C. Norman, *Astrophys. J.* **450**, 183 (1995).
6. J. Jijina, F. C. Adams, *Astrophys. J.* **462**, 874 (1996).
7. H. W. Yorke, C. Sonnhalter, *Astrophys. J.* **569**, 846 (2002).
8. M. M. Marinak *et al.*, *Phys. Rev. Lett.* **75**, 3677 (1995).
9. F. H. Shu, S. Tremaine, F. C. Adams, S. P. Ruden, *Astrophys. J.* **358**, 495 (1990).
10. M. R. Krumholz, R. I. Klein, C. F. McKee, *Astrophys. J.* **656**, 959 (2007).
11. K. M. Kratter, C. D. Matzner, M. R. Krumholz, *Astrophys. J.* **681**, 375 (2008).
12. B. D. Mason *et al.*, *Astron. J.* **115**, 821 (1998).
13. H. Sana, E. Gosset, Y. Nazé, G. Rauw, N. Linder, *Mon. Not. R. Astron. Soc.* **386**, 447 (2008).
14. C. F. McKee, J. C. Tan, *Nature* **416**, 59 (2002).
15. C. F. McKee, J. C. Tan, *Astrophys. J.* **585**, 850 (2003).
16. H. Beuther *et al.*, *Astron. Astrophys.* **466**, 1065 (2007).
17. A. A. Goodman, P. J. Benson, G. A. Fuller, P. C. Myers, *Astrophys. J.* **406**, 528 (1993).
18. R. I. Klein, *J. Comput. Appl. Math.* **109**, 123 (1999).
19. M. R. Krumholz, C. F. McKee, R. I. Klein, *Astrophys. J.* **611**, 399 (2004).
20. R. T. Fisher, thesis, University of California, Berkeley (2002).
21. M. R. Krumholz, R. I. Klein, C. F. McKee, J. Bolstad, *Astrophys. J.* **667**, 626 (2007).
22. A. I. Shestakov, S. S. R. Offner, *J. Comput. Phys.* **227**, 2154 (2008).
23. See supporting material on *Science* Online.
24. G. Lodato, W. K. M. Rice, *Mon. Not. R. Astron. Soc.* **358**, 1489 (2005).
25. M. R. Bate, *Mon. Not. R. Astron. Soc.* **314**, 33 (2000).
26. H. W. Yorke, P. Bodenheimer, *Astrophys. J.* **525**, 330 (1999).
27. C. D. Matzner, C. F. McKee, *Astrophys. J.* **545**, 364 (2000).
28. J. Alves, M. Lombardi, C. J. Lada, *Astron. Astrophys.* **462**, L17 (2007).
29. M. R. Krumholz, C. F. McKee, R. I. Klein, *Astrophys. J.* **618**, L33 (2005).
30. N. J. Turner, E. Quataert, H. W. Yorke, *Astrophys. J.* **662**, 1052 (2007).

31. Supported by NSF grants AST-0807739 (M.R.K.) and AST-0606831 (R.I.K. and C.F.M.); the Spitzer Space Telescope Theoretical Research Program, provided by NASA through a contract issued by the Jet Propulsion Laboratory (M.R.K.); NASA through Astrophysics Theory and Fundamental Physics Program grants NAG 05-12042 and NNG 06-GH96G (R.I.K. and C.F.M.); and the U.S. Department of Energy at Lawrence Livermore National Laboratory under contract B-542762 (R.I.K., S.S.R.O., and A.J.C.). This research used the Datastar system at the NSF San Diego Supercomputer Center (grant UCB267).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1165857/DC1
SOM Text
Figs. S1 to S5
Movie S1
References

12 September 2008; accepted 11 December 2008
Published online 15 January 2009;
10.1126/science.1165857
Include this information when citing this paper.

Confined Crystallization of Polyethylene Oxide in Nanolayer Assemblies

Haopeng Wang,[1] Jong K. Keum,[1] Anne Hiltner,[1]* Eric Baer,[1] Benny Freeman,[2] Artur Rozanski,[3] Andrzej Galeski[3]

The design and fabrication of ultrathin polymer layers are of increasing importance because of the rapid development of nanoscience and nanotechnology. Confined, two-dimensional crystallization of polymers presents challenges and opportunities due to the long-chain, covalently bonded nature of the macromolecule. Using an innovative layer-multiplying coextrusion process to obtain assemblies with thousands of polymer nanolayers, we discovered a morphology that emerges as confined polyethylene oxide (PEO) layers are made progressively thinner. When the thickness is confined to 20 nanometers, the PEO crystallizes as single, high-aspect-ratio lamellae that resemble single crystals. Unexpectedly, the crystallization habit imparts two orders of magnitude reduction in the gas permeability.

Crystalline polymers, such as polyethylene, polypropylene, poly(ethylene terephthalate), and nylon, have been broadly used as gas-barrier films in food, medicine, and electronics packaging, benefiting thereby from their low cost, easy processing, and mechanical toughness. Good barrier properties are imparted by the ability of polymer chains to crystallize into semicrystalline materials with both crystalline and amorphous phases (*1*). The efficiency of chain packing is such that the crystalline phase is generally regarded as impermeable to even small gas molecules, and gas transport is seen as occurring through the amorphous regions (*2*). The processing conditions can be readily varied to control the amount of crystallinity and chain orientation and to tune the barrier properties of the final product (*3*). With the growing use of polymers as thin and ultrathin films (*4*, *5*), morphologies have been found resulting from constrained two-dimensional (2D) polymer crystallization (*4*, *6*). These crystalline morphologies could possess gas permeability characteristics that are not expected from the bulk polymers.

Despite confinement, crystallization of polymer chains follows the conventional habit whereby polymer chains fold back and forth into stems to form crystalline lamellae with a thickness of ~10 to 20 nm. It is typical of crystallization from the isotropic melt that the lamellae are organized in a spherulitic morphology (*7*). However, the processes of nucleation and growth that control the crystallization kinetics can be profoundly affected by nanoscale confinement. The thickness of ultrathin polymer layers, usually a few tens of nanometers, is comparable to or a small multiple of the lamellar crystal thickness. Hence, the isotropic growth of lamellar crystals is greatly hampered, and crystallization under confinement can produce a specific lamellar crystal orientation. Often, the preferred lamellar crystal orientation is vertical to the layer (edge-on) (*8*, *9*, *10*). However, at the other extreme, lamellar crystal orientation parallel to the layer (flat-on) is observed (*6*, *11*–*14*). Although the mechanisms for the specific lamellar orientation during confined crystallization are still under investigation (*15*, *16*), it is believed that the confined crystals will show anisotropic properties.

The 2D crystallization of polymers is conventionally studied with thin films or block copolymers that contain a crystallizable block. In the former, crystallizable layers with nanometer to submicron thicknesses are prepared by a solution process such as spin-coating (*4*, *17*) or Langmuir-Blodgett (*18*) techniques. This approach is limited by the solvent requirement and by the small amount of material that can be fabricated. In the latter, a layered morphology on the nanometer scale is achieved as a consequence of microphase separation of the dissimilar blocks. If the crystallization temperature of the crystallizable block is below the order-disorder transition temperature (*19*), crystallization occurs with confinement in the layer-normal direction. A wide range of crystallizable blocks have been studied (*8*, *11*–*14*, *16*). However, elucidation of the structure-property relationships has been hindered by the need to synthesize the block copolymers and by the shear alignment that is required to construct the uniformly oriented phase structure (*20*).

In contrast to the self-assembled confinement created with microphase-separated block copolymers (*21*), layer-multiplying coextrusion uses forced assembly to create films with hundreds or thousands of alternating layers of two polymers (*22*, *23*). Almost any melt-processable polymer can be fabricated into kilometers of nanolayered films, and layers less than 10 nm in thickness have been made. Although the amount of material in a single layer is very small, the properties of the confined layer are multiplied many-fold by the number of identical layers in the assembly. This enables us to use conventional methods to probe size-scale-dependent properties. Polyethylene oxide (PEO) was coextruded with poly

[1]Department of Macromolecular Science and Engineering, Case Western Reserve University, Cleveland, OH 44106–7202, USA. [2]Department of Chemical Engineering, University of Texas at Austin, Austin, TX 78758, USA. [3]Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, 90-363 Lodz, Poland.

*To whom correspondence should be addressed. E-mail: ahiltner@case.edu

(ethylene-*co*-acrylic acid) (EAA), a copolymer with much lower crystallinity than PEO (fig. S1). Films with 33, 257, and 1025 alternating EAA and PEO layers were coextruded with various thicknesses and various composition ratios, including (EAA/PEO vol/vol) 50/50, 70/30, 80/20 and 90/10 (table S1). The nominal PEO layer thickness, which was calculated from the number of layers, the composition ratio, and the film thickness, varied from 3.6 μm to 8 nm. Nonlayered control films were also extruded.

Permeability to small molecules is an important performance property of polymer films. Gas transport also provides a probe into the solid-state structure (*3*). The oxygen permeability (*P*) was first measured on films where the PEO and EAA layers had the same thickness. The layer thickness was varied by changing the number of coextruded layers and the overall film thickness while maintaining the composition at 50/50. The results, plotted in Fig. 1A as a function of layer thickness, unexpectedly showed a steady decrease in the oxygen permeability as the EAA and PEO layers became thinner. A comparable reduction was found for carbon dioxide permeability.

The series model for layered assemblies gives the gas permeability as

$$P_{\parallel} = \left(\frac{\phi_{PEO}}{P_{PEO}} + \frac{1-\phi_{PEO}}{P_{EAA}}\right)^{-1} \quad (1)$$

where ϕ_{PEO} is the volume fraction of PEO, and P_{PEO} and P_{EAA} are the permeabilities of PEO and EAA control films, respectively. Using the determined values of 0.38 barrer (*24*) and 2.30 barrer for P_{PEO} and P_{EAA}, Eq. 1 gave the oxygen permeability of an EAA/PEO 50/50 layered assembly as 0.65 barrer. Only the film with the thickest layers conformed to this prediction.

Noting that PEO is substantially less permeable to oxygen than EAA, Eq. 1 predicts that $P_{//}$ will be quite sensitive to P_{PEO} even if ϕ_{PEO} is relatively small. To ascertain whether a dramatic change in the PEO permeability was responsible for the layer thickness effect, numerous films that varied in both the composition ratio and the layer thickness were tested. Although the measured values of *P* scattered, depending on the composition ratio, when an effective PEO permeability was extracted,

$$P_{PEO,eff} = \phi_{PEO}\left(\frac{1}{P} - \frac{1-\phi_{PEO}}{P_{EAA}}\right)^{-1} \quad (2)$$

the data collapsed to a single curve when $P_{PEO,eff}$ was plotted as a function of the PEO layer thickness (Fig. 1B). Only the results for the thicker PEO microlayers conformed to Eq. 2, with $P_{PEO,eff} \approx P_{PEO}$, as indicated by the dashed line. Deviation below the line was seen with 1-μm-thick PEO layers. The lowest $P_{PEO,eff}$ was achieved with 20-nm layers, and the value of 0.0052 barrer was almost 2 orders of magnitude less than P_{PEO}.

It seemed likely that crystallization in a confined space resulted in an unusual crystalline morphology that endowed the PEO nanolayers with exquisitely low permeability. However, differential scanning calorimetry revealed that even in the thinnest layers, both PEO and EAA possessed the same melting enthalpy and the same melting

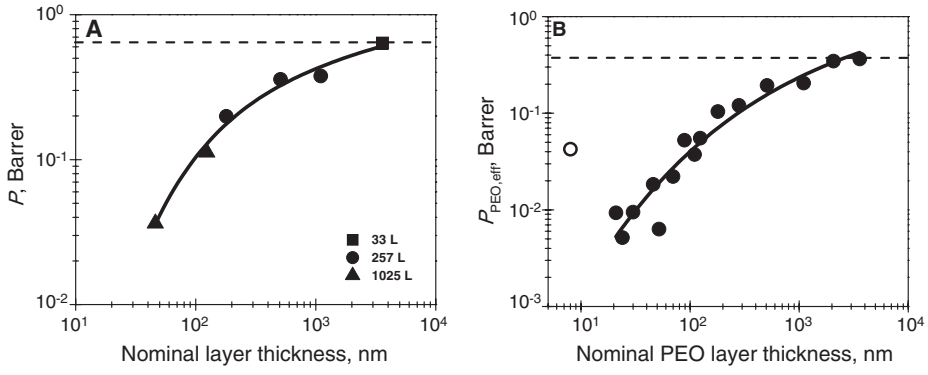
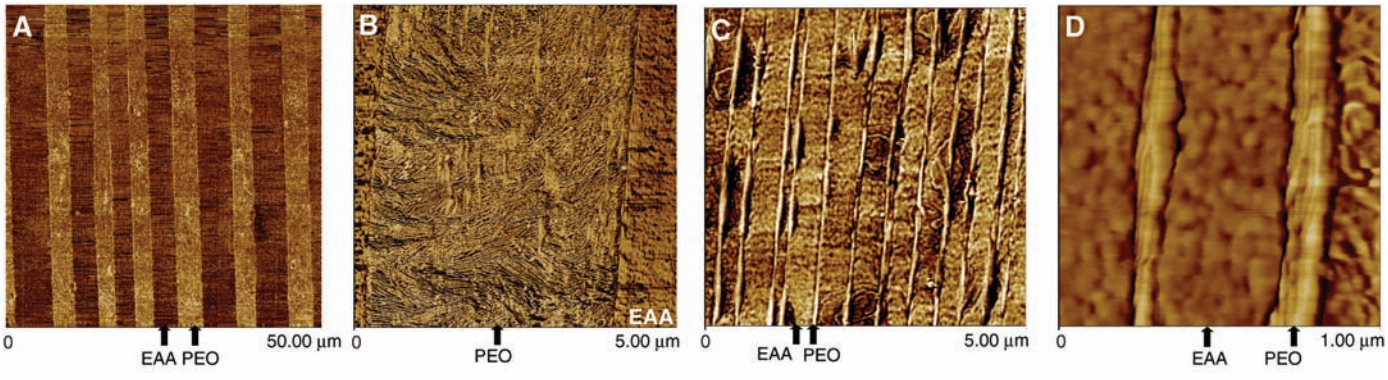


Fig. 1. The effect of layer thickness on oxygen permeability. (**A**) Oxygen permeability of films with equal volume fractions of EAA and PEO. The dashed line indicates $P_{//}$ calculated from Eq. 1. (**B**) Oxygen permeability of the PEO layers from films of varying composition calculated from Eq. 2. The dashed line indicates P_{PEO}. The open symbol is for a film with PEO layer breakup. The solid lines are drawn to guide the eyes.



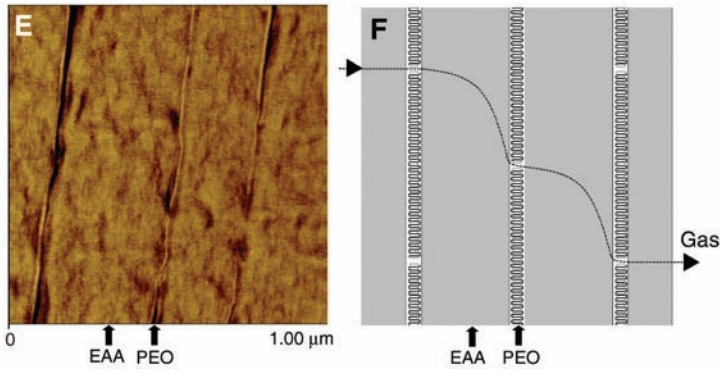


Fig. 2. AFM phase images of partial cross sections of the layered EAA/PEO films. The PEO layer has substantially higher crystallinity than the EAA layers and hence appears bright in the AFM images. (**A**) A low-resolution image of an EEA/PEO film with 50/50 composition, 33 alternating layers, and nominal PEO layer thickness of 3.6 μm. (**B**) A higher-resolution image showing the spherulitic morphology of the 3.6-μm-thick PEO layer. (**C**) A low-resolution image of an EAA/PEO film with 70/30 composition, 1025 alternating layers and nominal PEO layer thickness of 110 nm. (**D**) A higher-resolution image of the 110-nm-thick PEO layers showing the oriented stacks of PEO lamellae. (**E**) A high-resolution image of an EAA/PEO film with 90/10 composition, 1025 alternating layers, and nominal PEO layer thickness of 20 nm showing that the PEO layers crystallized as single, extremely large lamellae. (**F**) A schematic showing the gas diffusion pathway through the layered assembly with 20-nm-thick PEO layers. The arrows identify the EAA layers and PEO layers.

temperature as the control films, which were 153 J/g and 66°C for PEO, and 98 J/g and 98°C for EAA, respectively (fig. S2 and table S2). Thus, any unusual crystalline morphology that provided the very low permeability of PEO nanolayers was not accompanied by changes in the crystallinity or in the lamellar thickness.

The layers were viewed by microtoming the film through the thickness and examining the exposed surface in atomic force microscope (AFM). A region from the cross section of a film with 3.6-μm-thick PEO layers confirmed the layer continuity (Fig. 2A). Although there was some nonuniformity, the average layer thickness was close to the calculated nominal layer thickness. A higher magnification image showed the sharp boundaries between EAA and PEO layers (Fig. 2B). The spherulitic morphology of the bright PEO layer closely resembled that of PEO crystallized from the unconfined melt. It was expected that the properties of the PEO layers would also be the same and, indeed, the oxygen permeability of films with thick PEO layers closely conformed to Eq. 2 with $P_{PEO,eff} \approx P_{PEO}$.

Another pair of images in Fig. 2, C and D, compares a film with 110-nm-thick PEO layers. The images confirmed the continuity of the thin PEO layers and the close correspondence between the average layer thickness and the nominal thickness. At higher magnification, the effect of confinement on crystallization of the PEO layer shows that PEO crystallized as stacks of three to five long, thin lamellae oriented in the plane of the layer. When the PEO layer thickness was reduced to 20 nm, most of the PEO layers crystallized as single, extremely large lamellae whose lateral dimensions frequently exceeded the field of the AFM image (Fig. 2E). Coincidence between the layer thickness, which was determined by the extrusion conditions, and the thickness of PEO lamellae, about 20 nm, facilitated crystallization of the layers as single lamellae. The single lamellae could be thought of as very large single crystals (*25*). If the layer thickness was reduced to 8 nm, the layers broke up, which caused the increased permeability of the film with 8-nm PEO layers.

The lamellar crystalline core is considered impermeable, and the lamellar fold surfaces constitute the permeable amorphous regions. For a single layer of flat-on lamellae, as in the PEO nanolayer, the diffusion pathway depends on the frequency of defects such as lamellar edges (Fig. 2F). Structurally, the nanolayered assembly resembles a dispersion of impermeable platelets of given aspect ratio. If the platelets are oriented perpendicular to the flux, the permeability of the composite is expressed as (*26*)

$$P = P_{EAA}\left[1 + \frac{\alpha^2\phi^2}{4(1-\phi)}\right]^{-1} \quad (3)$$

where ϕ is the volume fraction of impermeable platelets and α is the aspect ratio of the platelets defined as length divided by width. In this case, ϕ was taken as the volume fraction of the PEO crystalline phase. For the thinnest PEO nanolayers, the aspect ratio from Eq. 3 was as high as 120, which meant a lateral dimension of more than 2 μm for lamellae 20 nm thick. Gradually increasing the PEO layer thickness relaxed the restriction on 3D growth, which allowed crystallization of lamellar stacks and permitted randomization of the lamellar orientation. Also, the increased area density of nuclei in individual layers reduced the lateral dimension of lamellae. The gradual change in crystallization habit from single lamellae to isotropic spherulites paralleled a gradual increase in the permeability of the PEO layer.

Confirmation of the oriented lamellar morphology and details of the global orientation were obtained with small-angle x-ray scattering (SAXS) and wide-angle x-ray scattering (WAXS). The SAXS examined the periodic arrangement of lamellar crystals within the constituent layers. By aligning the incident x-ray beam parallel to the normal direction (ND), the extrusion direction (ED), and the transverse direction (TD), the orientation of lamella was determined (fig. S3). The scattering patterns indicated that the long spacing of the PEO and EAA lamellae in layered films were 22.0 ± 0.6 nm and 10.8 ± 0.5 nm, respectively, nearly the same as in the control films. Isotropic scattering patterns in all three directions from 3.6 μm PEO layers indicated that the PEO layers were too thick for PEO lamellae to feel any substantial confinement effect. However, as the PEO layer thickness decreased to 110 nm, highly oriented meridional two-point scattering features appeared in the ED and TD patterns, which indicated that stacked PEO lamellae were oriented primarily parallel to the layer surface due to the confinement. The extremely weak first-order peak from 20 nm PEO layers indicated that they existed predominantly as single lamellae rather than as stacks (fig. S4).

The orientation of the PEO chains in the crystal was examined with 2D WAXS and pole

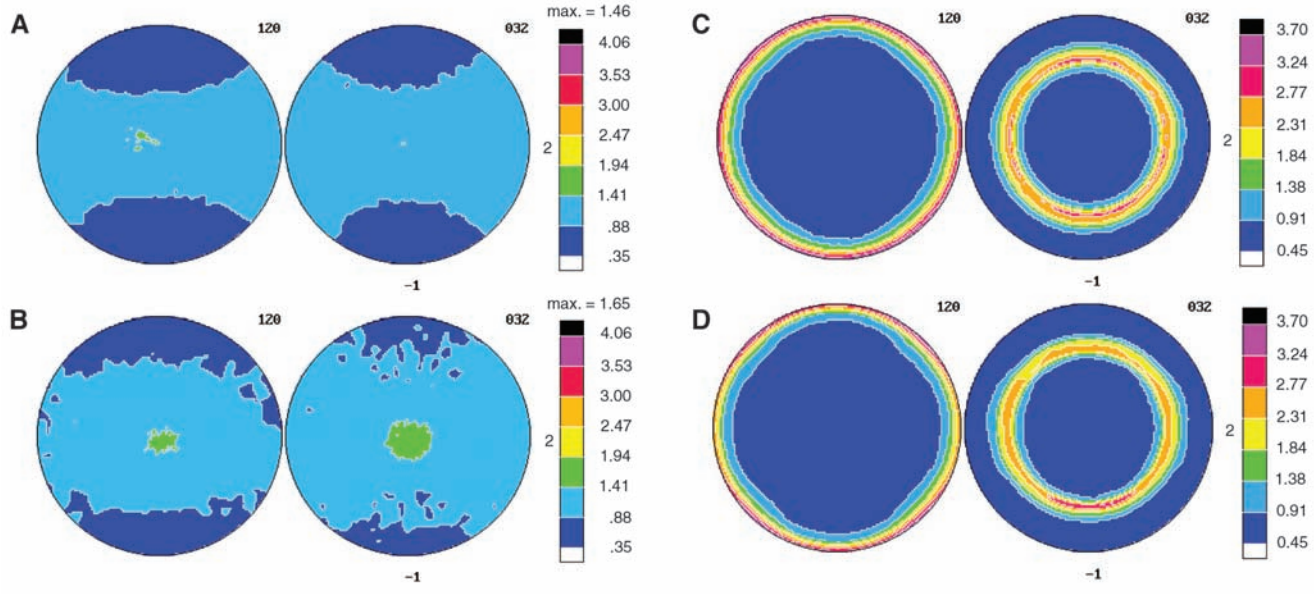


Fig. 3. Pole figures of normals to the (120) and (032) planes of the PEO monoclinic crystals (*27*). The extrusion direction is vertical and the transverse direction is horizontal. The normal direction is in the center of the pole figure. (**A**) The PEO control film. (**B**) An EAA/PEO film with 50/50 composition, 33 alternating layers, and nominal PEO layer thickness of 3.6 μm. (**C**) An EAA/PEO film with 70/30 composition, 1025 alternating layers, and nominal PEO layer thickness of 110 nm. Orientation of the (120) planes is perpendicular to the layer plane, and orientation of the (032) planes is at 67°. (**D**) An EAA/PEO film with 90/10 composition, 1025 alternating layers, and nominal PEO layer thickness of 20 nm.

figures. Consistent results were obtained by these two techniques. From the pole figures of normals to the (120) and (032) planes of PEO (*27*), it is seen that there is no preferred orientation of crystals in the control film except for a slight orientation due to the extrusion process (Fig. 3A). In Fig. 3B, the film with 3.6-μm-thick PEO layers also showed only very weak orientation. In contrast, the film with 110-nm-thick PEO layers showed a very strong orientation of the (120) and also the (032) planes (Fig. 3C). Nearly all the (120) planes that contained polymer chains were perpendicular to the film plane. This meant that the fold surfaces of the lamellar PEO crystals were in the plane of the layer. Upon decreasing the PEO layer thickness to 20 nm, the preferred orientation of PEO lamellae parallel to the layers seemed even stronger, as indicated by the narrower ring at the pole figure circumference (Fig. 3D). The (120) planes were distributed evenly in the plane of film, always being perpendicular to the film surface. The pole figures for (032) normal in Fig. 3, C and D, resembled rings exactly offset by 67°, as predicted by the crystallographic unit cell for the preferred orientation of PEO lamellae parallel to the layer interface (*16*). Again, the ring for the (032) normal in Fig. 3D was much narrower than in Fig. 3C.

The crystal orientation of PEO in confined nanolayers essentially reproduced the crystal structure of PEO blocks in self-assembled PS-*b*-PEO diblock copolymers (*14*, *16*). Comparing the sharpness of the WAXS pattern, considerably higher orientation was achieved by physically confining a high-molecular-weight PEO between force-assembled layers than by confining a low-molecular-weight PEO block between self-assembled lamellae with covalent links. When the thickness confinement occurred on the size scale of the usual lamellar thickness, the PEO layers crystallized as single lamellae with extremely large aspect ratios. It was suggested that the lamellae could be thought of as large, impermeable single crystals.

The coextrusion process, which operates with readily available polymers, now makes it possible to fabricate nanolayered polymeric structures in sufficient quantities to probe the structure-property relationships of the morphologies resulting from nanoscale confinement. For design and execution of packaging strategies, polymer nanolayers can be incorporated into conventional polymeric films with the right barrier properties for less cost, which in turn may reduce the environmental and energy impact.

References and Notes

1. D. C. Bassett, *Principles of Polymer Morphology* (Cambridge Univ. Press, Cambridge, 1981).
2. D. H. Weinkauf, D. R. Paul, in *Barrier Polymers and Structures*, W. J. Koros, Ed. (American Chemical Society, Washington, DC 1990), pp. 60–91.
3. A. Hiltner, R. Y. F. Liu, Y. S. Hu, E. Baer, *J. Polym. Sci. Pt. B Polym. Phys.* **43**, 1047 (2005).
4. C. W. Frank *et al.*, *Science* **273**, 912 (1996).
5. S. Goffri *et al.*, *Nat. Mater.* **5**, 950 (2006).
6. G. Reiter *et al.*, *Lect. Notes Phys.* **714**, 179 (2007).
7. E. Baer, A. Hiltner, H. D. Keith, *Science* **235**, 1015 (1987).
8. I. W. Hamley *et al.*, *Macromolecules* **29**, 8835 (1996).
9. T. E. Bernal-Lara, R. Y. F. Liu, A. Hiltner, E. Baer, *Polymer (Guildf.)* **46**, 3043 (2005).
10. Y. Jin *et al.*, *J. Polym. Sci. Pt. B Polym. Phys.* **42**, 3380 (2004).
11. R. M. Ho *et al.*, *Macromolecules* **37**, 5985 (2004).
12. Y. S. Sun *et al.*, *Macromolecules* **39**, 5782 (2006).
13. U. Mukai, R. E. Cohen, A. Bellare, R. J. Albalak, *J. Appl. Polym. Sci.* **70**, 1985 (1998).
14. M.-S. Hsiao *et al.*, *Macromolecules* **41**, 8114 (2008).
15. Y. Ma, W. Hu, G. Reiter, *Macromolecules* **39**, 5159 (2006).
16. L. Zhu *et al.*, *J. Am. Chem. Soc.* **122**, 5957 (2000).
17. S. Napolitano, M. Wübbenhorst, *J. Phys. Condens. Matter* **19**, 205121 (2007).
18. B. Li, A. R. Esker, *Langmuir* **23**, 2546 (2007).
19. F. S. Bates, G. H. Fredrickson, *Annu. Rev. Phys. Chem.* **41**, 525 (1990).
20. Z.-R. Chen, J. A. Kornfield, S. D. Smith, J. T. Grothaus, M. M. Satkowski, *Science* **277**, 1248 (1997).
21. G. M. Whitesides, B. Grzybowski, *Science* **295**, 2418 (2002).
22. R. Y. F. Liu, Y. Jin, A. Hiltner, E. Baer, *Macromol. Rapid Commun.* **24**, 943 (2003).
23. R. Y. F. Liu, T. E. Bernal-Lara, A. Hiltner, E. Baer, *Macromolecules* **37**, 6972 (2004).
24. 1 Barrer = $10^{-10}\frac{cm^3(STP)cm}{cm^2\ s\ cmHg}$, where STP is standard temperature and pressure.
25. P. H. Geil, *Polymer Single Crystals* (Wiley-Interscience, New York, 1963).
26. E. L. Cussler, S. E. Hughes, W. J. Ward III, R. Aris, *J. Membr. Sci.* **38**, 161 (1988).
27. The PEO reflection labeled (032) actually contains overlapped reflections from (032), ($\bar{1}$32), (112), ($\bar{2}$12), ($\bar{1}$24), ($\bar{2}$04), and (004), which have similar *d*-spacing of ~0.39 nm, with that from (032) being the strongest. They cannot be easily separated in (032) pole figure. Details can be found in (*16*).

28. This research was supported by the NSF Center for Layered Polymeric Systems (grant DMR-0423914).

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5915/757/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Tables S1 and S2
References

13 August 2008; accepted 26 November 2008
10.1126/science.1164601

Nitrogen-Doped Carbon Nanotube Arrays with High Electrocatalytic Activity for Oxygen Reduction

Kuanping Gong,[1] Feng Du,[1] Zhenhai Xia,[2] Michael Durstock,[3] Liming Dai[1,4]*

The large-scale practical application of fuel cells will be difficult to realize if the expensive platinum-based electrocatalysts for oxygen reduction reactions (ORRs) cannot be replaced by other efficient, low-cost, and stable electrodes. Here, we report that vertically aligned nitrogen-containing carbon nanotubes (VA-NCNTs) can act as a metal-free electrode with a much better electrocatalytic activity, long-term operation stability, and tolerance to crossover effect than platinum for oxygen reduction in alkaline fuel cells. In air-saturated 0.1 molar potassium hydroxide, we observed a steady-state output potential of −80 millivolts and a current density of 4.1 milliamps per square centimeter at −0.22 volts, compared with −85 millivolts and 1.1 milliamps per square centimeter at −0.20 volts for a platinum-carbon electrode. The incorporation of electron-accepting nitrogen atoms in the conjugated nanotube carbon plane appears to impart a relatively high positive charge density on adjacent carbon atoms. This effect, coupled with aligning the NCNTs, provides a four-electron pathway for the ORR on VA-NCNTs with a superb performance.

The oxygen reduction reaction (ORR) at the cathode of fuel cells (*1*) plays a key role in controlling the performance of a fuel cell, and efficient ORR electrocatalysts are essential for practical applications of the fuel cells (*2*, *3*). The ORR can proceed either through (i) a four-electron process to combine oxygen with electrons and protons directly, when coupled with oxidation on the anode, to produce water as the end product, or (ii) a less efficient two-step, two-electron pathway involving the formation of hydrogen peroxide ions as an intermediate (*2*, *3*). Alkaline fuel cells with platinum-loaded carbon as an electrocatalyst for the four-electron ORR were developed for the Apollo lunar mission in the 1960s (*4*), but their large-scale commercial application has been precluded by the high cost of the requisite noble metals. Apart from its high cost, the Pt-based electrode also suffers from its susceptibility to time-dependent drift (*5*) and CO deactivation (*6*).

Recent intensive research efforts in reducing or replacing Pt-based electrode in fuel cells have led to the development of new ORR electrocatalysts, including Pt-based alloys (*7*), transition metal chalcogenides (*8*), carbon nanotube–supported metal particles (*9*–*11*), enzymatic electrocatalytic systems (*12*), and even conducting poly(3,4-

[1]Departments of Chemical and Materials Engineering, University of Dayton, 300 College Park, Dayton, OH 45469, USA. [2]Department of Mechanical Engineering, University of Akron, Akron, OH 44325, USA. [3]Materials and Manufacturing Directorate, Air Force Research Laboratory, RXBP, Wright-Patterson Air Force Base, OH 45433, USA. [4]Department of Chemistry and University of Dayton Research Institute and Institute for the Development and Commercialization of Advanced Sensor Technology and Wright Brothers Institute, Dayton, OH 45469, USA.

*To whom correspondence should be addressed. E-mail: ldai@udayton.edu

ethylenedioxythiophene) (PEDOT)–coated membranes (*13*). Apart from their use as the noble metal–catalyst supports (*9*–*11*), aligned carbon nanotubes formed by high-temperature treatment of certain metal heterocyclic molecules (e.g., ferrocene/NH_3) have been recently demonstrated to show some ORR electrocatalytic activities (*10*). The observed electrocatalytic activity was attributed to the presence of FeN_2-C and/or FeN_4-C active sites in the nanotube structure (*10*). However, we found in this study that vertically aligned nitrogen-containing carbon nanotubes (VA-NCNTs) produced by pyrolysis of iron(II) phthalocyanine (a metal heterocyclic molecule containing nitrogen) (*14*), in either the presence or absence of additional NH_3 vapor (*15*), could be used as effective ORR electrocatalysts, even after a complete removal of the residual Fe catalyst by electrochemical purification [supporting online material (SOM) (*16*)]. These metal-free VA-NCNTs were shown to catalyze a four-electron ORR process with a much higher electrocatalytic activity, lower overpotential (the difference between thermodynamic and formal potentials), smaller crossover effect, and better long-term operation stability than that of commercially available or similar platinum-based electrodes (C2-20, 20% platinum on Vulcan XC-72R; E-TEK) in alkaline electrolytes (*17*, *18*).

A cross-sectional scanning electron microscopy (SEM) view for the as-synthesized VA-NCNT array is shown in Fig. 1A. The observed zigzag-like path along the nanotube length can be attributed to the integration of nitrogen into the graphitic structure to alter the nanotube surface from a straight cylinder geometry (*10*). The presence of structural nitrogen was confirmed by x-ray photoelectron spectroscopic (XPS) measurements (fig. S2) (*16*), which show the incorporation of pyridinic-like (399 eV) and pyrrolic-like (401 eV) nitrogen atoms within the nanotube graphene sheets with a N:C atomic ratio in the range of ~4 to 6 atomic % (table S1) (*16*). The aligned structure remained largely unchanged after the electrochemical purification (*16*, *19*).

The corresponding transmission electron microscopy (TEM) image shows that, after electrochemical purification, the individual NCNTs are free from residual iron catalyst particles (Fig. 1B) (*16*). These NCNTs are about 8 μm long and 25 nm in outer diameter and exhibit a bamboo-like structure such as that seen for other CNTs (*14*). Figure 1C shows a digital photograph of the electrochemically purified VA-NCNT film after having

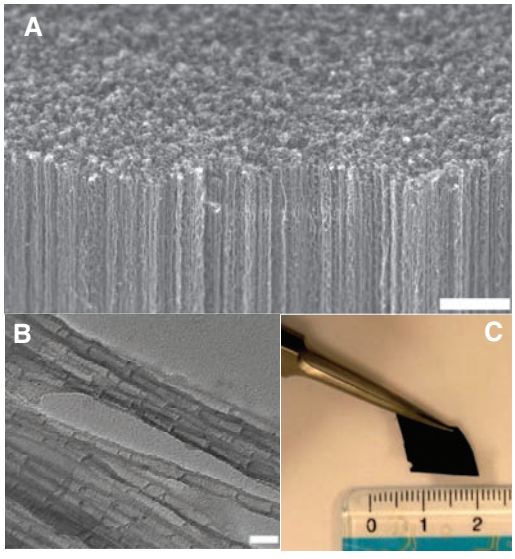


Fig. 1. (**A**) SEM image of the as-synthesized VA-NCNTs on a quartz substrate. (**B**) TEM image of the electrochemically purified VA-NCNTs. (**C**) Digital photograph of the VA-NCNT array after having been transferred onto a PS-nonaligned CNT conductive nanocomposite film (fig. S1) (*16*). Scale bars, 2 μm (A); 50 nm (B).

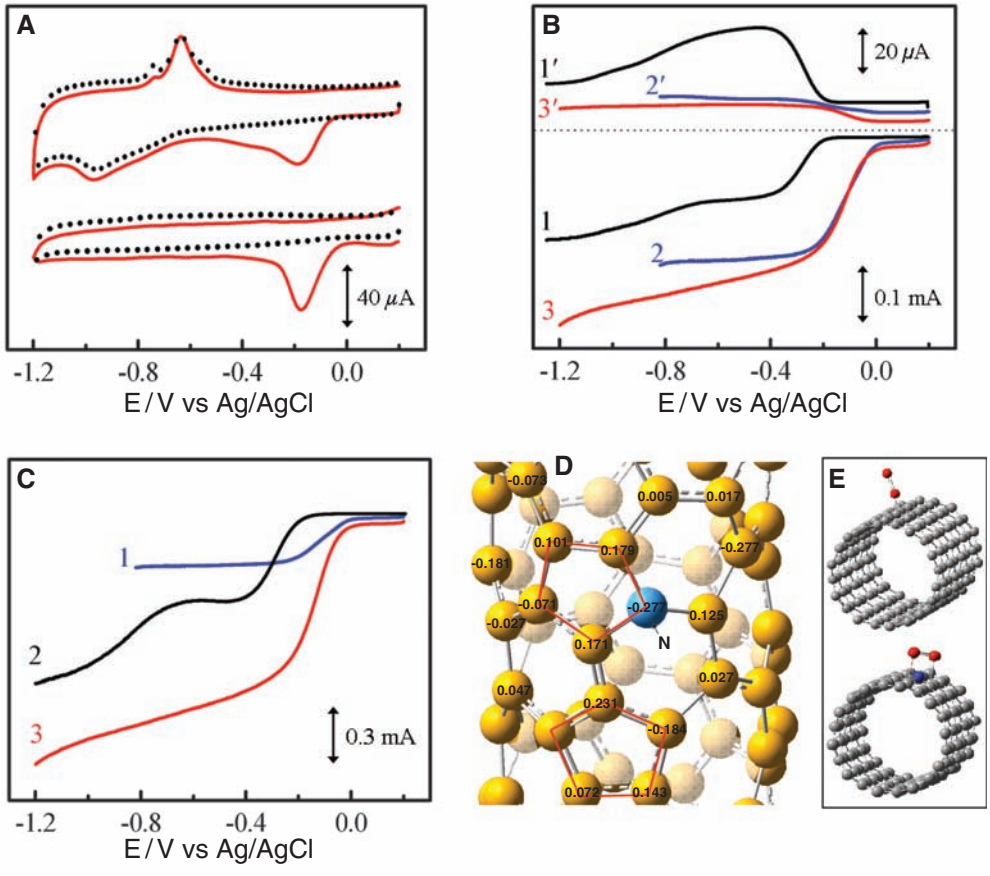


Fig. 2. (**A**) CVs for oxygen reduction at the unpurified (upper) and electrochemically purified (bottom) VA-NCNT/GC electrodes in the argon-protected (dotted curves) or air-saturated 0.1 M KOH (solid red curves) at the scan rate of 100 mVs^{-1}. Before the measurements, the unpurified VA-NCNT/GC electrode was repeatedly potentiodynamic swept from +0.2 V to −1.2 V in an Ar-protected 0.1 M KOH until a steady voltammogram curve was obtained. (**B**) RRDE voltammograms and the corresponding amperometric responses for oxygen reduction in air-saturated 0.1 M KOH at the NA-CCNT/GC (curves 1 and 1′), Pt-C/GC (curves 2 and 2′), and NA-NCNT/GC (curves 3 and 3′) electrodes at the scan rate of 10 mVs^{-1}. The electrode rotation rate was 1400 revolutions per minute (rpm), and the Pt ring electrode was poised at 0.5 V. (**C**) RRDE voltammograms for oxygen reduction in air-saturated 0.1 M KOH at the Pt-C/GC (curve 1), VA-CCNT/GC (curve 2), and VA-NCNT (curve 3) electrodes. Because of the technical difficulties associated with the sample mounting, amperometic responses with the Pt ring electrode were not measured for the vertically aligned carbon nanotubes. (**D**) Calculated charge density distribution for the NCNTs. (**E**) Schematic representations of possible adsorption modes of an oxygen molecule at the CCNTs (top) and NCNTs (bottom). The C atoms around the pyrrolic-like nitrogen could possess much higher positive charges than do the C atoms around the pyridinic-like nitrogen (fig. S6) (*16*).

been transferred onto a polystyrene (PS)–nonaligned carbon nanotube conductive composite film to be used to cover the active area of a glassy carbon electrode for ORR (fig. S1) (*16*). The size of the free-standing VA-NCNT film thus prepared is limited mainly by the size of the furnace used for the aligned nanotube growth.

The electrochemical purification was confirmed by voltammetric responses of the VA-NCNT electrode before and after electrochemical oxidation (*16*). As shown in Fig. 2A, the cyclic voltammogram (CV) of the glassy carbon–supported aligned unpurified VA-NCNT electrode (VA-NCNT/GC) (fig. S1) (*16*) in an aqueous solution of 0.1 M KOH under argon protection (top dot curve in Fig. 2A) exhibited two well-defined peaks at the potential of about –1.0 V and –0.65 V, arising from redox reactions associated with the Fe catalyst residues (*20*). In contrast, the corresponding voltammetric response for the same VA-NCNT/GC electrode after the electrochemical purification showed a featureless curve (bottom dot curve in Fig. 2A). This difference strongly suggests that the iron catalyst residues have been largely removed by the electrochemical oxidation, as also evidenced by XPS (fig. S2) (*16*) and thermogravimetric analyses (fig. S3) (*16*).

Also shown in Fig. 2A are CVs for oxygen reduction in 0.1 M KOH solution at the VA-NCNT/GC before (top solid red curve) and after (bottom solid red curve) the electrochemical purification. Unlike conventional carbon electrodes (*21*), a cathodic process with a rather high reduction potential (i.e., lower overpotential) of about –0.15 V was seen for the ORR at both the unpurified and electrochemically purified VA-NCNT/GC electrodes with some notable difference in the peak shape. To gain further insight on the ORR electrochemical procedures, we performed rotating ring-disk electrode (RRDE) voltammograms. Figure 2B shows the steady-state voltammograms for nitrogen-free nonaligned carbon nanotubes supported by a glassy carbon electrode (NA-CCNT/GC, curve 1), commercially available platinum-loaded carbon (Vulcan XC-72R) supported by a glassy carbon electrode (Pt-C/GC, curve 2), and glassy carbon–supported nonaligned nitrogen-containing carbon nanotubes (NA-NCNT/GC, curve 3) in air-saturated 0.1 M KOH electrolyte. The corresponding amperometric responses (curves 1′ to 3′) for the oxidation of hydrogen peroxide ions (HO_2^-) measured with a Pt ring electrode at the potential of 0.50 V (*16*) are also included in Fig. 2B. The NA-CCNT/GC electrode showed a two-step process for ORR with the onset potential of about –0.22 V and –0.70 V, respectively (curve 1 in Fig. 2B), consistent with previous reports (*8*, *21*). We attribute the first sharp step over –0.22 V to the two-electron reduction of O_2 to HO_2^-, as evidenced by a substantial concomitant increase in the oxidation current over about –0.22 to –0.41V at the ring electrode (curve 1′ in Fig. 2B). The subsequent gradual decrease in the ring current seen in curve 1′ of Fig. 2B corresponds to the reduced amount of HO_2^- reached to the ring electrode before being oxidized into HO^- at the disk electrode under the increased negative potentials.

Unlike the NA-CCNT/GC electrode, the NA-NCNT/GC electrode exhibited a one-step process for the ORR with the steady-state diffusion current that was almost twice that obtained at the NA-CCNT/GC electrode (curves 1 and 3 in Fig. 2B). Like the Pt-C/GC electrode, the observed one-step process suggests a four-electron pathway for the ORR at the NA-NCNT/GC electrode, as also supported by the corresponding negligible current for HO_2^- oxidation recorded at the Pt ring electrode (curve 3′ in Fig. 2B). The transferred electron number (n) per oxygen molecule involved in the ORR was calculated from Eq. 1 (*22*) to be 1.8 and 3.9 for the NA-CCNT/GC electrode (at the potential of –0.40 V) and the NA-NCNT/GC electrode (at the potential of –0.30 V), respectively.

$$n = 4I_D/(I_D + I_R/N) \quad (1)$$

where $N = 0.3$ is the collection efficiency (*16*), I_D is the faradic disk current, and I_R is the faradic ring current.

Figure 2C shows the steady-state voltammograms for Pt-C/GC (curve 1), VA-CCNT/GC (curve 2), and VA-NCNT/GC (curve 3) electrodes in the air-saturated 0.1 M KOH electrolyte.

The half-wave potentials for ORR at the NA-NCNT/GC (curve 3, Fig. 2B) and VA-NCNT/GC electrodes (curve 3 in Fig. 2C) are comparable to that at the Pt-C/GC electrode (–0.1 V), but a substantially enhanced steady-state diffusion current (~0.8 mA) was observed over a large potential range for the VA-NCNT/GC electrodes with respect to the Pt-C/GC electrode (Fig. 2C). Thus, the VA-NCNT/GC electrode is much better than the Pt-C/GC electrode for the ORR in an alkaline solution. Compared with the NA-NCNT/GC electrode, the better electrocatalytic performance of the VA-NCNT/GC electrode can be attributed to its well-defined large surface area with all of the nanotube top-ends falling on one plane at the interface between the aligned nanotube electrode and electrolyte solution to further facilitate the electrolyte/reactant diffusion (*23*, *24*). Similar current enhancement by the alignment structure was also observed for the VA-CCNT/GC electrode (curve 2 in Fig. 2C) with respect to its nonaligned counterpart (curve 1 in Fig. 2B), albeit with a relatively small effect and at a very high overpotential. The similar shape of the steady-state voltammograms for the aligned nanotube electrodes and their respective nonaligned counterparts indicates that the electrochemical mechanism for the ORR is insensitive to the alignment structure, although alignment can improve the electrokinetics. However, the much stronger diffusion-limited currents and lower overpotentials observed for the NCNT electrodes than their nitrogen-free counterparts indicate the importance of the nitrogen heteroatoms in the NCNTs (fig. S2) (*16*) to their electrocatalytic activities for the ORR.

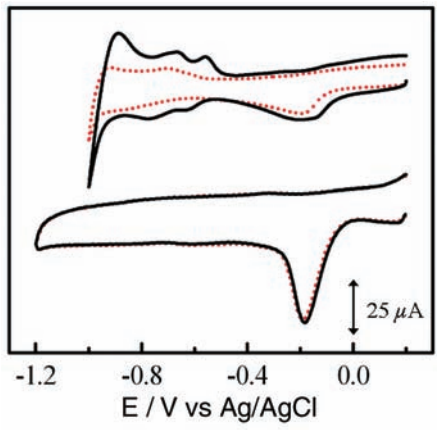


Fig. 3. CVs for the ORR at the Pt-C/GC (top) and VA-NCNT/GC (bottom) electrodes before (solid black curves) and after (dotted curves) a continuous potentiodynamic swept for ~100,000 cycles in an air-saturated 0.1 M KOH at room temperature (25 ± 1°C). Scan rate, 100 mV s^{-1}. The wavelike bands over −1.0 to −0.5 V seen for the pristine Pt-C/GC electrode are attributable to hydrogen adsorption/desorption.

Quantum mechanics calculations with B3LYP hybrid density functional theory (Gaussian 03) (*25*) indicate that carbon atoms adjacent to nitrogen dopants possess a substantially high positive charge density to counterbalance the strong electronic affinity of the nitrogen atom (Fig. 2D and figs. S5 and S6) (*16*). Together with the recent work on the metal-free conjugated PEDOT ORR electrode (*13*), this result prompted us to put forward an O_2 reduction mechanism on the NCNT electrodes. A redox cycling process reduces the carbon atoms that naturally exist in an oxidized form by the action of the electrochemical cycling, followed by reoxidation of the reduced carbon atoms to their preferred oxidized state upon O_2 absorption. The nitrogen-induced charge delocalization could also change the chemisorption mode of O_2 from the usual end-on adsorption (Pauling model) at the CCNT surface (top, Fig. 2E) to a side-on adsorption (Yeager model) onto the NCNT electrodes (bottom, Fig. 2E) (*26*). The parallel diatomic adsorption could effectively weaken the O–O bonding to facilitate ORR at the NCNT/GC electrodes. As such, doping carbon nanotubes with nitrogen heteroatoms as in the NCNT electrodes can efficiently create the metal-free active sites for electrochemical reduction of O_2.

To investigate the stability of the VA-NCNT/GC electrode toward ORR, we performed continuous potential cycling between +0.2 and –1.2 V for the VA-NCNT/GC electrode with Pt-C/GC as reference in air-saturated 0.1 M KOH for ~100,000 cycles. As can be seen in Fig. 3, the deterioration of Pt occurred, apart from the ORR, on the commercial Pt-C/GC (Vulcan XC-72) electrode (top curves) (*27*). The continuous po-

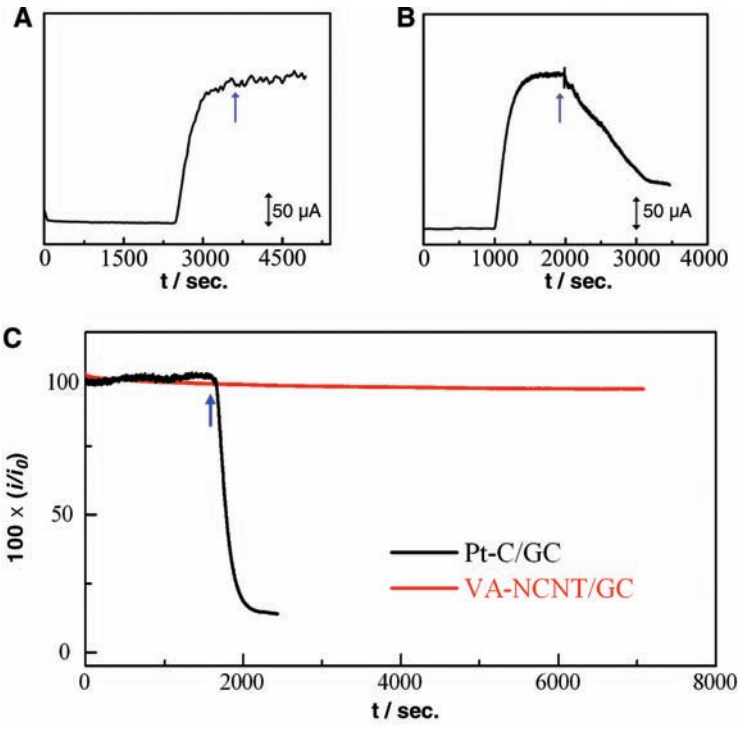


Fig. 4. (**A**) *i-t* chronoamperometric responses obtained at the VA-NCNT/GC electrode at −0.25 V in 0.1 M KOH under magnetic stirring (1000 rpm, MR 3001 K, Heidolph) and Ar-protection over 0 to 2400 s, followed by an immediate introduction of air. The arrow indicates the sequential addition of 3.0 M glucose, 3.0 M methanol, and 3.0 M formaldehyde, respectively, into the air-saturated electrochemical cell (fig. S1B) (*16*). (**B**) Corresponding *i-t* chronoamperometric response obtained at the Pt-C/GC electrode with the addition of 3.0 M methanol (as indicated by the arrow) after exposure to air at 1000 s under the same conditions as in (**A**) for comparison. (**C**) CO-poison effect on the *i-t* chronoamperometric response for the Pt-C/GC and VA-NCNT/GC electrodes. The arrow indicates the addition of 55 mL/min CO gas into the 550 mL/min O_2 flow; the mixture gas of ~9% CO (volume/volume) was then introduced into the electrochemical cell (fig. S1B) (*16*). i_0, initial current.

tential cycling might have caused migration/aggregation of the Pt nanoparticles and subsequent loss of the specific catalytic activity. In contrast, the VA-NCNT/GC electrode showed almost identical voltammetric responses before and after the continuous potential cycling under the same condition (bottom curves in Fig. 3).

The VA-NCNT/GC electrode was further subjected to testing the possible crossover and poison effects in the presence of fuel molecules (e.g., methanol) and CO, respectively (*27*–*29*). To examine the possible crossover effect, we measured electrocatalytic selectivity of the VA-NCNT/GC electrode against the electrooxidation of various commonly used fuel molecules, including hydrogen gas, glucose, methanol, and formaldehyde (Fig. 4A). For comparison, the corresponding current-time (*i-t*) chronoamperometric response for a Pt-C/GC electrode given in Fig. 4B shows a sharp decrease in current upon the addition of 3.0 M methanol. However, the strong and stable amperometric response from the ORR on the VA-NCNT/GC electrode remained unchanged after the sequential addition of hydrogen gas, glucose, methanol, and formaldehyde (Fig. 4A). Such high selectivity of the VA-NCNT/GC electrode toward the ORR and remarkably good tolerance to crossover effect for the VA-NCNT/GC electrode can be attributed to the much lower ORR potential than that required for oxidation of the fuel molecules (*30*) and is further evidenced by the cathodic polarization measurements in the presence and absence of 3.0 M methanol (fig. S4) (*16*).

The effect of CO on the electrocatalytic activity of the VA-NCNT/GC electrode was also tested because CO poisoning of most noble-metal electrodes is a major issue in the current fuel cell technology. Being nonmetallic, the VA-NCNT/GC electrode was insensitive to CO poisoning even after adding about 10% CO in oxygen (red line in Fig. 4C), whereas the Pt-C/GC electrode was rapidly poisoned under the same conditions (black curve in Fig. 4C). This result indicates that the VA-NCNT/GC electrode is free from residual iron particles, as iron-based catalysts would otherwise be readily poisoned by CO (*6*).

The high–surface area, good electrical and mechanical properties, and superb thermal stability intrinsically characteristic of aligned carbon nanotubes (*23*, *31*, *32*) provide additional advantages for the nanotube electrode to be used in fuel cells under both ambient and harsh conditions. Although cost depends on many factors, scarcity is not an intrinsic issue for carbon nanotubes. As the number of industrial-scale facilities for the relatively low-cost production of carbon nanotubes continues to grow, the price of carbon nanotubes is expected to further decrease (*33*). The role of nitrogen-doping demonstrated in this study could be applied to the design and development of various other metal-free efficient ORR catalysts, and these nitrogen-containing carbon nanotube electrodes are clearly of practical importance.

References and Notes

1. B. C. H. Steele, A. Heinzel, *Nature* **414**, 345 (2001).
2. S. Basu, Ed., *Recent Trends in Fuel Cell Science and Technology* (Springer, New York, 2007).
3. B. Sljukic, C. E. Banks, R. G. Compton, *J. Iranian Chem. Soc.* **2**, 1 (2005).
4. http://americanhistory.si.edu/fuelcells/alk/alk3.htm (copyright 2001, Smithsonian Institution).
5. X. Yu, S. Ye, *J. Power Sources* **172**, 145 (2007).
6. M. Winter, R. J. Brodd, *Chem. Rev.* **104**, 4245 (2004).
7. J. Zhang, K. Sasaki, E. Sutter, R. R. Adzic, *Science* **315**, 220 (2007).
8. K. Gong, P. Yu, L. Su, S. Xiong, L. Mao, *J. Phys. Chem. C* **111**, 1882 (2007).
9. G. Che, B. B. Lakshmi, E. R. Fisher, C. R. Martin, *Nature* **393**, 346 (1998).
10. J. Yang, D.-J. Liu, N. N. Kariuki, L. X. Chen, *Chem. Commun. (Cambridge)* 329 (2008).
11. A. Kongkanand, S. Kuwabata, G. Girishkumar, P. Kamat, *Langmuir* **22**, 2392 (2006).
12. J. P. Collman *et al.*, *Science* **315**, 1565 (2007).
13. B. Winther-Jensen, O. Winther-Jensen, M. Forsyth, D. R. MacFarlane, *Science* **321**, 671 (2008).
14. L. Dai *et al.*, *ChemPhysChem* **4**, 1150 (2003), and references therein.
15. L. S. Panchakarla, A. Govindaraj, C. N. R. Rao, *ACS Nano* **1**, 494 (2007).
16. Supporting materials are available on *Science* Online.
17. Z. Chen, M. Waje, W. Li, Y. Yan, *Angew. Chem. Int. Ed.* **46**, 4060 (2007).
18. S. H. Joo *et al.*, *Nature* **412**, 169 (2001).
19. M. Gao *et al.*, *Angew. Chem. Int. Ed.* **39**, 3664 (2000).
20. U. Casellato, N. Comisso, G. Mengoli, *Electrochim. Acta* **51**, 5669 (2006).
21. M. Zhang *et al.*, *Langmuir* **20**, 8781 (2004).
22. S. L. Gupta, D. Tryk, M. Daroux, W. Aldred, E. Yeager, in *Proceedings of the Symposium on Load Levelling and Energy Conservation in Industrial Processes*, D. Chin, Ed. (Electrochemical Society, Pennington, NJ, 1986), p. 207.
23. L. Dai, Ed., *Carbon Nanotechnology: Recent Developments in Chemistry, Physics, Materials Science and Device Applications* (Elsevier, Amsterdam, 2006).
24. L. Dai, *Aust. J. Chem.* **60**, 472 (2007), and references therein.
25. M. J. Fisch *et al.*, *Gaussian 03*, Revision B.04 (Gaussian, Pittsburgh, PA, 2003).
26. Z. Shi, J. Zhang, Z. Liu, H. Wang, D. P. Wilkinson, *Electrochim. Acta* **51**, 1905 (2006).
27. E. H. Yu, K. Scott, R. W. Reeve, *Fuel Cells (Weinheim)* **3**, 169 (2003).
28. The methanol oxidation on the anode is a competitive reaction to oxygen reduction at the cathode in a direct methanol fuel cell. Because the oxidation of fuels on noble-metal electrodes often commences at a potential lower than that of the ORR under the working conditions, methanol crossover from the anode to the cathode, if not eliminated, could diminish the cathodic performance through the depolarizing effect. Furthermore, it may even cause the whole fuel-cell system to be paralyzed through poisoning the ORR electrocatalyst by CO-like species generated as intermediates of the methanol oxidation at the cathode (*29*).
29. A. S. Arico, S. Srinivasan, V. Antonucci, *Fuel Cells (Weinheim)* **1**, 133 (2001).
30. H.-F. Cui, J.-S. Ye, X. Liu, W.-D. Zhang, F.-S. Sheu, *Nanotechnology* **17**, 2334 (2006).
31. J. J. Gooding *et al.*, *Electrochem. Commun.* **9**, 1677 (2007).
32. V. H. Ebron *et al.*, *Science* **311**, 1580 (2006).

33. The price for multiwall carbon nanotubes has dropped from a few hundred U.S. dollars per gram in the 1990s to about 100 U.S. dollars per kilogram. See, for example, http://www.azonano.com/details.asp?ArticleID=1108.
34. We gratefully acknowledge financial support for this work from the Air Force Office of Scientific Research (grant FA9550-06-1-0384). K.G. thanks L. Qu and J. Zhu for help with some initial work.

Supporting Online Material
www.sciencemag.org/cgi/content/full/323/5915/760/DC1
Materials and Methods
Figs. S1 to S6
Table S1
References

4 November 2008; accepted 16 December 2008
10.1126/science.1168049

Anomalously Metal-Rich Fluids Form Hydrothermal Ore Deposits

Jamie J. Wilkinson,[1,2]*† Barry Stoffell,[1]‡ Clara C. Wilkinson,[1,2]† Teresa E. Jeffries,[2] Martin S. Appold[3]

Hydrothermal ore deposits form when metals, often as sulfides, precipitate in abundance from aqueous solutions in Earth's crust. Much of our knowledge of the fluids involved comes from studies of fluid inclusions trapped in silicates or carbonates that are believed to represent aliquots of the same solutions that precipitated the ores. We used laser ablation inductively coupled plasma mass spectrometry to test this paradigm by analysis of fluid inclusions in sphalerite from two contrasting zinc-lead ore systems. Metal contents in these inclusions are up to two orders of magnitude greater than those in quartz-hosted inclusions and are much higher than previously thought, suggesting that ore formation is linked to influx of anomalously metal-rich fluids into systems dominated by barren fluids for much of their life.

Hydrothermal ore deposits, formed from the flow of hot solutions through porous or fractured rocks, are the principal source of metals in Earth's crust (*1*). Such large accumulations of metal require concentration of elements hundreds or thousands of times above natural abundance, implying high-mass fluxes through small volumes of rock coupled with efficient precipitation. A fundamental control on the formation of hydrothermal deposits is the ability of the fluid to carry metals in solution (*2*). Yet, paradoxically, for most deposit types formed at low-to-intermediate temperatures, both direct analysis of fluid inclusions and theoretical calculation indicate that the concentrations of dissolved metals are likely to be low, on the order of tens of parts per million (*3*). Also, samples of modern crustal fluids, such as those from oil fields or mid-ocean ridges, typically contain only a few parts per million of Cu, Zn, and Pb (*4*, *5*), although there are exceptions, such as the Salton Sea geothermal brines in California (*6*) and oilfield waters from central Mississippi (*4*). A consequence is that the other parameters that govern total metal flux in ore formation (average flow velocity and system lifetime) tend toward their likely geological limits in both numerical simulations and empirical models based on geological and geochronological constraints (*7*, *8*). As a result, it has been suggested that higher-than-normal concentrations of metal in fluids may be required to form large ore bodies (*9*).

For several decades, a key source of information on the physical and chemical conditions of hydrothermal ore formation has been fluid inclusions trapped during mineral growth (*10*). In most deposits, metalliferous ore minerals (commonly opaque sulfides) occur together with uneconomic transparent phases (gangue). Because fluid inclusions in the opaque phases are not easily studied by traditional transmitted light microscopy and microanalytical methods, the nature of ore-forming fluids and the conditions of ore-mineral precipitation have often been inferred from the properties of inclusions trapped in the associated gangue minerals. However, it is often difficult to provide unequivocal evidence for coprecipitation based on textural observations or isotopic measurements (*11*); consequently, uncertainty remains concerning the temporal and, therefore, genetic relationship between gangue-hosted inclusions and the ore-forming process. Several studies that used infrared light microscopy to observe inclusions in opaque minerals such as wolframite and cassiterite have shown that the properties of these fluid inclusions may, indeed, be different (*12*).

We analyzed fluid inclusions in sphalerite (ZnS) from two zinc-lead ore systems with the use of laser ablation inductively coupled plasma mass spectrometry (LA-ICPMS). Primary inclusions in sphalerite must represent the ore-forming fluid because they are trapped during growth of the ore mineral itself. Unlike bulk analytical studies that are limited to a few major elements and that may sample multiple populations of inclusions (*13*), LA-ICPMS allows determination of trace elements (including ore metals) in single, texturally constrained inclusions.

We selected samples from two well-studied hydrothermal ore systems. The Northern Arkansas district of the Ozark Plateau, North America, is an example of low-temperature Mississippi Valley–Type (MVT) zinc-lead mineralization, thought to have formed by continent-scale basinal brine migration (*14*). The Midlands Basin orefield in Ireland contains several large zinc-lead(-barium) ore deposits formed from moderate temperature fluids generated by deep crustal circulation of seawater-derived brines during continental rifting (*15*, *16*). Both systems are economically noteworthy but provide a contrast in terms of sources of metals, sulfur, and hydrological regime. Lead is of particular interest because it needs to be concentrated above average crustal abundance more than any other common ore-forming element (~4000 times) to form a potentially economic accumulation.

Samples from Northern Arkansas were collected from exposures in the Monte Cristo and Philadelphia Mines of the Rush subdistrict and from the Lucky Dog Mine of the Tomahawk Creek subdistrict. They comprise fine- to coarse-grained crystalline quartz and medium- to coarse-grained pale yellow–to–brown sphalerite. Regionally, precipitation of sphalerite typically overlapped with that of jasperoid and finely crystalline quartz, and more coarsely crystalline quartz formed later (*17*). Samples from Ireland were collected from historic mine exposures and drill core from the Silvermines deposit, as well as from quarry outcrop of quartz-sulfide veins nearby. The deposit samples are composed of massive sulfide dominated by coarse-grained brown sphalerite that replaces early disseminated granular and framboidal pyrite. The vein sample is composed of quartz and ankerite, as well as minor sphalerite and galena, and was selected as a representative example of a regionally developed set of feeder veins developed underneath the ore deposits (*18*, *19*).

Salinity data derived from freezing experiments (*20*) show that the Northern Arkansas mineralization formed from brines, typical of MVT deposits (Fig. 1). Assuming fluids were trapped at hydrostatic pressure at depths of <2 km, the inferred depth of ore formation (*14*), we calculated an isochoric correction of <+10°C to recorded homogenization temperature (T_h) values to give true trapping temperatures. Thus, T_h can be regarded as a reasonable approximation of fluid temperature during mineral growth. Inclusions from the Irish samples display lower salin-

[1]Department of Earth Science and Engineering, Imperial College London, South Kensington Campus, Exhibition Road, London SW7 2AZ, UK. [2]Department of Mineralogy, Natural History Museum, Cromwell Road, London SW7 5BD, UK. [3]Department of Geological Sciences, University of Missouri–Columbia, 101 Geological Sciences Building, Columbia, MO 65211, USA.

*To whom correspondence should be addressed. E-mail: j.wilkinson@imperial.ac.uk
†Present address: Australian Research Council Centre of Excellence in Ore Deposits (CODES), Private Bag 126, University of Tasmania, Hobart, Tasmania 7001, Australia.
‡Present address: Rio Tinto Mining and Exploration Limited, 2 Eastbourne Terrace, London W2 6LG, UK.

ity and higher T_h values than the MVT fluids (Fig. 1), typical for the Irish orefield (*11*). For the Irish ores, which formed at shallow depth (*16*, *19*), any correction to homogenization temperatures will again be small so that T_h values can be regarded as a good proxy for fluid trapping temperature.

Laser ablation analyses were carried out with the use of a New Wave UP213AI, 213-nm aperture-imaged laser ablation system (*20*–*22*) on primary fluid inclusions interpreted to have formed during initial mineral growth based on conventional petrographic criteria (*10*). Some secondary inclusions, formed during later fracturing and annealing, were analyzed for comparison (Fig. 2). Lead and other elements of interest in hydrothermal systems such as Ba and Mn are clearly present in the fluid phase, as indicated by their good correlation with Cl in inclusion signals (Fig. 3). Full data are reported in tables S1 and S2 (*20*).

Zn concentrations in primary quartz-hosted inclusions from Northern Arkansas are low, ranging from 0.12 to 12.3 parts per million (ppm). In Irish quartz, except for one primary inclusion (14 ppm), Zn was below the limit of detection (mean = 37.6 ppm) because of the small inclusion size. Such low Zn concentrations are consistent with previously reported bulk analyses of 3.4 to 6.0 ppm (*16*). Unfortunately, we were not able to measure the Zn concentration of inclusions trapped in sphalerite because of the overwhelming host mineral contribution to the laser ablation signal. A similar problem also occurred for Cu (Fig. 3). However, Pb can be used as an indicator of the ore metal content of these inclusions, as it does not substitute appreciably into sphalerite, and its presence there can be corrected for (*20*).

In Northern Arkansas, Pb concentrations display a marked bimodal distribution, ranging from 0.2 to 3.5 ppm in quartz and primarily from 10 to 400 ppm in sphalerite (Fig. 4). This excludes five sphalerite-hosted inclusions that fall in the lower population, interpreted to represent unrecognized secondary inclusions trapping fluid related to the later quartz. The quartz-hosted inclusion data are consistent with a 266-nm LA-ICPMS study that found that fluid inclusions in gangue minerals from the Southeast Missouri MVT district contained Pb, Zn, and Cu concentrations below instrumental detection limits of ~10 ppm (*23*). The Monte Cristo and Philadelphia sphalerites have similar mean Pb concentrations (80 ppm) that are lower than the Lucky Dog sphalerites (119 ppm).

In Ireland, we also observed a distinction between the Pb content of primary inclusions in quartz (3.6 to 26 ppm) and sphalerite (22 to 890 ppm) (Fig. 4). The quartz-hosted inclusion data are consistent with bulk fluid inclusion analyses for Irish feeder veins that gave Pb concentrations of 11.4 to 19.8 ppm (*16*). The more saline primary inclusions in sphalerite 75-85-104 have a higher mean Pb concentration (430 ppm) than the

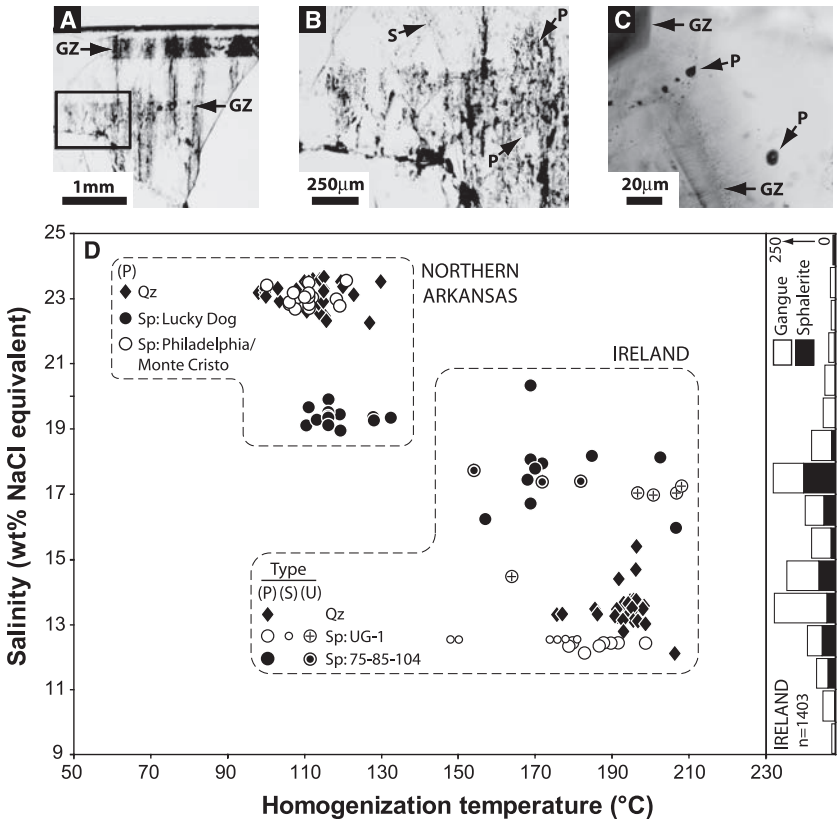


Fig. 1. (**A**) Quartz wafer from Northern Arkansas showing primary growth zones (GZ). (**B**) Magnification of inset shown in (A) illustrating complex distribution of fluid inclusions, together with some secondary trails (S) and selected primary inclusions (P) within growth clusters. (**C**) Sphalerite wafer from Ireland showing growth zones defined by fine fluid inclusions and color banding, together with euhedral primary inclusions. (**D**) Plot of fluid temperature and salinity data derived from microthermometry. Salinity was estimated from the freezing point depression of ice, modeled in the NaCl-H_2O system (*20*). For Northern Arkansas [data previously shown in (*17*)], quartz from Monte Cristo and sphalerite (Sp) from Monte Cristo and Philadelphia contain apparently identical primary inclusions. Slightly lower salinity primary inclusions are found in sphalerite from the Lucky Dog mine (20 km to the southwest), indicating geographic variability in brine composition. In the Irish Orefield, UG-1 sphalerite contains primary and secondary inclusions of a less saline brine [12 to 15 weight percent (wt %) NaCl equivalent] and a trail of inclusions of more saline fluids (16 to 18 wt % NaCl equivalent). This cuts a growth zone boundary (Fig. 2), indicating that these fluids are younger than this surface and its associated primary inclusions, but the exact timing with respect to other inclusions in the sample is uncertain (U). Fluids in the more saline population are trapped as primary inclusions in sphalerite 75-85-104. The data display a bimodal salinity distribution that mirrors the distribution observed in regional-scale fluid inclusion studies (see histogram at right), suggesting that these modes reflect multiple pulses of districtwide flow affecting a rock volume estimated at >130,000 km^3. Analogous brine pulses have been inferred in the recent history of the Salton Trough geothermal field (*6*). The evidence noted above, together with crosscutting relations observed in other samples, suggests that the higher salinity fluid pulse is later and is associated with the majority of the sphalerite in the district. The inclusions analyzed by LA-ICPMS in sphalerite sample 75-85-104 are not plotted because homogenization experiments could not be carried out due to problems with leakage.

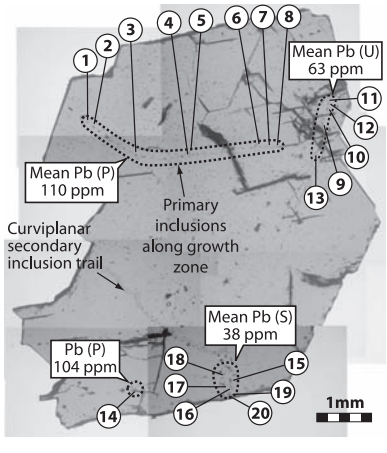


Fig. 2. Transmitted light digital photomontage of double-polished fluid inclusion wafer (~100 μm thick) of Irish sphalerite UG-1. Shown are individual fluid inclusions analyzed (numbered) and average determined Pb concentrations for each population of primary, secondary, or uncertain inclusions (same as those referred to in Fig. 1).

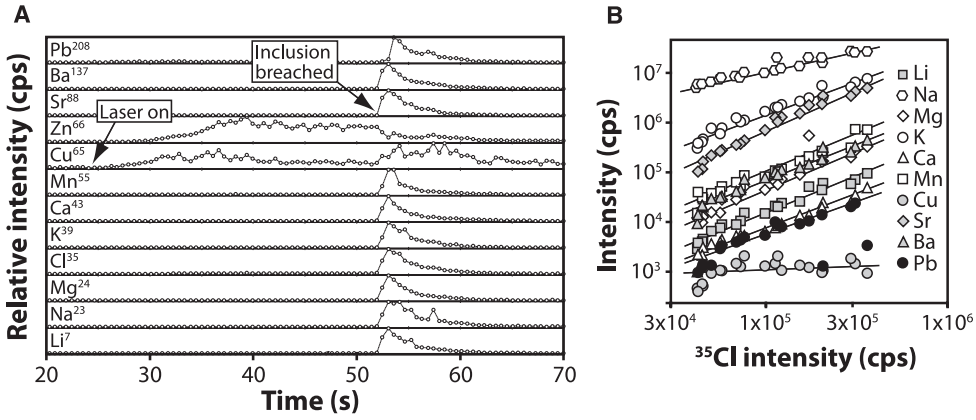


Fig. 3. (**A**). Example of a time-resolved laser ablation spectrum for a fluid inclusion in sphalerite (primary inclusion 9, Irish sphalerite 75-85-104). Initially, gas background was acquired and then the laser was turned on at 24 s. Signals for ^{66}Zn and ^{65}Cu increase as sphalerite begins to be ablated. The inclusion was breached at ~52 s. The signal was integrated offline over an ~16-s interval. The *y* axis is scaled from maximum to minimum recorded counts per second (cps) for each individual isotope. (**B**) The good correlations between the intensity for ^{35}Cl (only present in the fluid inclusion) and all isotopes measured (except ^{65}Cu) through the integration interval confirms their predominance in the fluid phase. This analysis returned 160 ppm Li, 530 ppm Mg, 7740 ppm K, 17500 ppm Ca, 530 ppm Mn, 2750 ppm Sr, 1770 ppm Ba, and 61 ppm Pb.

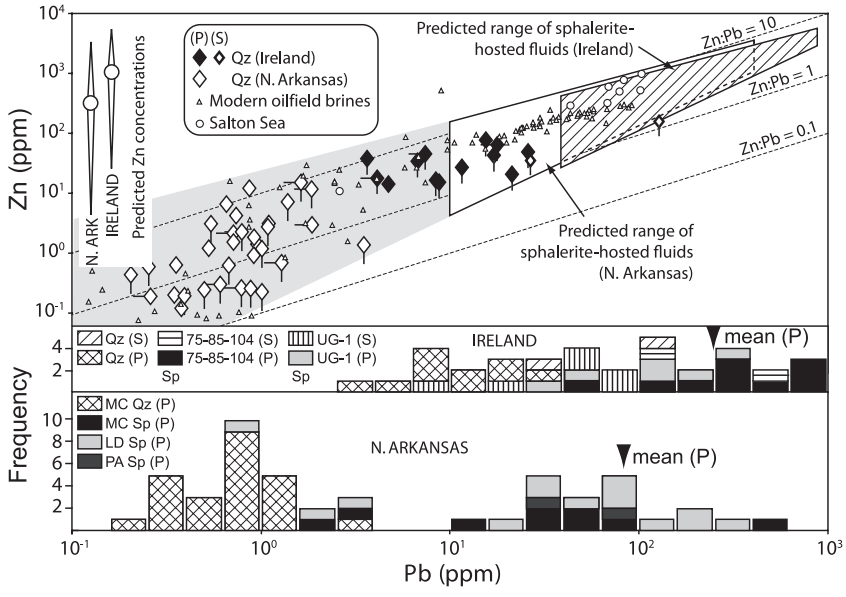


Fig. 4. Concentrations of Zn versus Pb in quartz-hosted fluid inclusions and histograms showing measured Pb concentrations in quartz- and sphalerite-hosted inclusions, compared with natural brine data from broadly analogous environments. Predicted Pb-Zn covariation for sphalerite-hosted inclusions indicated by fields, based on the extrapolation of empirical Pb-Zn concentrations, including data from modern oil-field waters (*24*, *25*) adjusted to mass/mass units and from the Salton Sea geothermal field (*6*). Both of these data sets have values that fall in our "high-metal" fluid fields, suggesting that they are anomalously metalliferous and viable ore-forming fluids. Typical Zn:Pb ratios (by mass) for natural fluids are in the range of 1 to 10 (dashed lines) and tend toward higher values at elevated concentrations. Five data points from the Northern Arkansas data set with low detected Pb are inferred secondary inclusions. Short bars on symbols indicate that the plotted value is a maximum (limit of detection) value for that element. Qz, quartz; Sp, sphalerite; MC, Monte Cristo; PA, Philadelphia; LD, Lucky Dog. Some of the Northern Arkansas data shown are presented in (*17*), figures 6 and 7.

inclusions in sphalerite UG-1 (120 ppm); secondary inclusions in both samples display similar, lower concentrations. Two secondary inclusions in quartz contained higher metal concentrations (27 to 128 ppm) (Fig. 4) than the primary inclusions, indicating overprinting by later, more metalliferous fluids such as those trapped in sphalerite.

Pb and Zn concentrations are commonly correlated in hydrothermal solutions as a result of their similar geochemical behavior and potential buffering by their respective sulfides (*4*, *24*). This is illustrated by their empirical covariation in modern basin brines (*4*, *6*, *24*–*26*) and in the quartz-hosted inclusions in this study (above ~1 ppm) where both metals were determined (Fig. 4). This relation allows us to model the likely range of Zn concentrations to be expected in the sphalerite-hosted inclusions at the time of trapping: up to 5000 ppm in Irish sphalerite and 3000 ppm in Northern Arkansas. The latter estimate is ~60% of the value predicted theoretically from thermodynamic data using measured Pb concentrations, assuming galena and sphalerite saturation and making reasonable assumptions about pH and oxygen fugacity (*17*). Although these are only order-of-magnitude estimates, it is clear that the fluids precipitating sphalerite were markedly enriched in Pb and Zn compared with those precipitating quartz in the two systems.

It could be argued that the fluid inclusions in quartz represent spent ore fluids, trapped after sulfide precipitation had already taken place. This is thought to be unlikely in the Irish deposits because the quartz was sampled from largely barren vein systems that formed beneath the ore deposits, in hydrothermal upflow zones (*16*), and the fluids show no signs of having substantially cooled or mixed (Fig. 1), as is known to occur during mineralization (*16*, *19*). In the Tri-State MVT district, we found sphalerite- and quartz-hosted brine inclusions to have distinct halogen signatures, indicating that the fluids had separate origins (*17*). The fertile ore fluids appear to have originated during strong evaporation of seawater at Earth's surface, before later burial and expulsion. This observation contrasts with the barren fluids that evolved from less strongly evaporated seawater (*17*). The metalliferous fluids are therefore linked to the paleoclimate and the specific characteristics of the sedimentary aquifer within which they were trapped. The development of low-pH surface brines in the U.S. mid-continent in the Permian via sulfide oxidation (*27*) is an intriguing possible origin for such metal-prone fluids. Not only will unusually low pH enhance metal solubilization, but the reservoir rocks will be depleted in reduced sulfur that would otherwise limit metal take-up. The nonideal behavior of Cl at high ionic strengths, coupled with low pH and chloride complexing, has been cited as a control of high Zn and Pb concentrations in the most saline modern oil-field waters (*6*), but this does not account for the low metal concentrations observed in similar salinity brines in some oil fields or those trapped in gangue minerals in this study (Fig. 4). Irrespective of the origin, both numerical models (*8*) and empirical observation

of brine heterogeneity in modern sedimentary basins (*6*) imply inefficiency of mixing and the potential for preservation of individual, metal-charged brine reservoirs that could be tapped at some later time.

In the case of Ireland, the origin of metal-enriched fluids is uncertain, although a deepening convective flow system (*15*) has the capability to extract higher concentrations of metal later in the life of the system due to increasing temperature and possibly also progressive exhaustion of the buffer capacity for pH (by feldspar-mica) or the activity of H_2S (by pyrite-Fe silicate) on the convective flow path. The observation that the texturally later brines have higher metal contents is consistent with this model, although their higher salinity is also likely to have contributed to enhanced metal transport. The high Ba content of the metalliferous fluids (up to 6000 ppm) indicates that the oxidized sulfur content must have been low, limited by barite saturation. Combined with high base metal concentrations that imply low reduced sulfur concentrations, we conclude that a key property of these fluids was low ΣS (total sulfur concentration).

High metal concentrations may pertain in other types of hydrothermal ore systems, such as epithermal or volcanic-hosted massive sulfide deposits. In these environments, periodic injection of metalliferous magmatic fluids may be responsible for the bulk of metal introduction (*28*–*30*) into systems otherwise dominated by barren geothermal waters. A number of large, high-temperature, magmatic-hydrothermal deposits are also known to have formed from magmatic fluids that contained very high concentrations of ore metals (*31*–*33*). Accepting that hydrothermal ores may form specifically from anomalously metal-rich batches of fluid implies geochemically specialized source regions and an episodicity and potentially short duration of ore-forming events that may be controlled by changes in hydrology. Although the existence of an efficient trap for these metals remains a fundamental prerequisite for hydrothermal ore formation, our interpretation contrasts with the view that many crustal fluids are viable ore fluids subject to the right perturbations in physicochemical conditions to cause efficient deposition (*24*).

References and Notes

1. H. L. Barnes, A. W. Rose, *Science* **279**, 2064 (1998).
2. D. L. Huston, *J. Aust. Geol. Geophys* **17**, 15 (1998).
3. T. M. Seward, H. L. Barnes, in *Geochemistry of Hydrothermal Ore Deposits*, H. L. Barnes, Ed. (Wiley, New York, ed. 3, 1997), pp. 435–486.
4. J. S. Hanor, in *Carbonate-Hosted Lead-Zinc Deposits*, Society of Economic Geologists Spec. Publ. 4 (Society of Economic Geologists, Littleton, CO, 1996), pp. 483–500.
5. K. L. Von Damm, *Annu. Rev. Earth Planet. Sci.* **18**, 173 (1990).
6. M. A. McKibben, L. A. Hardie, in *Geochemistry of Hydrothermal Ore Deposits*, H. L. Barnes, Ed. (Wiley, New York, ed. 3, 1997), pp. 877–936.
7. G. Garven, M. S. Appold, V. I. Toptygina, T. J. Hazlett, *Hydrogeol. J.* **7**, 108 (1999).
8. M. S. Appold, G. Garven, *Econ. Geol.* **95**, 1605 (2000).
9. C. J. Hodgson, in *Giant Ore Deposits*, Society of Economic Geologists Spec. Publ. 2 (Society of Economic Geologists, Queens Univ., Golden, CO, 1993), pp. 1–2.
10. E. Roedder, *Fluid Inclusions*, *Reviews in Mineralogy*, vol. 12 (Mineralogical Society of America, Washington, DC, 1984).
11. J. J. Wilkinson, *Lithos* **55**, 229 (2001).
12. A. R. Campbell, K. S. Panter, *Geochim. Cosmochim. Acta* **54**, 673 (1990).
13. J. G. Viets, A. H. Hofstra, P. Emsbo, in *Carbonate-Hosted Lead-Zinc Deposits*, Society of Economic Geologists Spec. Publ. 4 (Society of Economic Geologists, Littleton, CO, 1996), pp. 465–482.
14. D. L. Leach, in *Sediment-Hosted Zn-Pb Ores* (Springer, Berlin, 1994), pp. 104–138.
15. M. J. Russell, in *Geology and Genesis of Mineral Deposits in Ireland* (Irish Association for Economic Geology, Dublin, 1986), pp. 545–554.
16. J. J. Wilkinson, C. E. Everett, A. J. Boyce, S. A. Gleeson, D. M. Rye, *Geology* **33**, 805 (2005).
17. B. Stoffell, M. S. Appold, J. J. Wilkinson, N. A. McClean, T. E. Jeffries, *Econ. Geol.* **103**, 1411 (2008).
18. C. E. Everett, J. J. Wilkinson, D. M. Rye, in *Fractures, Fluid Flow and Mineralization*, Geological Society of London Spec. Publ. 155 (Geological Society of London, London, 1999), pp. 247–276.
19. I. M. Samson, M. J. Russell, *Econ. Geol.* **82**, 371 (1987).
20. Materials and methods are available as supporting material on *Science* Online.
21. B. Stoffell, J. J. Wilkinson, T. E. Jeffries, *Am. J. Sci.* **304**, 533 (2004).
22. T. E. Jeffries, S. E. Jackson, H. P. Longerich, *J. Anal. At. Spectrom.* **13**, 935 (1998).
23. M. S. Appold, T. J. Numelin, T. J. Shepherd, S. R. Chenery, *Econ. Geol.* **99**, 185 (2004).
24. B. Yardley, *Econ. Geol.* **100**, 613 (2005).
25. A. B. Carpenter, M. L. Trout, E. E. Pickett, *Econ. Geol.* **69**, 1191 (1974).
26. Y. K. Kharaka *et al.*, *Appl. Geochem.* **2**, 543 (1987).
27. K. C. Benison, R. H. Goldstein, B. Wopenka, R. C. Burruss, J. D. Pasteris, *Nature* **392**, 911 (1998).
28. S. F. Simmons, K. L. Brown, *Science* **314**, 288 (2006).
29. C. A. Heinrich, *Science* **314**, 263 (2006).
30. T. Ulrich, S. D. Golding, B. S. Kamber, K. Zaw, *Ore Geol. Rev.* **22**, 61 (2003).
31. T. Ulrich, D. Günther, C. A. Heinrich, *Nature* **399**, 676 (1999).
32. A. C. Harris, V. S. Kamenetsky, N. C. White, E. van Achterbergh, C. G. Ryan, *Science* **302**, 2109 (2003).
33. B. Rusk, M. Reed, J. H. Dilles, L. Klemm, C. A. Heinrich, *Chem. Geol.* **210**, 173 (2004).

34. This work was supported by an Imperial College Albert Julius Bursary to B.S. and National Environment Research Council grant GR9/03047. We thank B. Coles and R. Garcia-Sanchez for laboratory support and the Natural History Museum–Imperial College London Joint Analytical Facility for access to instrumentation. The constructive comments of three anonymous reviewers are appreciated.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5915/764/DC1
Materials and Methods
Tables S1 and S2
References

8 August 2008; accepted 4 November 2008
10.1126/science.1164436

Signature of the End-Cretaceous Mass Extinction in the Modern Biota

Andrew Z. Krug,[1] David Jablonski,[1] James W. Valentine[2]

The long-term effects of mass extinctions on spatial and evolutionary dynamics have been poorly studied. Here we show that the evolutionary consequences of the end-Cretaceous [Cretaceous/Paleogene (K/Pg)] mass extinction persist in present-day biogeography. The geologic ages of genera of living marine bivalves show a significant break from a smooth exponential distribution, corresponding to the K/Pg boundary. The break reflects a permanent increase in origination rates, intermediate between the Mesozoic rate and the post-extinction recovery pulse. This global rate shift is most clearly seen today in tropical bioprovinces and weakens toward the poles. Coupled with the modern geographic distributions of taxa originating before and after the K/Pg boundary, this spatial pattern indicates that tropical origination rates after the K/Pg event have left a permanent mark on the taxonomic and biogeographic structure of the modern biota, despite the complex Cenozoic history of marine environments.

The major extinctions of the geologic past, each of which removed >50% of well-preserved genera and perhaps >70% of their species (*1*, *2*), irreversibly restructured the taxonomic composition of the global biota. Although the broad macroevolutionary consequences of mass extinctions are well known (as in the dinosaurs-mammals changeover), their long-term effects on the temporal and spatial dynamics of clades and biotas are rarely investigated. For example, the good fit of modern biodiversity to local temperatures and to refugia from recent glaciations (*3*, *4*) might imply that the recovery from the most recent major extinction, at the end of the Cretaceous Period 65 million years ago (Ma) [the Cretaceous/Paleogene (K/Pg) extinction], is largely obscured by subsequent events.

Here, we show the lasting influence of the K/Pg mass extinction on the evolutionary and biogeographic structure of modern biotas, using backward survivorship curves (BSCs, also called prenascence curves), which plot the proportion of taxa within a cohort that originated before some reference time (*5*, *6*). Assuming that rates are time-specific and taxonomically homogeneous (*5*), a BSC defines an exponential probability function governed only by the origination rate (λ) of the cohort (*7*). The slope of a BSC is insen-

[1]Department of Geophysical Sciences, University of Chicago, 5734 South Ellis Avenue Chicago, IL 60637, USA. [2]Department of Integrative Biology and Museum of Paleontology, University of California, Berkeley, Berkeley, CA 94720, USA.

sitive to variations in extinction rate, so any significant changes in that slope signal major shifts in the origination rate (*5*). BSCs are particularly useful here because they let us (i) estimate intrinsic rates within bivalve cohorts, as opposed to a building a time series of individual events; and (ii) analyze the structure of age-frequency distributions in modern biotic provinces in comparison to the global distributions.

We analyzed (*7*) a taxonomically standardized, spatially explicit, global database of shallow-water living marine bivalves (*8*, *9*), which now includes 73 families (table S1), 854 genera and subgenera (711 with a fossil record), 5132 species, and 28,264 spatial occurrences from 322 localities (fig. S1). Marine bivalves are one of the few groups that allow a direct analysis of long-term evolutionary and biogeographic dynamics: They have a rich and spatially extensive fossil record, their systematics are increasingly well known and standardized, and imperfections in their record are increasingly understood in terms of preservational controls (*10*–*15*). Their diversity patterns today and through the Cenozoic correspond closely to those of other major marine clades, both spatially and temporally (*16*, *17*), and so they are a robust model system for probing the evolutionary and biogeographic history of the global biota. Localities were grouped into provinces according to sources cited in (*7*). BSCs were constructed (*5*, *18*, *19*) using the geologic ages of genera in an extensively revised and updated version of Sepkoski's compendium (*8*, *10*, *20*). We tested the fit of one-parameter, three-parameter, and five-parameter probability functions (*7*) to the global and provincial data, using the *optim* function in the software package R.

The BSC for extant bivalves globally shows a distinct inflection point at ~65 Ma (Fig. 1, A and B), with an increase in origination rate (λ) after the mass extinction (from 0.015 to 0.029 genera per lineage per million years) that continues to the present. A three-parameter model, incorporating two probability functions (with rate parameters λ_1 and λ_2; see methods in supporting online material) describing the BSC on either side of the inflection point (t_{crit}), is statistically supported over a single exponential fit (table S1). The optimizations place the inflection point near the K/Pg boundary, maximally supporting an age of 69 million years (My), not distinct from the K/Pg boundary, given the resolution of our binning scheme (*7*); the difference between the support of a critical age of 65 versus 69 My is not statistically meaningful (fig S2). This slight offset may also reflect less intense study in the 75- to 100-My interval, which could artificially concentrate originations in the last 10 My of the Cretaceous (*2*). This result indicates a significant and permanent increase in the origination rate of extant marine bivalve clades after the K/Pg extinction.

A discrete rebound phase with a steeper slope is also visible in semi-log space for the global data set (Fig. 1B), verified by a five-parameter (three rate parameters and two inflection points) model with a slope of 0.065 (four times the Mesozoic value and twice the overall Cenozoic value) and bounded at 65 and ~50 Ma. This short-term origination pulse is consistent with patterns described for post-K/Pg bivalves (*21*) and for other post-extinction intervals, and patterns of diversification in recovery episodes clearly stand as distinct from those in more quiescent times. However, this transient increase in origination does not bias the estimated rate differential between the broader Mesozoic and Cenozoic intervals. The rate estimated for the final 50 My of the Cenozoic (0.026) is very similar to that for the entire Cenozoic (0.030), indicating that the overall Cenozoic rate, λ_1 estimated with the three-parameter model above (table S2), is mathematically dominated by the shift in "normal" origination rates rather than the post-extinction rebound interval.

The evolutionary signature of the K/Pg event is also seen at the provincial scale (Figs. 1C and 2), with a strong latitudinal component despite the apparent global homogeneity of the extinction itself (*22*). The three-parameter function, with an inflection point near the K/Pg boundary, is the optimal fit for 19 of the 27 provinces; the provincial curves show some additional variation, but minor inflection points are not supported when the five-parameter model is used (perhaps due to small sample sizes in provinces with additional non-K/Pg inflections, such as the Lusitanian; Fig. 2). Provincial values of λ decrease with latitude (Fig. 1C), so that a one-parameter exponential function is best supported for two polar and six high-latitude provinces (table S2). This latitudinal effect is seen most clearly when a critical time of 65 My is imposed on all 27 provinces and the probability functions are fit to the data: The difference between the post- and pre-K/Pg rates ($\lambda_1 - \lambda_2$) declines significantly as the latitudinal midpoint of the province increases (Fig. 3A), a pattern driven predominantly by the

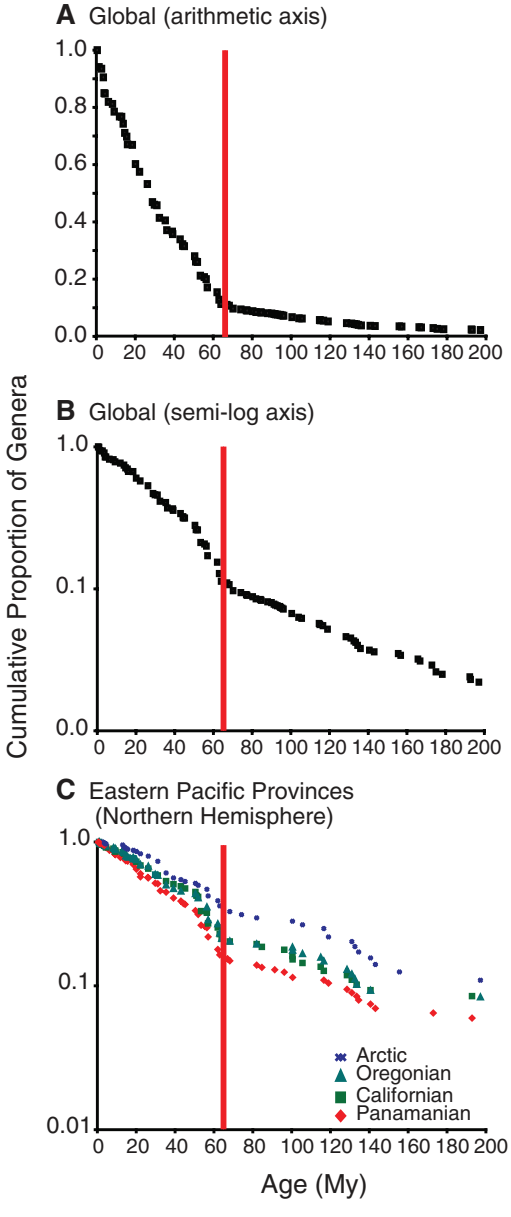


Fig. 1. Global and regional BSCs for living marine bivalves; the vertical red line marks the K/Pg boundary (65 Ma). (**A**) Global BSC plotted on arithmetic axes. (**B**) Global BSC in semi-log space showing constant rates and different slopes of the pre- and post-K/Pg portions of the curve. Straight lines in semi-log space represent constant exponential rates. (**C**) Decrease in BSC slopes with increasing latitude for the Northern Hemisphere bioprovinces of the eastern Pacific.

shallowing of the estimated Cenozoic rate (λ_1) in high-latitude provinces (Fig. 3B). Pre-K/Pg rates show no such latitudinal trend (Fig. 3C). The tropical West Pacific and Arctic provinces show the extremes of this contrast (fig. S3): The pre-K/Pg slopes of these two provinces differ only slightly (many of these older taxa are shared by the two regions), but the difference after the K/Pg extinction is dramatic.

The ages used here are the global first appearance of a genus (as opposed to the first occurrence of a genus in a given province), so that the slope of the provincial curves represents the outcome of origination plus invasion minus local extinction, rather than simple in situ origination. However, the late Cenozoic fossil record constrains these data by showing that marine bivalve genera tend to appear first in the tropics, with range expansion being overwhelmingly unidirectional from the tropics to the poles (*8*, *9*, *23*). The shallowing of Cenozoic age-frequency curves from tropics to poles thus appears to reflect the decreasing probability for genera to reach and remain established in progressively higher latitudes (*9*). Therefore, though the end-Cretaceous extinction was equally severe in low and high latitudes (*22*), the spatial pattern of origination and spread affects the taxonomic structures of these regions differently; the post-K/Pg slopes of BSCs primarily reflect origination rates in tropical provinces and invasion rates in polar provinces. This view is supported by the contrasting geographic distributions of pre- and post-K/Pg genera (Fig. 4). The distribution of post-K/Pg genera today forms a hollow curve, with 32% of the 638 genera being present exclusively in tropical or warm-temperate provinces (Fig. 4). Pre-K/Pg genera have significantly wider distributions, with only 14% of 73 genera being exclusive to warm-water provinces (significantly fewer than in the post-K/Pg sample, log-likelihood ratio G test, $P < 0.001$), the majority (82%) inhabiting both warm- and cold-water regions (Fig. 4).

Because the present day is our reference time, our BSCs are reminiscent of lineage-through-time (LTT) plots, which estimate diversity dynamics through time using the divergence times of lineages in a phylogenetic tree constructed for

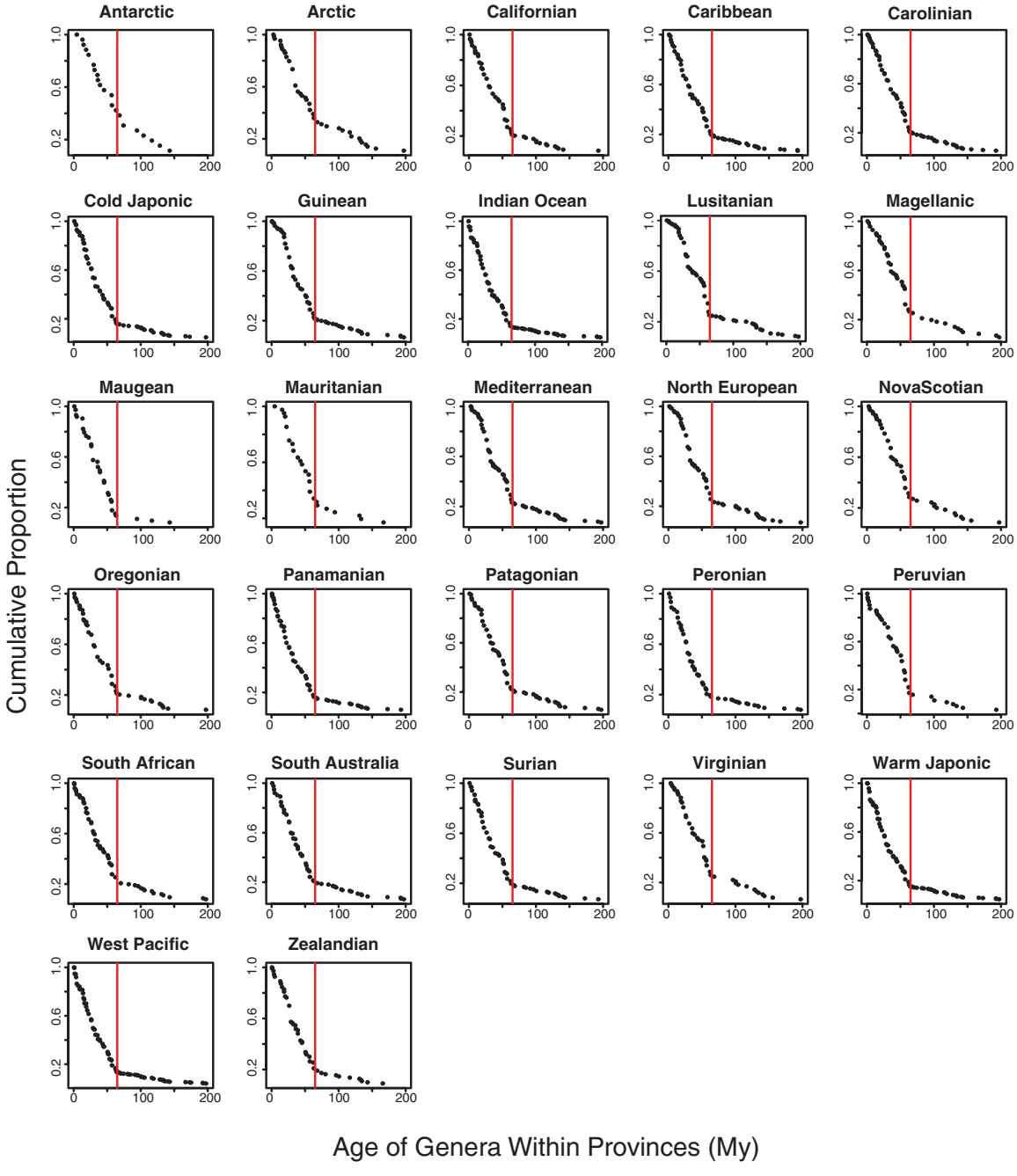


Fig. 2. Backward survivorship curves for the 27 provincial faunas of the present-day shelf biota. Plots are on arithmetic axes to enable visualization at low diversities; vertical red lines mark the K/Pg boundary (65 Ma).

living species (*24*–*27*). BSCs differ importantly from LTTs, however, in that the frequency distribution of branch lengths (here the age of the genus) is the raw data, rather than the length of time between nodes on a phylogenetic tree (*28*). Therefore, although changes in extinction rate can influence the slopes of LTTs (*24*, *28*), these effects are mathematically cancelled out in BSCs (*5*, *28*), given the assumptions noted above.

Nevertheless, we verified our global results by calculating the average per-taxon origination rates (*29*) [termed "per-capita rates" in (*29*), but units are actually per lineage per million years] of all genera, living and extinct, within extant bivalve families for 5-My intervals within the post-recovery Cenozoic (0 to 50 Ma) and for the 50 My before the K-Pg boundary. This form of per-capita rate, derived from branching theory, is robust to variation in the true extinction rate, including events as dramatic as mass extinctions (*29*). Focusing on shorter 5-My intervals also limits the biasing effect of taxonomic heterogeneity on the estimated rates, so this analysis, including extinct genera, makes a useful comparison to the results for living genera alone. These per-capita origination rates are remarkably close to those determined for the living genera, with the rate for the 50 My before the K/Pg at 0.020 genera per lineage per million years (versus 0.014 for the living cohort) and the post-K/Pg rate at 0.026 per lineage per million years (versus 0.026 for the living cohort), a statistically significant increase (Student's t test, $P = 0.02$, normality of data confirmed by Shapiro-Wilk test; a nonparametric Wilcoxon signed-rank test gives a similar result, $P = 0.025$) that corroborates the shift in origination rate at the K/Pg boundary. The per-taxon rate calculations do not specifically address the assumptions underlying the rates estimated from the BSC of the living cohort, but the agreement of the two results suggests that the inference of an inflection point in the BSC is robust. Spatially explicit LTT plots and paleontological cohort analyses will be valuable next steps but are difficult to generate because local extinction rates can include strong sampling effects.

The sharp increase in the slope of the bivalve BSC after the K/Pg event is unlikely to derive entirely from improved preservation and sampling in the younger part of the record and in the present day. Our optimizations find the changes in slope at the K/Pg boundary and in the early Eocene, not in the late Cenozoic when lithification and dissolution impinge least on fossil preservation (*30*). Additionally, an analysis restricted to bivalves having shell microstructures most robust to geochemical dissolution (with low organic content, consisting predominantly of calcite) and thus producing records least subject to distortion over time (*11*) shows a pattern consistent with that of the entire bivalve cohort (fig. S4).

These results provide a long-term perspective on the consequences of the K/Pg mass extinction, with a significant shift in origination rate that outlasted the accelerated recovery phase, and a strong spatial component to diversification, such that tropical and warm-temperate regions carry the strongest signature of post-K/Pg evolution. The lasting increase in origination may reflect the removal of competitors or predators that were damping the Mesozoic diversification of the living families; alternatively, the Cenozoic diversification of shell-penetrating predators, particularly in the tropics (*31*, *32*), might have increased origination rates in their prey during the post-recovery interval. These hypotheses need to be tested directly; for example, by ecological and functional analyses of victims and post-extinction diversifiers. The declining strength of the apparent shift in global origination rates with latitude underscores how the taxonomic composition of high-latitude provinces depends on tropical origination rates. The K/Pg extinction was large, but its downstream effect has also been important

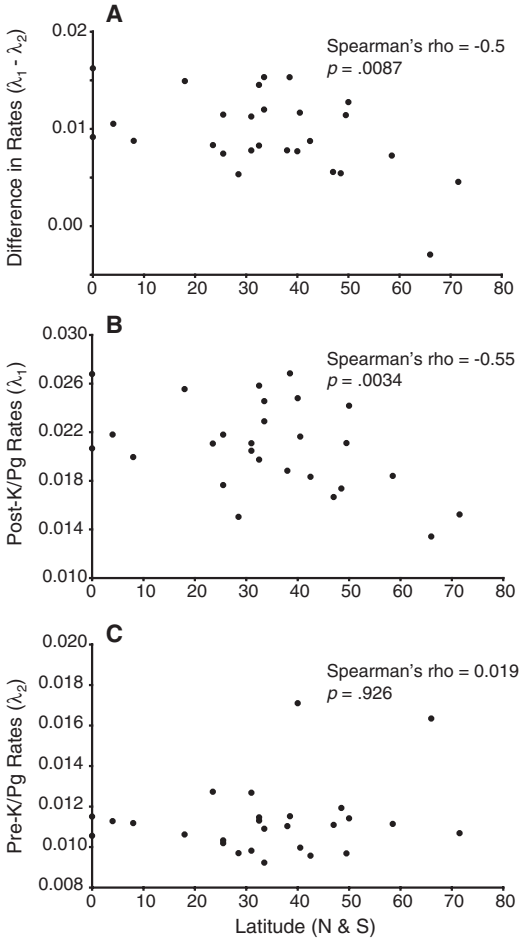


Fig. 3. Relationship between the estimated rates for the 27 biogeographic provinces and latitude. Rates were estimated using an inflection point of 65 My. (**A**) The difference between post- and pre-K/Pg origination rates ($\lambda_1 - \lambda_2$) declines significantly with increasing latitude. (**B**) Post-K/Pg origination rates (λ_1) decrease significantly with latitude. (**C**) Pre-K/Pg origination rates (λ_2) show no significant trend with latitude.

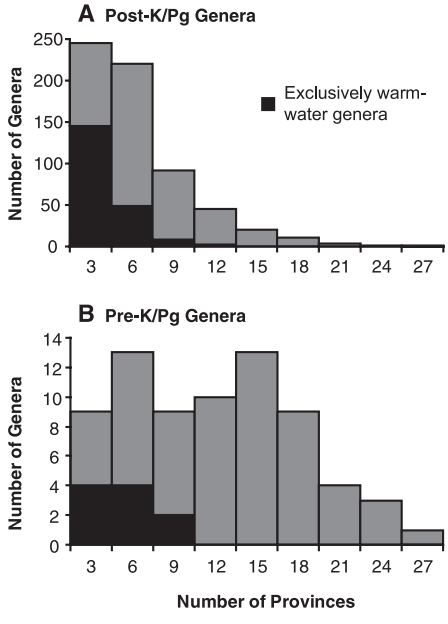


Fig. 4. Present-day geographic ranges of bivalve genera that originated (**A**) after or (**B**) before the K/Pg extinction. Genera originating before the K/Pg extinction (B) have significantly greater geographic ranges today, determined as the number of bioprovinces in which a genus occurs (median range of pre-K/Pg genera = 10 provinces, median range of post-K/Pg genera = 4 provinces, Kolmogorov-Smirnov test, $P = 2.16 \times 10^{-5}$), and significantly fewer exclusively warm-water genera than (A) genera originating after the K/Pg. Genera that exist exclusively in provinces with tropical or warm-temperate climates are denoted in black. All other genera have ranges extending into (but not necessarily restricted to) cold-temperate or polar provinces. The vast majority of such genera (91%) also occur in warm-water provinces.

and ongoing, so that the long-term shift in origination rate, and the spatial pattern of those originations, continues to structure the biological provinces of the modern world. Despite 65 My of evolutionary and environmental change, the modern marine biota, globally and regionally, bears the evolutionary and biogeographic imprint of the K/Pg event.

References and Notes

1. J. J. Sepkoski Jr., in *Global Events and Event Stratigraphy*, O. H. Walliser, Ed. (Springer, Berlin, 1996), pp. 35–52.
2. M. Foote, *J. Geol.* **111**, 125 (2003).
3. G. G. Mittelbach *et al.*, *Ecol. Lett.* **10**, 315 (2007).
4. R. E. Ricklefs, *Ecol. Lett.* **7**, 1 (2004).
5. M. Foote, in *Evolutionary Patterns*, J. B. C. Jackson, S. Lidgard, M. L. McKinney, Eds. (Univ. of Chicago Press, Chicago, 2001), pp. 245–294.
6. C. M. Pease, *Paleobiology* **13**, 484 (1987).
7. See supporting material on *Science* Online.
8. D. Jablonski, K. Roy, J. W. Valentine, *Science* **314**, 102 (2006).
9. J. W. Valentine, D. Jablonski, A. Z. Krug, K. Roy, *Paleobiology* **34**, 169 (2008).
10. D. Jablonski, K. Roy, J. W. Valentine, R. M. Price, P. S. Anderson, *Science* **300**, 1133 (2003).
11. S. M. Kidwell, *Science* **307**, 914 (2005).
12. J. W. Valentine, D. Jablonski, S. Kidwell, K. Roy, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6599 (2006).
13. J. A. Crame, *Paleobiology* **28**, 184 (2002).
14. A. K. Behrensmeyer *et al.*, *Paleobiology* **31**, 607 (2005).
15. J. A. Crame, *Paleobiology* **26**, 188 (2000).
16. M. L. Reaka, P. J. Rodgers, A. U. Kudla, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11474 (2008).
17. S. T. Williams, T. F. Duda, *Evolution* **62**, 1618 (2008).
18. D. M. Raup, *Paleobiology* **4**, 1 (1978).
19. D. M. Raup, *Paleobiology* **1**, 333 (1975).
20. J. J. Sepkoski Jr., *Bull. Am. Paleontol.* **363**, 1 (2002).
21. A. I. Miller, J. J. Sepkoski Jr., *Paleobiology* **14**, 364 (1988).
22. D. M. Raup, D. Jablonski, *Science* **260**, 971 (1993).
23. K. Roy, E. E. Goldberg, *Am. Nat.* **170**, S71 (2007).
24. R. E. Ricklefs, *Trends Ecol. Evol.* **22**, 601 (2007).
25. R. E. Ricklefs, J. B. Losos, T. M. Townsend, *J. Evol. Biol.* **20**, 1751 (2007).
26. S. Nee, *Annu. Rev. Ecol. Evol. Syst.* **37**, 1 (2006).
27. S. Nee, *Evolution* **55**, 661 (2001).
28. S. Nee, *Paleobiology* **30**, 172 (2004).
29. M. Foote, *Paleobiology* **26** (suppl.), 74 (2000).
30. J. Alroy *et al.*, *Science* **321**, 97 (2008).
31. S. M. Stanley, *Paleobiology* **34**, 1 (2008).
32. G. J. Vermeij, *Evolution and Escalation* (Princeton Univ. Press, Princeton, NJ, 1987).

33. We thank M. Foote, S. M. Kidwell, T. D. Price, K. Roy, and three anonymous reviewers for comments and M. Foote for programming assistance. This research was supported by NASA.

Supporting Online Material

www.sciencemag.org/cgi/content/full/523/5915/767/DC1
Methods
Figs. S1 to S4
Tables S1 and S2
References

20 August 2008; accepted 3 December 2008
10.1126/science.1164905

A Great-Appendage Arthropod with a Radial Mouth from the Lower Devonian Hunsrück Slate, Germany

Gabriele Kühl,[1] Derek E. G. Briggs,[2] Jes Rust[1]

Great-appendage arthropods, characterized by a highly modified anterior limb, were previously unknown after the Middle Cambrian. One fossil from the Lower Devonian Hunsrück Slate, Germany, extends the stratigraphic range of these arthropods by ~100 million years. *Schinderhannes bartelsi* shows an unusual combination of anomalocaridid and euarthropod characters, including a highly specialized swimming appendage. A cladistic analysis indicates that the new taxon is basal to crown-group euarthropods and that the great-appendage arthropods are paraphyletic. This new fossil shows that features of the anomalocaridids, including the multisegmented raptorial appendage and circular plated mouth, persisted long after the initial radiation of the euarthropods.

A range of Cambrian arthropods share a prominent limb at the front of the head referred to as the "great appendage" (*1*). They include the Anomalocarididae and Cambrian euarthropods such as *Yohoia*, *Leanchoilia*, *Jiangfengia*, and *Fortiforceps* and are usually interpreted as a paraphyletic group within the crown, basal to chelicerates, and the great appendage to be homologous with the chelicera of living chelicerates (*1*–*6*). Alternatively, they have been considered basal to the crown-group euarthropods, in which case the great appendage has been lost in living taxa (*7*, *8*). Here, we describe a great-appendage arthropod from the famous Lower Devonian Hunsrück Slate (*9*). This arthropod shows a previously unseen combination of the characters that occur in Cambrian anomalocaridids and euarthropods.

Schinderhannes bartelsi, a new genus and species, is a representative of the stem lineage of the Euarthropoda. The name *Schinderhannes* is derived from an 18th-century bandit in the Hunsrück area and *bartelsi* is a tribute to Christoph Bartels of Bochum, Germany, an authority on the Hunsrück Slate fossils. The material is a single specimen: PWL 1994/52-LS is the holotype, located at the Naturhistorisches Museum Mainz. The specimen has been x-radiographed (Fig. 1F), drawn (Fig. 1, B and D), and reconstructed (Fig. 1G). It was discovered in the Hunsrück Slate, Germany, at the Eschenbach-Bocksberg Quarry in Bundenbach (*10*); it was found in the Lower Devonian Kaub Formation.

Diagnosis. The head bears a preoral great appendage with at least nine podomeres, a radial mouth, and huge eyes. Podomeres three to nine of the great appendage bear an elongate lateral spine, and podomeres four to nine have a long, ventral, comblike projection. An additional large triangular flaplike appendage at the rear of the head is thickened along its margin. The trunk consists of 12 segments: The first 10 have tergites and bear biramous appendages, the 11th bears a lateral flukelike appendage, and the 12th is apodous. The trunk terminates in a long tail spine.

Description. The total length, including the tail spine, is 98 mm. The specimen (Fig. 1, A and B) is oriented parallel-oblique to the bedding so that the ventral side shows the left eye in outline and reveals the outer surface of the left edge of the trunk tergites (fig. S1, C and D). Some features, such as the head shield, are obscured by incomplete pyritization and phosphatization. The outline of the large postoral head appendage has been affected by flattening: The left is foreshortened as compared with the right, which shows striking evidence of wrinkling. This appendage is difficult to interpret without additional examples preserved in other attitudes to the bedding.

The body is divided into head and trunk. The head bears a pair of preoral great appendages (Fig. 1, B to E, and fig. S1A), a circular mouth structure, a pair of huge eyes, and a pair of flaplike limbs at the rear. Although there is no clear evidence of a head shield, the presence of dorsal tergites in the trunk region suggests that one was present (fig. S1D).

The great appendage (A1) (Fig. 1, B and E, and fig. S1A) consists of up to nine podomeres. Podomeres one and two are short and wide, whereas podomere three is three times longer and bears a lateral spine (Fig. 1, E and G). Podomeres four to nine also bear a long inwardly projecting lateral spine and a long parallel-sided ventral projection that extends at a high angle to the appendage (Fig. 1, B, E, and G, and fig. S1A). This projection decreases in length from the fourth to the ninth podomere. Each one bears a series of up to five or six regularly spaced orthogonal spines along its length, decreasing in number on the more distal podomeres. A subcircular mouth ring (Fig. 1, F and H, and fig. S1B) is concealed by the ventral projections. A series of small subtriangular features, particularly on the left side, may represent tooth plates. Differences in width may reflect its attitude to the bedding.

The stalked eyes are positioned laterally behind the great appendages. The left eye indicates a suboval outline, although it is affected by flattening. The eyes are covered in numerous tiny, close-packed, hexagonal lenses (fig. S1C).

[1]Division of Palaeontology, Steinmann Institute, University of Bonn, Nussallee 8, 53115 Bonn, Germany. [2]Department of Geology and Geophysics, and Yale Peabody Museum of Natural History, Yale University, Post Office Box 208109, New Haven, CT 06520–8109, USA.

A subcircular darker area between the eyes consists of apatite, presumably indicating the position of the gut.

A second highly differentiated appendage (A2) is interpreted as belonging to the head (Fig. 1, A and B). It is a wide, triangular, flaplike structure with a broad proximal area (Fig. 1F). The ridgelike appearance of the margins represents thickening, and they preserve evidence of short spines (Fig. 1B).

The trunk consists of 12 segments (Fig. 1B). The first 10 bear a pair of homogeneous biramous appendages. The endopods are flat and made up of at least three wide podomeres (fig. S1F). The proximal two podomeres are short and similar in length; the third is twice as long and terminates in several elongations. The exopod consists of an unsegmented axis bearing numerous flaps (fig. S1, D and E). The appendages correspond in position to the tergites. The last two segments of the trunk lack biramous appendages. Segment 11 bears a pair of lateral flukelike limbs with thickened outer and posterior margins; they partially overlap segment 12 (Fig. 1, A and B).

A concentration of apatite along the axis of the trunk may reflect the position of the gut. The anus is located at the posterior margin of the last trunk segment, indicated by a circle of pyrite. The trunk terminates in a long unsegmented spine with a median keel.

The spiny great appendage suggests that *Schinderhannes* was a predator [for example, (*2*)]. The large eyes imply a substantial visual capability. The flukes at the rear of the trunk indicate that swimming was the main mode of locomotion. Thrust was presumably generated by the large postoral appendage (A2), which may have functioned as a paddle or "wing," as was deduced for the lobes in the trunk of

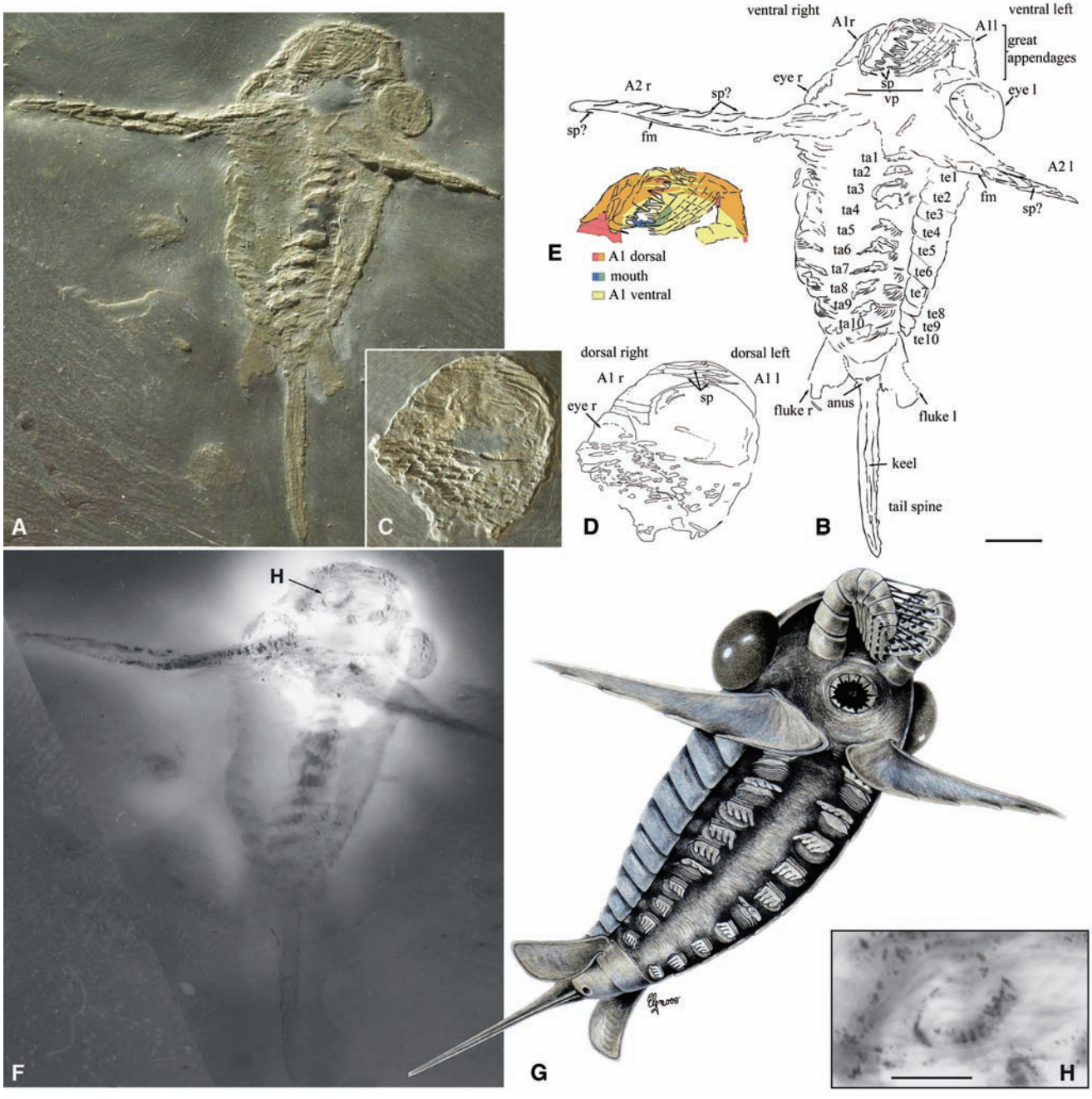


Fig. 1. Holotype of *Schinderhannes bartelsi*. (**A**) Ventral. (**B**) Interpretative drawing of ventral side. l, left; r, right; A1, great appendage; A2, flaplike appendage; sp, spine; fm, flap margin; te, tergite; ta, trunk appendage. (**C**) Partly exposed dorsal side, horizontally mirrored. (**D**) Interpretative drawing of dorsal side. (**E**) Interpretative drawing of great appendages, combining information from the dorsal and ventral sides. (**F**) Radiograph. (**G**) Reconstruction. Scale bar, 10 mm [for (A) to (G)]. (**H**) Mouth-part. Scale bar, 5 mm.

Anomalocaris (*11*). Lift was also generated by the tail flukes.

Schinderhannes shares several characters with anomalocaridids. The morphology and position of the great appendage, which lies at the front of the head and anterior of the eyes, are very similar to that of "Appendage F" (*12*), the frontal appendage of *Laggania cambria* (*11*, *13*). The circular mouth is characteristic of anomalocaridids (*13*), and flap-like appendages similar to that at the rear of the head of *S. bartelsi* occur in the trunk of *Anomalocaris canadensis* (*11*), *Parapeytoia yunnanensis* (*14*), *Pambdelurion whittingtoni* (*15*), and *Kerygmachela kierkegaardi* (*16*). Other trunk features of *Schinderhannes* are characteristic of euarthropods, including "short–great-appendage" arthropods. These include the tergites, biramous trunk appendages, and tail spine.

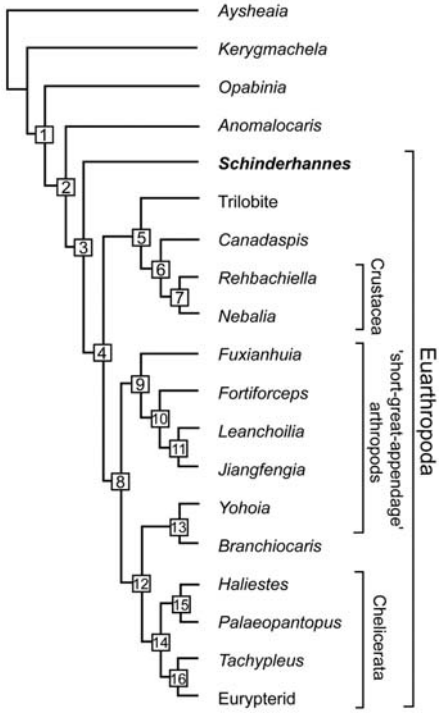


Fig. 2. Cladogram; tree length, 87. Consistency index, 0.5402; retention index, 0.6552. (1) Peytoia-like mouth sclerites, terminal mouth position, lateral lobes, loss of lobopod limbs, and stalked eyes. (2) Great appendages. (3) Sclerotized tergites, head shield, loss of lateral lobes, and biramous trunk appendages. (4) Stalked eyes in front and loss of radial mouth. (5) Post-antennal head appendages biramous and antenna in first head position. (6) Free cephalic carapace, carapace bivalved, and two pairs of antennae. (7) Maxilla I and II. (8) Exopods simple oval flap. (9) Two pre-oral appendages and a multisegmented trunk endopod. (10) Post-antennal head appendages biramous and tail appendages fringed with setae. (11) Long flagellae on great appendage and exopods fringed with filaments. (12) Trunk appendages uniramous and eyes not stalked. (13) No posterior tergites. (14) Tail spines and chelicere/chelifore on first head position. (15) Proboscis. (16) Six post-antennal head appendages.

A cladistic analysis (*17*) places *Schinderhannes* between *Anomalocaris* and other arthropods (Fig. 2). The short–great-appendage taxa are paraphyletic, and the monophyly of a taxon "Megacheira" (*1*) was not confirmed. The position of the short–great-appendage arthropods as stem lineage representatives of the Chelicerata, however, is consistent with the majority of recent analyses (*2*, *3*, *5*, *18*–*22*) and supports the interpretation of the great appendage as homologous with the chelicera of living chelicerates (*1*, *3*–*7*).

The discovery of *Schinderhannes* emphasizes the importance of exceptionally preserved deposits (Konservat-Lagerstätten) in revealing the evolutionary history of arthropods. It shows that features of the giant Cambrian anomalocaridids survived for about 100 million years after the Middle Cambrian. The Hunsrück Slate also yields examples of Marrellomorpha (*23*), a clade well known from the Cambrian (*24*) and more recently discovered in exceptionally preserved fossil deposits from the Silurian and the Ordovician (*25*, *26*). Thus, the rarity of post-Cambrian great-appendage arthropods may be a result in part of the decline of Burgess Shale–type preservation after the Middle Cambrian (*27*).

References and Notes

1. X.-G. Hou, J. Bergström, *Fossils Strata* **45**, 1 (1997).
2. J. Chen, D. Waloszek, A. Maas, *Lethaia* **37**, 3 (2004).
3. A. Maas *et al.*, *Prog. Nat. Sci.* **14**, 158 (2004).
4. G. Scholtz, G. D. Edgecombe, in *Crustacea and Arthropod Relationships,* S. Koenemann, R. Jenner, Eds. (Taylor and Francis, New York, 2005), pp. 139–165.
5. G. Scholtz, G. D. Edgecombe, *Dev. Genes Evol.* **216**, 395 (2006).
6. M. Jager *et al.*, *Nature* **441**, 506 (2006).
7. G. Budd, *Nature* **417**, 271 (2002).
8. A. Maxmen, W. E. Browne, M. Q. Martindale, G. Giribet, *Nature* **437**, 1144 (2005).
9. C. Bartels, D. E. G. Briggs, G. Brassel, *Fossils of the Hunsrück Slate: Marine Life in the Devonian* (Cambridge Univ. Press, Cambridge, 1998).
10. O. E. Sutcliffe, S. L. Tibbs, D. E. G. Briggs, *Metalla (Bochum)* **9**, 89 (2002).
11. H. B. Whittington, D. E. G. Briggs, *Philos. Trans. R. Soc. London Ser. B* **309**, 569 (1985).
12. D. E. G. Briggs, *Palaeontology* **22**, 77 (1979).
13. D. Collins, *J. Paleontol.* **70**, 280 (1996).
14. X.-G. Hou, J. Bergström, P. Ahlberg, *Geol. Foren. Stockh. Forh.* **117**, 163 (1995).
15. G. E. Budd, in *Arthropod Relationships*, R. A. Fortey, R. H. Thomas, Eds. (Chapman and Hall, London, 1998), vol. 55, pp. 125–138.
16. G. E. Budd, *Trans. R. Soc. Edinb. Earth Sci.* **89**, 249 (1999).
17. Materials and methods are available as supporting material on *Science* Online.
18. J. Bitsch, C. Bitsch, *Acta Zool.* **88**, 317 (2007).
19. T. J. Cotton, S. J. Braddy, *Trans. R. Soc. Edinb. Earth Sci.* **94**, 169 (2004).
20. D. Waloszek, J. Chen, A. Maas, X.-Q. Wang, *Arthropod Struct. Dev.* **34**, 189 (2005).
21. D. Waloszek, A. Maas, J. Chen, M. Stein, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **254**, 273 (2007).
22. J. A. Dunlop, *Acta Zool. Bulg.* **1** (suppl.), 9 (2006).
23. G. Kühl, J. Bergström, J. Rust, *Palaeontographica* **286**, 123 (2008).
24. H. B. Whittington, *Geol. Surv. Can. Bull.* **209**, 1 (1971).
25. P. Van Roy, thesis, University of Ghent, Ghent, Belgium (2006).
26. D. J. Siveter, R. A. Fortey, M. D. Sutton, D. E. G. Briggs, D. J. Siveter, *Proc. R. Soc. London Ser. B* **274**, 2223 (2007).
27. N. J. Butterfield, *Lethaia* **28**, 1 (1995).

28. We thank the German Science Foundation (grant PN RU 665/5-1) and the Humboldt Foundation (Research Award to D.E.G.B.) for support; the Museum of Natural History in Mainz for access to the material; C. Bartels for information; G. Oleschinski for radiographs and photography; E. Gröning for the reconstruction; and A. Maas, R. Willmann, G. D. Edgecombe, and two other reviewers for comments.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5915/771/DC1
Materials and Methods
SOM Text
Fig. S1
Table S1
References

30 September 2008; accepted 15 December 2008
10.1126/science.1166586

Two Thresholds, Three Male Forms Result in Facultative Male Trimorphism in Beetles

J. Mark Rowland[1]* and Douglas J. Emlen[2]

Male animals of many species deploy conditional reproductive strategies that contain distinct alternative phenotypes. Such facultatively expressed male tactics are assumed to be due to a single developmental threshold mechanism switching between the expression of two alternative phenotypes. However, we discovered a clade of dung beetles that commonly expresses two threshold mechanisms, resulting in three alternative phenotypes (male trimorphism). Once recognized, we found trimorphism in other beetle families that involves different types of male weapons. Evidence that insects assumed to be dimorphic can express three facultative male forms suggests that we need to adjust how we think about animal mating systems and the evolution of conditional strategies.

Alternative reproductive tactics in male animals, such as guards versus sneaks or callers versus satellites, have served as models for examining the evolution of threshold-mediated traits (e.g., polyphenisms). Thousands of empirical examples of alternative reproductive tactics (*1*), numerous theoretical models of threshold evolution (*2*–*5*), and many examples of poly-

phenic regulatory developmental mechanisms (*1*, *6*) are known. These studies assume that two phenotypic alternatives are separated by a single regulatory threshold, which results in a dominant tactic used by the largest, healthiest, or highest quality males and an alternative, generally less-aggressive tactic adopted by subordinate males. Models for the evolution of these and other threshold traits incorporate a single threshold that yields two reproductive tactics.

A few well-studied animal species contain three male tactics (male trimorphism), including alpha, beta, and gamma males in isopods (*7*) and fish (*8*); blue, yellow, and orange males in side-blotched lizards (*9*, *10*); and independent, satellite, and faeder males in birds (*11*). In all of these species, male reproductive phenotypes are determined primarily by the inheritance of alleles of large effect (i.e., the morphs behave like genetic polymorphisms), rather than by facultative mechanisms incorporating thresholds. The evolution of male trimorphism has yet to be examined in the context of conditional alternative tactics and polyphenism.

Because animal development is rife with thresholds and because the expression of complex morphological and behavioral phenotypes may be sensitive to many environmental or circumstantial stimuli at multiple periods during development (*6*), there is no reason to expect that a single threshold should be the norm. We provide examples of beetles that incorporate two developmental thresholds into the regulation of expression of a single male trait (weapons), yielding three male forms (facultative male trimorphism).

Dung beetles form a monophyletic clade of the Scarabaeidae that originated at least 65 million years ago but radiated extensively with mammals (*12*). More than 5000 species have been described worldwide, and many of these bear horns. Beetle horns are rigid outgrowths of the exoskeleton that function in intrasexual combat over access to resources. In most dung beetles, horns are confined to males, and males of many of these species are dimorphic, with distinct major and minor phenotypes (*13*, *14*). Typically, males larger than a threshold body size (major males) produce large horns. Males smaller than this threshold body size (minor males) have disproportionally smaller horns or are hornless. Male weapon morphology is determined facultatively, on the basis of an interaction between an intrinsic and heritable threshold and the nutritional history (environment) of the animal, resulting in maturation at a body size that is either above or below this critical threshold size (*14*–*16*). Major and minor male dung beetles have been found to use alternative reproductive tactics to mate with females (*17*–*19*). Major males use their large body sizes and long horns to defend entrances to burrows containing females, guarding these tunnels from rival males; minor males sneak into these tunnels, either by slipping past the guarding male or by digging side tunnels that intercept tunnels beneath the guarding males. Thus, dung beetles appear to fit the traditional model of conditionally expressed alternative male tactics regulated by a single developmental threshold mechanism.

However, at least two qualitatively different conditional threshold mechanisms are now known to regulate horn expression in dung beetles. One mechanism (threshold mechanism 1; Fig. 1A, left) appears to truncate production of the horn entirely, either by suppressing proliferation in the cells that form the horns (*20*, *21*) or by reabsorbing horn tissue during metamorphosis (*22*). Populations incorporating this regulatory

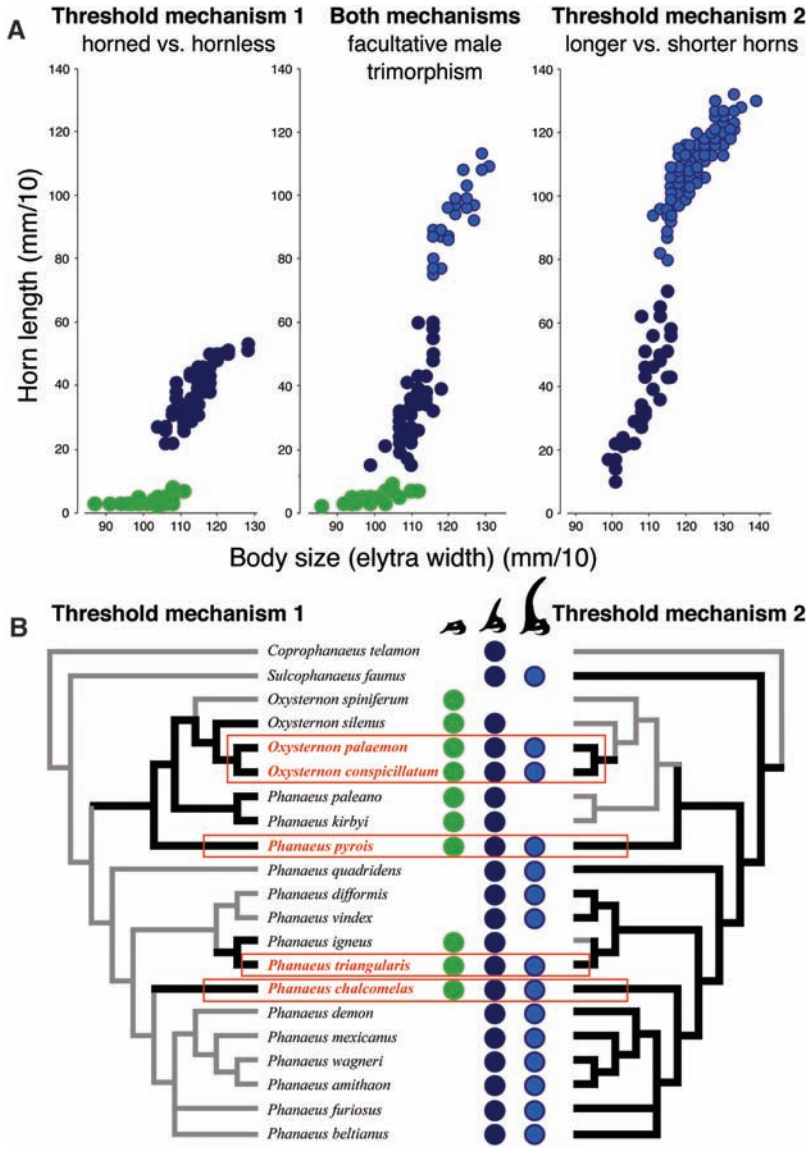


Fig. 1. Two threshold mechanisms regulate expression of horns in male dung beetles. (**A**) Horn length/body size scaling relationships for natural populations of adult males of *Phanaeus igneus* (left), *P. triangularis* (middle), and *P. vindex* (right) reveal these processes as an abrupt and size-dependent change in scaling relationship slope (threshold mechanism 1, left), a size-dependent shift in intercept (threshold mechanism 2, right), or a combination of the two mechanisms (middle). Because among-individual variation in body size in scarab beetles is influenced primarily by larval nutrition, species simultaneously incorporating both of these threshold mechanisms are facultatively trimorphic. From this, we define alpha (light blue), beta (dark blue), and gamma (green) male forms, which were discriminated by the likelihood method for normal distributions (*25*). (**B**) Evolution of two thresholds mapped onto a phylogeny of 22 species of phanaeine scarabs [tree topology from (*34*)], with the presence of each threshold mechanism (bold branches) estimated by parsimony (*35*). Species containing alpha, beta, and gamma male forms indicated as above. Facultative trimorphism appears to have been gained at least four separate times in the period covered by this phylogeny (red boxes).

[1]Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA. [2]Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA.

*To whom correspondence should be addressed. E-mail: rowland@unm.edu

mechanism can be identified by a dramatic switch in the slope of the scaling relationship (allometry) between horn length and body size. Individuals producing horns have horn lengths that scale positively and often linearly with among-individual variation in body size. Individuals not producing horns yield flat scaling relationships when the corresponding parts of the head are measured. A different mechanism (threshold mechanism 2; Fig. 1A, right) modulates the total amount of horn growth, resulting in morphologically similar but disproportionally shorter horns. Populations incorporating this second threshold mechanism typically have horn length/body size scaling relationships with similar (e.g., parallel) slopes but shifted *y* intercepts. Both types of threshold mechanism have been shown to couple horn growth with nutrition in scarab beetles and are considered facultative (polyphenic) regulatory processes (*14*, *16*, *23*, *24*), and preliminary evidence suggested that they might sometimes both be present in the same species (*25*).

We looked at a taxonomic array of phanaeine dung beetles to determine whether both threshold mechanisms could cooccur within the males of a single species. Remarkably, we found that simultaneous incorporation of both threshold mechanisms was common in these beetles and that male trimorphism resulted from this cooccurrence. We were thus able to identify distinct alpha, beta, and gamma males within these species (Figs. 1 and 2). Additionally, we found that both temporal and geographic changes in population body sizes in trimorphic species produced changes in the relative proportions of the three male morphs, as predicted for facultative threshold mechanisms (fig. S1, A and B). Our data suggest that facultative trimorphism evolved through the stepwise accumulation of independent threshold mechanisms (e.g., monomorphism → threshold mechanism 2 → threshold mechanisms 2 + 1; Fig. 1B). An alternative possibility that we did not test is that independent, dimorphic lineages with different threshold mechanisms hybridized to form trimorphic populations. Data from additional species will be needed to test this evolutionary scenario.

We observed male trimorphism in 5 of the 21 species of phanaeine dung beetles studied, with striking trimorphism in several (Fig. 1 and 2). These results coupled with literature suggesting that isolated species of other insects may have three male forms (*26*–*29*) led us then to look for trimorphism in other families of beetles. Here again, a limited study identified male trimorphism in the allied family Lucanidae (stag beetles) and also in the more distantly related Curculionidae (weevils). We found that trimorphism is expressed in three different organ systems in these three families of beetles: head horns in Scarabaeidae, mandibles in Lucanidae, and sternal spines in Curculionidae (Fig. 2). It now appears that several groups long considered to be male dimorphic actually contain trimorphic species.

We detected the occurrence of trimorphism in dung beetles by using phylogenetic reconstruction of the evolution of each of two developmental threshold mechanisms, scoring trimorphic taxa as having both mechanisms. Trimorphism in several of these species was evident in the nonlinear allometry of their head horns. However, none of the statistical methods used to detect thresholds (*13*, *25*, *30*, *31*) was able to consistently identify more than a single threshold in these taxa, reflecting the fact that they were designed to test for dimorphism, not trimorphism. In addition, we showed that trimorphism can occur in other groups of beetles without dramatic or even detectable inflections in allometry. For example, in lucanids allometry of mandible scaling alone reveals two male trajectories, yet we found that at least some lucanid species express three qualitatively different male phenotypes on the basis of armature of the mandibles (Fig. 2B and fig. S2). This suggests that in some taxa gamma male morphs will be difficult to detect with current methods and raises the possibility that gamma males are present but as yet undetected in many animal taxa.

How are three male forms maintained in these populations? Studies of allelically trimorphic taxa may be especially informative because al-

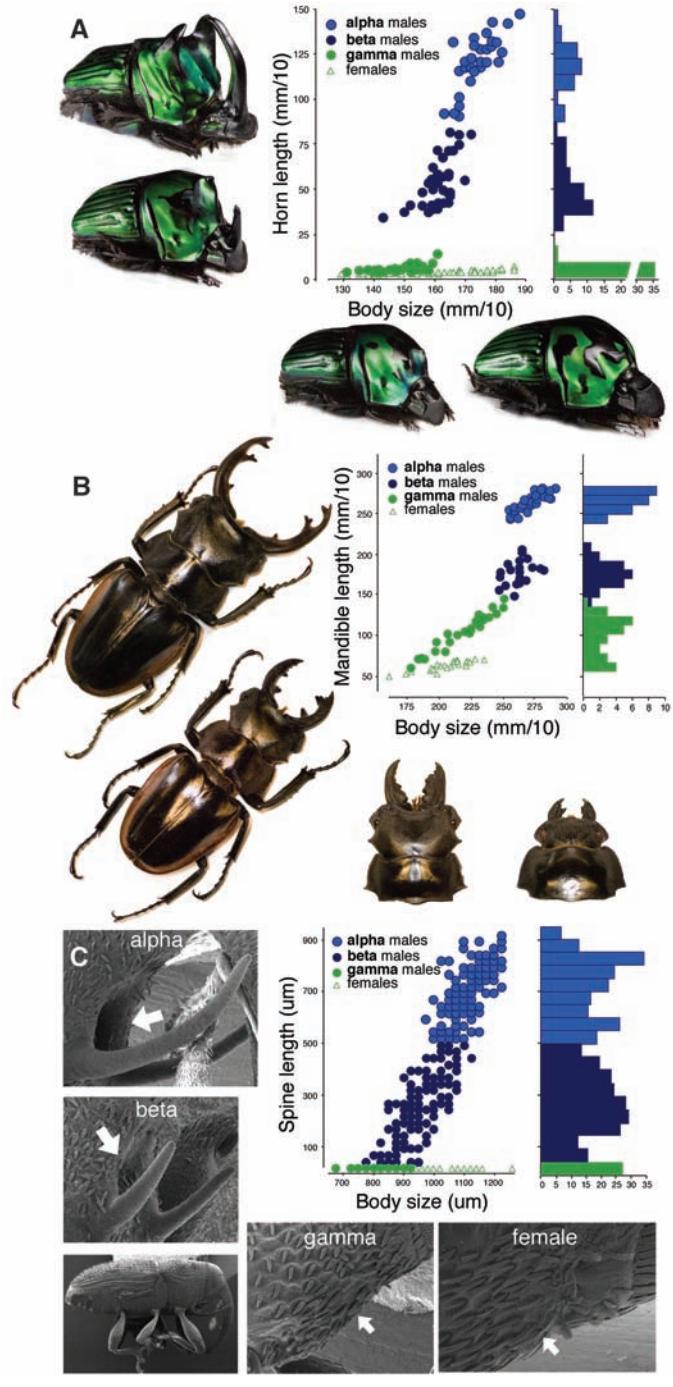


Fig. 2. Male trimorphism in three beetle families. (**A**) In scarabaeid dung beetles (*Oxysternon conspicillatum*), alpha and beta males both produce head horns, but the relative sizes of these weapons differ (shift in intercept of the scaling relationship between horn length and body size, threshold mechanism 2). Gamma males and females lack head horns entirely (change in scaling relationship slope, threshold mechanism 1). Alpha, beta, and gamma males were discriminated by the likelihood method for normal distributions (*25*). (**B**) In lucanid beetles (*Odontolabis cuvera*), male mandibles occur in three discrete anatomical conformations, which identify the alpha, beta, and gamma male forms (also fig. S2). (**C**) In weevils [*Parisoschoenus expositus*; measures from (*26*)], males produce long ventral spines that are outgrowths of the sternum and flank a deep invagination of cuticle, the sheath. As with the dung beetles, alpha and beta males produce similar weapons that differ in relative size (threshold mechanism 2), whereas gamma males lack weapons and resemble females. In *P. expositus*, gamma males and females lack both the ventral spines and the sheath (white arrows) used in male combat (*26*).

pha, beta, and gamma males are well known to use discrete alternative mating behaviors (*7*–*10*). In these species, negative frequency-dependent selection with intransitive fitness interactions among tactics (e.g., rock-paper-scissors) appear to maintain trimorphism within populations (*9*, *10*, *32*). It is unknown how male reproductive behavior varies in species with facultative trimorphism, and it will be important to determine whether similar fitness intransitivity of tactics applies in these cases. It is interesting that the smallest male forms (gamma males), at least in the beetle families studied so far, invariably resemble females. This suggests that male reproductive tactics may include a dominant (fight/guard) tactic, a subordinate (sneak) tactic, and a female-mimicry tactic. This is a striking parallel to allelically trimorphic fish (*8*), isopods (*7*), and birds (*11*) and would support predictions from recent rock-paper-scissors models that suggest that many taxa will contain cryptic (undetected) female-mimicking males (*10*, *32*).

Recognizing that there are at least two distinct thresholds also affects studies of the evolution of these developmental mechanisms. It is already clear that there are many developmental routes to the polyphenic regulation of male weapon systems in beetles, and, as these mechanisms are discovered and described, they are routinely compared across species (*20*–*22*, *33*). We suggest that the most meaningful comparisons will be those that explicitly consider the type of threshold mechanism involved and treat these accordingly as distinct and evolutionarily independent processes. Our findings raise the possibility that horned male majors in species with threshold mechanism 1 (Fig. 1A, left) are actually, developmentally, beta males, whereas the horned majors in species with threshold mechanism 2 (Fig. 1A, right) are alpha males. Acknowledging this distinction can help us better understand the full complexity of their rich behavioral repertoires, as well as more appropriately study the developmental and genetic architectures of their facultative regulatory mechanisms.

References and Notes

1. R. Oliveira, M. Taborsky, H. J. Brockmann, *Alternative Reproductive Tactics: An Integrative Approach* (Cambridge Univ. Press, Cambridge, UK, 2008), p. 507.
2. W. N. Hazel, R. Smock, M. D. Johnson, *Proc. R. Soc. London Ser. B.* **242**, 181 (1990).
3. S. J. Plaistow, R. A. Johnstone, N. Colegrave, M. Spencer, *Behav. Ecol.* **15**, 534 (2004).
4. J. Repka, M. R. Gross, *J. Theor. Biol.* **176**, 27 (1995).
5. S. M. Shuster, M. J. Wade, *Mating Systems and Strategies* (Princeton Univ. Press, Princeton, NJ, 2003).
6. M. J. West-Eberhard, *Developmental Plasticity and Evolution* (Oxford Univ. Press, Oxford, 2003).
7. S. M. Shuster, *Evolution* **43**, 1683 (1989).
8. M. J. Ryan, C. M. Pease, M. R. Morris, *Am. Nat.* **139**, 21 (1992).
9. B. Sinervo, C. M. Lively, *Nature* **380**, 240 (1996).
10. B. Sinervo, K. R. Zamudio, *J. Hered.* **92**, 198 (2001).
11. J. Jukema, T. Piersma, *Biol. Lett.* **2**, 161 (2006).
12. A. L. V. Davis, C. H. Scholtz, T. K. Philips, *J. Biogeogr.* **29**, 1217 (2002).
13. W. G. Eberhard, E. E. Gutierrez, *Evolution* **45**, 18 (1991).
14. D. J. Emlen, *Proc. R. Soc. London Ser. B* **256**, 131 (1994).
15. A. P. Moczek, H. F. Nijhout, *Evol. Dev.* **4**, 252 (2002).
16. D. J. Emlen, H. F. Nijhout, *J. Insect Physiol.* **45**, 45 (1999).
17. A. P. Moczek, D. J. Emlen, *Anim. Behav.* **59**, 459 (2000).
18. J. Rasmussen, *J. Insect Behav.* **7**, 67 (1994).
19. J. Hunt, L. W. Simmons, *Proc. R. Soc. London Ser. B* **268**, 2409 (2001).
20. D. J. Emlen, Q. Szafran, L. S. Corley, I. Dworkin, *Heredity* **97**, 179 (2006).
21. D. J. Emlen, J. Hunt, L. W. Simmons, *Am. Nat.* **166**, S42 (2005).
22. A. P. Moczek, *Am. Nat.* **168**, 711 (2006).
23. Y. Iguchi, *Ann. Entomol. Soc. Am.* **91**, 845 (1998).
24. K. Karino, N. Seki, M. Chiba, *Ecol. Res.* **19**, 663 (2004).
25. J. M. Rowland, C. R. Qualls, *Evol. Ecol. Res.* **7**, 421 (2005).
26. W. G. Eberhard, J. M. Garcia-C, J. Lobo, *Proc. R. Soc. London Ser. B* **267**, 1129 (2000).
27. J. C. Moore, J. Pienaar, J. M. Greeff, *Behav. Ecol.* **15**, 735 (2004).
28. C. Kelly, *Behav. Ecol.* **19**, 1018 (2008).
29. Y. Iguchi, *Kogane* **1**, 21 (2000).
30. J. S. Kotiaho, J. L. Tomkins, *Behav. Ecol.* **12**, 553 (2001).
31. J. L. Tomkins, J. S. Kotiaho, N. R. LeBas, *Am. Nat.* **165**, 389 (2005).
32. B. Sinervo *et al.*, *Am. Nat.* **170**, 663 (2007).
33. A. P. Moczek, D. Rose, W. Sewell, B. R. Kesselring, *Dev. Genes Evol.* **216**, 655 (2006).
34. D. L. Price, *Insect Syst. Evol.* **38**, 1 (2007).
35. D. R. Maddison, W. P. Maddison, *MacClade 4.0: Analysis of Phylogeny and Character Evolution* (Sinauer, Sunderland MA, 2000).

36. We thank S. Spector, T. Larsen, A. Solís, T. Gardner, J. Louzada, P. Skelley, D. Almquist, C. Gillett, D. Edmonds, D. Lewis, L. Herman, T. Fincher, M. Barclay, C. O'Brien, J. Prena, F. Génier, G. and J. Lewallern, R. Veal, D. Heinicke, and B. Raber for providing critical samples; assistance of the Scarabaeine Research Network; K. Bright, W. Eberhard, E. Greene, A. Kodric-Brown, T. Maginnis, E. McCullough, C. Qualls, and two anonymous reviewers for ideas and comments on the manuscript; J. Driver and the University of Montana Electron Microscopy facility for the weevil images; and the NSF (IOS-0642409) (to D.J.E.) for funding.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5915/773/DC1
Materials and Methods
Figs. S1 and S2

17 October 2008; accepted 7 January 2009
10.1126/science.1167345

Sequential Sympatric Speciation Across Trophic Levels

Andrew A. Forbes,[1]* Thomas H.Q. Powell,[1] Lukasz L. Stelinski,[2] James J. Smith,[3] Jeffrey L. Feder[1]†

A major cause for biodiversity may be biodiversity itself. As new species form, they may create new niches for others to exploit, potentially catalyzing a chain reaction of speciation events across trophic levels. We tested for such sequential radiation in the *Rhagoletis pomonella* (Diptera: Tephritidae) complex, a model for sympatric speciation via host plant shifting. We report that the parasitic wasp *Diachasma alloeum* (Hymenoptera: Braconidae) has formed new incipient species as a result of specializing on diversifying fly hosts, including the recently derived apple-infesting race of *R. pomonella*. Furthermore, we show that traits that differentially adapt *R. pomonella* flies to their host plants have also quickly evolved and serve as ecological barriers to reproduction, isolating the wasps. Speciation therefore cascades as the effects of new niche construction move across trophic levels.

The idea that species induce speciation has been inferred to explain current and past patterns of biodiversity by paleontologists, ecologists, and evolutionary biologists alike (*1*–*3*). However, this hypothesis of sequential radiation is difficult to directly test in nature. Examples such as adaptive radiations after mass extinctions (*4*), species richness in the tropics (*1*), and the increased diversity of insects having herbivorous life styles (*5*, *6*) have mainly been investigated on the basis of phylogenetic inference and/or correlative analyses.

Host plant–specific phytophagous insects and their parasites may be good candidates for testing the sequential radiation hypothesis (*7*). This is because new resource opportunities become available when a plant-eating insect diversifies by shifting and adapting to a novel host plant, with its guild of associated parasites potentially following suit and speciating in kind. Unfortunately, a lack of historical and biogeographic information concerning host shifting and the absence of a free-living parasite life stage often complicate our understanding of plant-insect-parasite systems. In these cases, cocladogenesis (cospeciation resulting from parallel allopatry of interacting organisms) rather than the cascading effects of shifting host ecology across trophic levels could trigger codiversification (*8*). One cannot rule out that insect and parasitoids became separated in tandem from other conspecifics and evolved into new species as a consequence of their shared physical isolation.

[1]Department of Biological Sciences, University of Notre Dame, Galvin Life Sciences Building, Notre Dame, IN 46556, USA. [2]Department of Entomology and Nematology, University of Florida Citrus Research and Education Center, 700 Experiment Station Road, Lake Alfred, FL 33850, USA. [3]Department of Entomology and Lyman Briggs College, Michigan State University, East Lansing, MI 48824, USA.

*Present address: Department of Entomology, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA. To whom correspondence should be addressed. E-mail: aaforbes@ucdavis.edu

†Present address: Wissenschaftskolleg zu Berlin Institute for Advanced Study, Wallotstrasse 19, D14193 Berlin, Germany.

The *Rhagoletis pomonella* sibling species complex is a model for speciation in the absence of geographic isolation via host plant shifting. The natural history of these flies, including the recent sympatric host shift of the species *R. pomonella* from hawthorn to introduced, domesticated apple within the past 150 years (*9*), suggests that they did not evolve as a result of passive cocladogenesis. Therefore, we tested for sequential radiation of its specialist parasitoid wasp, *Diachasma alloeum* [see supporting online material (SOM) text for parallels between fly and wasp biology], by examining whether wasps attacking the ancestral hawthorn and the derived apple races of *R. pomonella*, as well as the closely related sibling species *R. mendax* (host: blueberry, *Vaccinium* spp.) and *R. zephyria* (snowberry, *Symphoricarpos* spp.), display patterns of host-related genetic variation similar to those of the flies. *D. alloeum* only attacks *R. pomonella* complex flies found on these four host plants (*10*). We also investigated whether wasps differed by the same type of host plant–specific mating and diapause traits that are ecologically isolating the flies.

The pattern of genetic differentiation among *D. alloeum* populations was similar to that for *R. pomonella* flies (*11*, *12*). Mitochondrial DNA (mtDNA) cytochrome oxidase I (COI) sequences displayed only modest host-related differentiation for the wasps (Fig. 1A, figs. S1 and S2, and tables S1 and S2). Therefore, the wasps are apparently of relatively recent origin and do not represent highly genetically diverged cryptic sibling species. However, we did find a mtDNA haplotype in snowberry wasps that was not found in the other wasp populations (fig. S2 and table S2). In addition, a common mtDNA haplotype in apple, hawthorn, and blueberry wasps was not present in the snowberry population. Thus, in a mtDNA genetic distance network, the snowberry wasp was offset from the other taxa (Fig. 1A).

Nine of 21 microsatellite loci analyzed (DA003, DA013, DA019, DA150, DA174, DA183, DA192, DA202, and DA205) displayed consistent host-related allele frequency differences between at least two of the wasp populations (table S3 and SOM appendix), much like allozymes for *R. pomonella* flies (*13*, *14*). Locus DA003 showed particularly pronounced frequency differences (see Fig. 2 for graph of combined DA003 196+200+204 allele frequencies). Indeed, allele 196 was common in apple (mean frequency = 0.292; n = 291 total alleles scored), blueberry (0.179, n = 246), and snowberry (0.257, n = 101) wasp populations but absent from all hawthorn wasps (n = 385 alleles scored; SOM appendix). The lack of allele 196 in hawthorn wasps for locus DA003 suggests that there is little or no effective interhost gene flow from apple, blueberry, or snowberry wasp populations into the hawthorn population. Neighbor-joining trees for the nine loci displaying host-related divergence (Fig. 1B) and the full 21-loci data set (fig. S3) separated hawthorn and blueberry wasp populations at different ends of the networks. The lone snowberry wasp site analyzed was genetically most closely associated with blueberry wasps, whereas the apple wasps were placed at an intermediate position between blueberry and hawthorn populations (Fig. 1B).

We next investigated whether the same host-related adaptations ecologically isolating *R. pomonella* flies were responsible for genetically differentiating the wasps. Because *Rhagoletis* flies mate on or near the fruit of their respective host plants (*15*–*17*), host choice generates prezygotic isolation and facilitates the evolution of host-specific performance traits. A field study of free-living adult *D. alloeum* in a blueberry patch in Fennville, MI, suggested that the wasps also use host fruit as a rendezvous site for mating. Before copulation, all male and female wasps that formed the 24 mating pairs recorded in the study were first observed flying toward blueberries from distances of 1 to 3 m away. In all cases, wasps flew close to blueberries, sometimes making contact with the fruit, before they initiated coupling on nearby host leaves [mean distance mating from fruit = 7.4 ± 6.3 cm (SD), n = 24, range = 3 to 30 cm]. The average time from first observation to mating was 172.2 ± 94.7 s SD.

Rhagoletis flies use the volatile compounds emitted from the surface of ripening fruit as olfactory cues to both find and discriminate among host plants for mating and oviposition (*18*). We thus tested whether wasps show similar discriminatory behavior for surface host fruit volatiles by using a y-tube olfactometer (fig. S4). We found that naïve adult wasps from apple, hawthorn, and blueberry populations positively oriented to the arm of the y-tube containing their natal fruit odor and were antagonized by nonnatal volatiles (Table 1). There was no qualitative difference between males and females in their behavioral responses to surface fruit volatiles. However, female hawthorn wasps displayed stronger preferences for natal hawthorn fruit volatiles (Fisher's exact test, P = 0.006, 1 df) and more pronounced antagonism to nonnatal apple volatiles (P = 0.030, 1 df) than male hawthorn wasps did. Snowberry wasps showed behavioral specificity for their natal fruit volatiles that was similar to that of the other parasites, but too few snowberry wasps were available for testing to draw definitive conclusions (Table 1). All wasps exhibited behavioral antagonism to the volatiles of flowering dogwood fruit (*Cornus florida*), a host for an undescribed sister taxon to *R. pomonella* that *D. alloeum* does not attack (*19*). These data suggest that fruit odor discrimination may act in a similar manner in the wasps as in the flies and may generate host-specific mating, resulting in an ecological barrier to gene flow in sympatry. Because we cannot mate or rear *D. alloeum* for multiple generations in the laboratory, it is unknown whether maternal effects and/or larval conditioning may contribute to genetically based fruit volatile discrimination. In the rare instances in which larval experience was inferred to influence an adult insect's preference, the effect was mediated via direct larval contact with plant compounds (*20*). But, because *D. alloeum* has no direct contact with surface fruit volatiles at any stage of its life before adult eclosion, larval conditioning is unlikely.

Diapause life-history differences represent a second critical ecological barrier to gene flow among *Rhagoletis* flies (*21*). Sympatric blueberry, apple, and hawthorn flies all eclose as adults at different times in the spring and/or summer, matching the timing of fruit ripeness on their respective

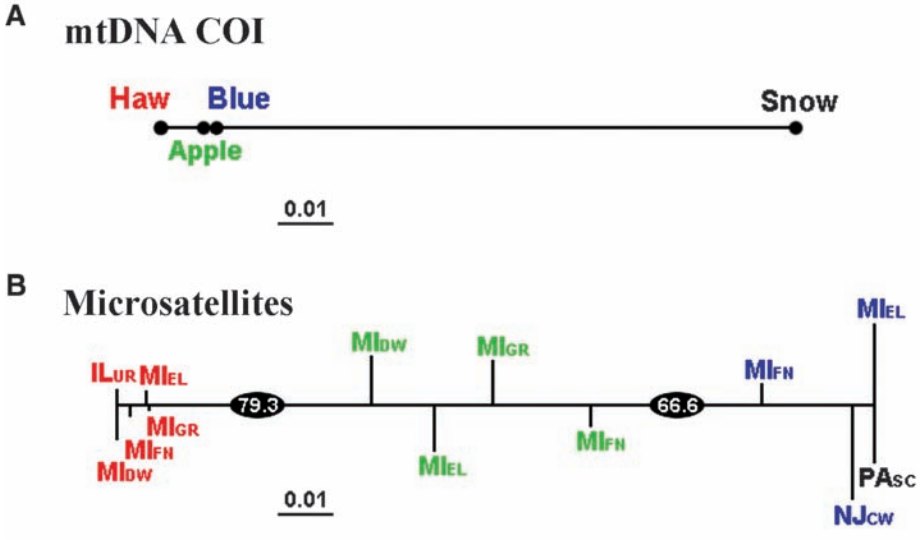


Fig. 1. Genetic distance networks depicting relationships of *D. alloeum* wasp populations attacking blueberry (blue), apple (green), hawthorn (red), and snowberry flies (black). (**A**) Fitch network created on the basis of Nei genetic distances of mtDNA haplotype frequencies in wasp populations. (**B**) Neighbor-joining Nei genetic distance network of the nine microsatellite loci showing consistent site-to-site host-related frequency differences. Black ovals give bootstrap confidence values (10,000 replicates). Site abbreviations are MI_{GR} = Grant, MI; MI_{HL} = Holland, MI; MI_{EL} = East Lansing, MI; MI_{FN} = Fennville, MI; MI_{DW} = Dowagiac, MI; IL_{UR} = Urbana, IL; PA_{SC} = State College, PA; NJ_{CW} = Chatsworth, NJ. See fig. S1 and table S1 for a map and descriptions, respectively, of study sites.

host plants (Fig. 3). Because *Rhagoletis* flies have only one generation per year and live for about a month in the field, the differences in the timing of eclosion can allochronically isolate the flies. Laboratory rearing experiments and field studies of *D. alloeum* indicated that similar eclosion time differences exist among blueberry, apple, and hawthorn wasps (Fig. 3 and table S4). The lone exception was a late fruiting apple from the Dowagiac, MI site where apple wasps had a relatively late mean eclosion time of 112.9 ± 1.01 days (SE) (n = 102) that overlapped with that of hawthorn wasps (111.1 ± 1.13 days, n = 44). Mean longevity of adult *D. alloeum* was determined to be less than 2 weeks in the laboratory (12.9 ± 0.4 days SE, n = 286). Assuming wasps have a similar lifespan in nature (a likely overestimate), eclosion time differences may decrease seasonal overlap and mating opportunities between blueberry and apple wasps at the sympatric Fennville, MI site by 28.9%, between apple and hawthorn wasps by 29.8%, and between blueberry and hawthorn wasps by 74.5%.

For *R. pomonella*, the allozymes displaying host-related frequency differences correlate with eclosion time (*9*), tying together the genetics of host race formation with a trait that allochronically isolates the flies. We tested for similar relationships for the microsatellite loci separately in male and female wasps because of sex-related differences in eclosion time (male wasps eclose several days earlier than females). For *D. alloeum*, the microsatellite loci were significantly correlated with variation in eclosion time for hawthorn (r^2 stepwise multiple regression for females = 0.504, $P < 0.0001$, 52 df; r^2 males = 0.525, $P < 0.0001$, 46 df), apple (r^2 females = 0.163, $P = 0.0097$, 54 df; r^2 males = 0.657, $P < 0.0001$, 30 df), and blueberry (r^2 females = 0.686, $P < 0.0001$, 29 df; r^2 males = 0.511, $P < 0.0001$, 53 df; table S5) wasps. In particular, the locus DA003 was significantly correlated with eclosion time variation in blueberry wasps (r^2 females = 0.378, $P = 0.0003$, 29 df; r^2 males = 0.511, $P < 0.0001$, 53 df) but only moderately in female apple wasps (r^2 females = 0.101, $P = 0.0182$, 54 df) and was not a significant predictor for hawthorn wasps. The alleles 196+200+204 at locus DA003 were associated with earlier eclosion times in both female and male blueberry and apple wasps (fig. S5). These three alleles were present in highest frequency for blueberry wasps (0.790, n = 205 total alleles scored), intermediate for apple (0.485, n = 260), and lowest for hawthorn wasps (0.029, n = 385), corresponding to the order of eclosion from earliest to latest for these populations (Fig. 3). Locus DA003 therefore may represent a naturally segregating, major effect quantitative trait locus for diapause that is associated with ecological reproductive isolation among *D. alloeum* populations.

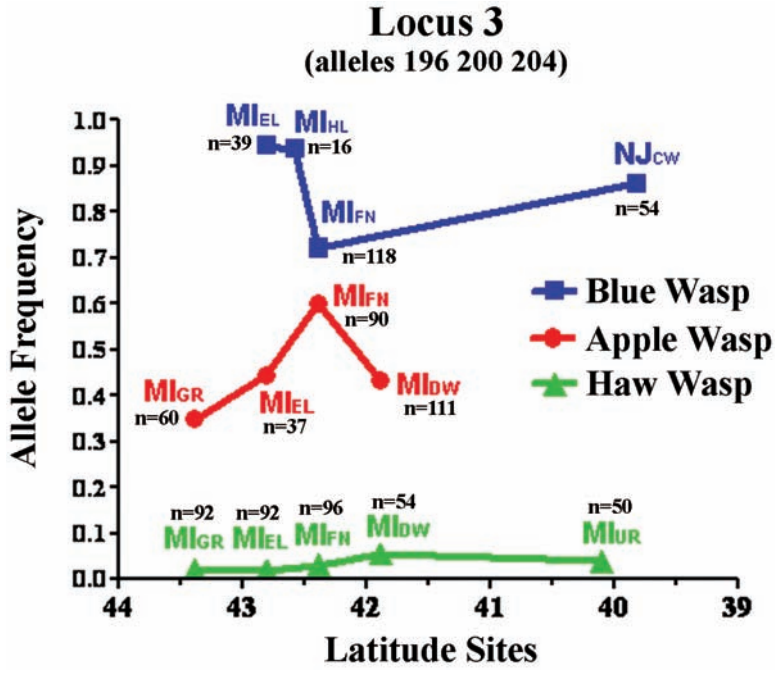


Fig. 2. Combined 196+200+204 allele frequencies for locus DA003 for *D. alloeum* wasp populations attacking blueberry (blue sites), apple (green sites), and hawthorn (red sites) arrayed by latitude in the United States. n is number of alleles scored for site. See Fig. 1 legend for site abbreviations.

Table 1. Percentages of increase (+ indicates preference) or decrease (–, antagonism or avoidance) in the behavioral orientation of populations of *D. alloeum* wasps to the arm of a y-tube olfactometer containing the indicated host fruit volatiles compared with blank control experiments. Control values (bottom row) denote the percentage of wasps orienting to either the right or the left arm of the y-tube in experiments when no odor was present in both arms. There was no directional preference for either arm of the y-tube in the blank control runs, and the remaining ~23% of wasps that did not orient right or left failed to move up the y-tube from their release point. $^{**}P < 0.01$, $^{***}P < 0.001$, and †not significant.

Host fruit volatiles	Apple wasps n = 201	Blueberry wasps n = 110	Hawthorn wasps n = 330	Snowberry wasps n = 9
Apple	+54%***	–29%**	–18%**	–14%†
Blueberry	–23%***	+42%***	–28%***	–14%†
Hawthorn	–33%***	–34%***	+44%***	–14%†
Snowberry	–46%***	–11%†	–77%***	0%†
Dogwood	–36%***	–53%***	–16%***	–14%†
Control	38%	39%	38%	39%

Our results imply that sympatric host shifts of *R. pomonella* onto new plants may have triggered a reciprocal and rapid starburst of adaptive radiation for its parasitoid, *D. alloeum*. We found that host-related ecological effects initiating speciation for *Rhagoletis* rippled through the community and may have amplified diversity for the wasp. Although we were unable to unambiguously resolve the source of the recently formed apple wasp race in this study, the absence of the 196 allele at microsatellite locus DA003 in the hawthorn wasp race suggests that the hawthorn wasp race is not the sole progenitor of the apple wasps. Together with the mtDNA data, allele 196 and its association with early eclosion instead suggest that blueberry wasps may have given rise to the apple wasp population. It is possible, however, that apple wasps are hybrids resulting from crosses between hawthorn and blueberry wasps on the basis of both their intermediate position in the microsatellite distance network and their intermediate overlapping eclosion time. Regardless, our study demonstrates that the origin of the apple wasp was not a result of strict 1:1 cocladogenesis with apple-infesting *Rhagoletis* flies. One intriguing possibility is that the wasps themselves may be promoting *Rhagoletis* host shifts. Lower parasitism rates for *R. pomonella* in derived hosts like apple [i.e., enemy-free space (*22*)] could favor race formation for the fly, creating new niches promoting wasp divergence. We must also determine whether non–host-related premating and intrinsic postzygotic isolation exist in the genus *Diachasma*. The sequential radiation hypothesis predicts that if nonecological barriers exist they should arise after host-related barriers evolve.

We have presented one case study supporting sequential speciation. Other studies have identified potential examples (*7*), and, because *D. alloeum* is just one member of a guild of parasitic braconids attacking *R. pomonella*, it is possible that sympatric host races of *Diachasmimorpha mellea*

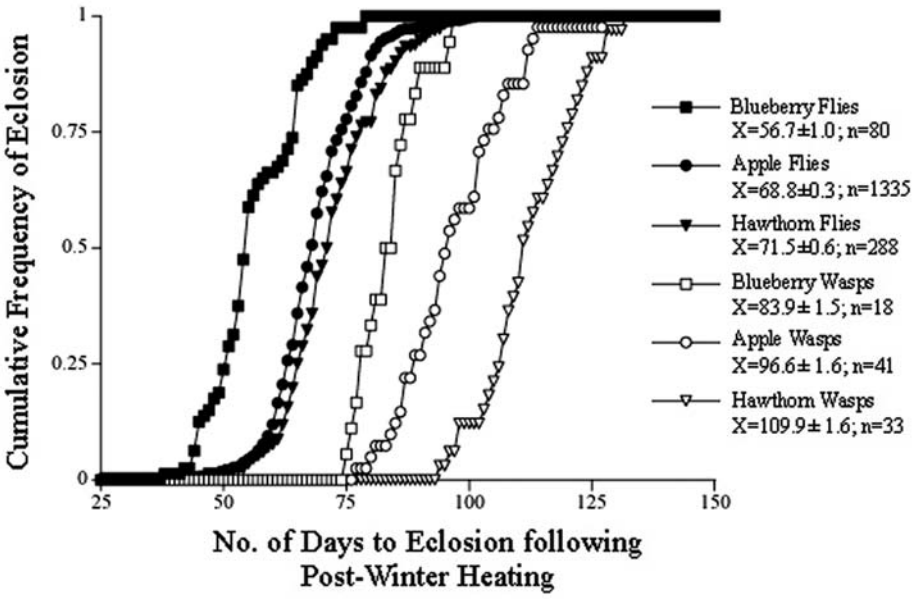


Fig. 3. Cumulative eclosion curves for blueberry-, apple-, and hawthorn-infesting *Rhagoletis* flies (solid shapes) and their associated *D. alloeum* wasp parasites (open shapes) at a sympatric field site near Fennville, MI. Mean times to adult emergence after overwintering (in days) and sample sizes (*n*) are given in the figure key.

and *Utetes canaliculatus* wasps are also associated with *R. pomonella*. Given that over half of all animals may be parasites in a broad sense (*23*), that there are more phytophagous insects than any other life form (*23*), and that 20% of all insects may be parasitic wasps (*24*), there is a world of opportunity for sequential speciation in nature.

References and Notes

1. B. C. Emerson, N. Kolm, *Nature* **434**, 1015 (2005).
2. D. H. Erwin, *Science* **308**, 1752 (2005).
3. C. Mitter, B. Farrell, B. Wiegmann, *Am. Nat.* **132**, 107 (1988).
4. N. J. Butterfield, *The Ecology of the Cambrian Region* (Columbia Univ. Press, New York, 2001).
5. D. Dilcher, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7030 (2000).
6. D. R. Strong, J. A. Lawton, T. R. E. Southwood, *Insects on Plants: Community Patterns and Mechanisms* (Blackwell Scientific, Oxford, 1984).
7. W. G. Abrahamson, C. P. Blair, in *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects*, K. J. Tilmon, Ed. (Univ. California Press, Berkeley, CA, 2008), pp. 188–202.
8. J. A. Coyne, H. A. Orr, *Speciation* (Sinauer, Sunderland, MA, 2004).
9. J. L. Feder, in *Endless Forms: Species and Speciation*, D. J. Howard, S. H. Berlocher, Eds. (Oxford Univ. Press, New York, 1998).
10. R. A. Wharton, P. M. Marsh, New world Opiinae (Hymenoptera: Braconidae) parasitic on Tephritidae (Diptera). *J. Wash. Acad. Sci.* **68**, 147 (1978).
11. J. J. Smith, G. L. Bush, *Mol. Phylogenet. Evol.* **7**, 33 (1997).
12. S. H. Berlocher, *Evolution* **54**, 543 (2000).
13. J. L. Feder, C. A. Chilcote, G. L. Bush, *Nature* **336**, 61 (1988).
14. B. A. McPheron, D. C. Smith, S. H. Berlocher, *Nature* **336**, 64 (1988).
15. J. L. Feder *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7990 (1994).
16. R. J. Prokopy, E. W. Bennett, G. L. Bush, *Can. Entomol.* **103**, 1405 (1971).
17. R. J. Prokopy, E. W. Bennett, G. L. Bush, *Can. Entomol.* **104**, 97 (1972).
18. C. Linn *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11490 (2003).
19. S. H. Berlocher, *Heredity* **83**, 652 (1999).
20. T. C. J. Turlings, F. L. Wäckers, L. E. M. Vet, W. J. Lewis, J. H. Tumlinson, in *Insect Learning: Ecological and Evolutionary Perspectives*, D. R. Papaj, A. C. Lewis, Eds. (Chapman and Hall, New York, 1993), p. 398.
21. J. L. Feder, S. Berlocher, S. B. Opp, in *Genetic Structure in Natural Insect Populations: Effects of Host Plants and Life History*, S. Mopper, S. Strauss, Eds. (Chapman and Hall, New York, 1998), pp. 408–441.
22. J. L. Feder, *Ecology* **76**, 801 (1995).
23. P. W. Price, *Evolutionary Biology of Parasites* (Princeton Univ. Press, Princeton, NJ, 1980).
24. J. La Salle, I. D. Gauld, *Redia* **74**, 315 (1991).

25. We thank M. Aluja, S. H. Berlocher, H. R. Dambroski, F. Footman, T. M. Forbes, E. Luecke, A. P. Michel, M. A. F. Noor, J. Rull, D. Schwarz, S. Velez, F. Wang, and R. A. Wharton for their technical, physical, intellectual, and moral support. Funded by a NSF doctoral dissertation improvement grant to A.A.F. and J.L.F., an American Museum of Natural History Theodore Roosevelt Fund Grant to A.A.F., and support to J.L.F. from the Wissenschaftskolleg zu Berlin. GenBank accession numbers for the mtDNA COI sequences are EU881512 to EU881682.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5915/776/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Tables S1 to S5
SOM Appendix

8 October 2008; accepted 9 December 2008
10.1126/science.1166981

Evolution of the *Drosophila* Nuclear Pore Complex Results in Multiple Hybrid Incompatibilities

Shanwu Tang[1] and Daven C. Presgraves[1,2]

Speciation often involves the evolution of incompatible gene interactions that cause sterility or lethality in hybrids between populations. These so-called hybrid incompatibilities occur between two or more functionally divergent loci. We show that the *nucleoporin 160kDa* (*Nup160*) gene of the fruitfly *Drosophila simulans* is incompatible with one or more factors on the *D. melanogaster* X chromosome, causing hybrid lethality. *Nup160* encodes a nuclear pore complex protein and shows evidence of adaptive evolution. Furthermore, the protein encoded by *Nup160* directly interacts with that of another hybrid lethality gene, *Nup96*, indicating that at least two lethal hybrid incompatibility genes have evolved as byproducts of divergent coevolution among interacting components of the *Drosophila* nuclear pore complex.

As species diverge from one another, they accumulate genetic substitutions that function within their own genomic background but, when brought together in hybrids, can disrupt gametogenesis (causing hybrid sterility) or development (causing hybrid lethality) (*1*, *2*). The evolution and genetics of these hybrid incompatibilities have been shown to follow specific rules. For instance, hybrid incompatibilities tend to accumulate gradually as species diverge (*3*); behave as partial recessives in hybrids (*4*); follow Haldane's rule [that is, the preferential sterility or inviability of hybrids of the heterogametic (XY or ZW) sex (*5*)]; and accumulate disproportionately on the X chromosome (that is, the large X-effect) (*6*, *7*). The molecular biology of hybrid incompatibilities has revealed that five of six hybrid incompatibility genes identified so far show signatures of recurrent adaptive evolution (*8*–*13*), which suggests that hybrid sterility and inviability generally evolve as incidental byproducts of positive natural selection.

To test whether this emerging molecular rule of speciation holds for additional hybrid incompatibility loci, we performed a genetic screen for lethal hybrid incompatibilities between *Drosophila melanogaster* and *D. simulans*, two species that diverged ~3 million years ago. Because all hybrids between *D. melanogaster* and

D. simulans are sterile, we performed a chromosomal deletion (deficiency) screen to identify recessive lethal hybrid-incompatibility factors in the *D. simulans* autosomal genome in the F_1 generation (Fig. 1A) (*14*). By screening ~70% of the *D. simulans* autosomal genome with ~200 deficiencies, we identified 20 small regions that cause hybrid lethality when combined with a hemizygous *D. melanogaster* X chromosome (*14*, *15*). In one of these 20 regions, we mapped a new hybrid lethality gene using chromosomal deficiencies from *D. melanogaster* to a region that includes 21 genes (Fig. 1B, and Table 1, lines 1 to 5) and then performed single-locus complementation tests (table S1). Only one lesion, *PBac{RB}RfC38^{e00704}*, uncovers hybrid lethality (Table 1, line 6). To confirm that hybrid lethality is uncovered by the *PBac{RB}RfC38^{e00704}* mutation (a *piggyBac* transposon insertion) and not an unidentified linked mutation, we generated revertant chromosomes that were genetically identical to the original *PBac{RB}RfC38^{e00704}* chromosome, except that the *piggyBac* insertion is precisely excised (fig. S1). We recovered eight independent revertant chromosomes that lacked the *piggyBac* insertion (*PBac{RB}RfC38^{e00704}-R1* to *-R8*), and all were viable in hybrid males (table S2), confirming that the *piggyBac* insertion unmasks hybrid lethality.

To determine which gene (or genes) is disrupted by *PBac{RB}RfC38^{e00704}*, we analyzed the genomic flanking sequence of the insertion (*16*) and found that the overlapping 3′ untranslated regions of two genes, *CG4738* and *RfC38*, are disrupted (Fig. 1B). Of these two genes, *CG4738* is the stronger candidate on the basis of three observations: DNA sequences are more diverged between species at *CG4738* than at *RfC38* (58 fixed replacement differences versus 1, respectively); disruption of *RfC38^{mel}* is not generally associated with hybrid lethality because two other transposable element insertions that disrupt *RfC38* but not *CG4738* (*P{lacW}RfC38^{k13807}* and *PBac{WH}RfC38^{f07177}*) do not unmask hybrid lethality (table S1, lines 2 and 9); and *CG4738* encodes the predicted protein sequence of the *Drosophila* homolog of nucleoporin 160kDa (NUP160), a protein component of the nuclear pore complex. The latter finding is especially intriguing because we previously found that another nucleoporin gene, *Nup96*, causes hybrid lethality between these species (*11*).

Although compelling, all three lines of evidence are circumstantial. We reasoned that if the *D. simulans* allele, *Nup160^{sim}*, causes recessive hybrid lethality, then introduction of the *D. melanogaster* allele, *Nup160^{mel}*, should rescue otherwise inviable hybrid males. To test this hypothesis, we cloned the *Nup160^{mel}* cDNA into a vector that carries a mini-*white^{+}* (*w^{+}*) eye-color marker (conferring red eyes) and a GAL4-inducible *upstream activation sequence* (*UAS*) promoter (hereafter, we refer to this construct as *P{UAS-Nup160^{mel}}*) (*15*). We transformed *P{UAS-Nup160^{mel}}* into *D. simulans w; Lhr* flies and recovered four independent stocks with autosomal insertions of the *P{UAS-Nup160^{mel}}* construct. We then crossed *D. melanogaster PBac{RB}RfC38^{e00704}/CyO; P{Tub-Gal4}/TM3, Ser* females carrying a ubiquitously expressed GAL4-driver of *P{UAS-Nup160^{mel}}* expression with *D. simulans w; Lhr P{UAS-Nup160^{mel}}* males. These crosses produce four male zygotic genotypes in equal proportions (1:1:1:1), all of which inherited the *P{UAS-Nup160^{mel}}* construct. As expected, the two kinds of hybrid males inheriting *CyO* are viable (Table 2, line 1), and hybrid males inheriting the *PBac{RB}RfC38^{e00704}* insertion but no induced *Nup160^{mel}* expression (that is, no *P{Tub-Gal4}*) were lethal (Table 2, line 1). However, in contrast to their *PBac{RB}RfC38^{e00704}/sim; TM3/sim* brothers, hybrid males inheriting the *PBac{RB}RfC38^{e00704}* insertion and induced *Nup160^{mel}* expression, *PBac{RB}RfC38^{e00704}/sim; P{Tub-Gal4}/sim*, were viable (Table 2, line 1). Moreover, the ratio of the three viable classes of hybrid males is not significantly different from 1:1:1 ($\chi^2 = 3.69$, $P = 0.158$). These data show that ubiquitous expression of *D. melanogaster Nup160* rescues the hybrid lethality caused by the insertion of *PBac{RB}RfC38^{e00704}*, and that *D. simulans Nup160* is a lethal hybrid-incompatibility gene.

To confirm that *Nup160^{sim}* is incompatible with one or more factors on the *D. melanogaster* X chromosome, we switched the species origin of the X in hybrids. We constructed a *D. melanogaster C(1)A, y/Y; PBac{RB}RfC38^{e00704}/CyO* stock, which possesses two X^{mel} chromosomes fused to a single centromere that are transmitted together, and crossed females from this stock with *D. simulans w; Lhr* males. These crosses produce progeny with the same autosomal genotypes (and cytoplasm) as those in Fig. 1A, except that hybrid males inherit their X from *D. simulans*, whereas hybrid females inherit both X chromosomes from *D. melanogaster* (fig. S2). [Attached-X^{mel} hybrid females also inherit a Y^{sim} chromosome, but the Y does not determine sex in *Drosophila* and is not essential for viability in these species (*17*).] These crosses show that, whereas *X^{mel}/Y^{sim}; PBac{RB}RfC38^{e00704}/sim* hybrid males are lethal (Table 1, line 6), *X^{sim}/Y^{mel}; PBac{RB}RfC38^{e00704}/sim* hybrid males are viable (Table 1, line 7; Fisher's exact test, $P < 10^{-40}$). We also found that although *X^{mel}/X^{sim}; PBac{RB}RfC38^{e00704}/sim* hybrid females are viable (Table 1, line 6), attached-X^{mel}, *C(1)A,*

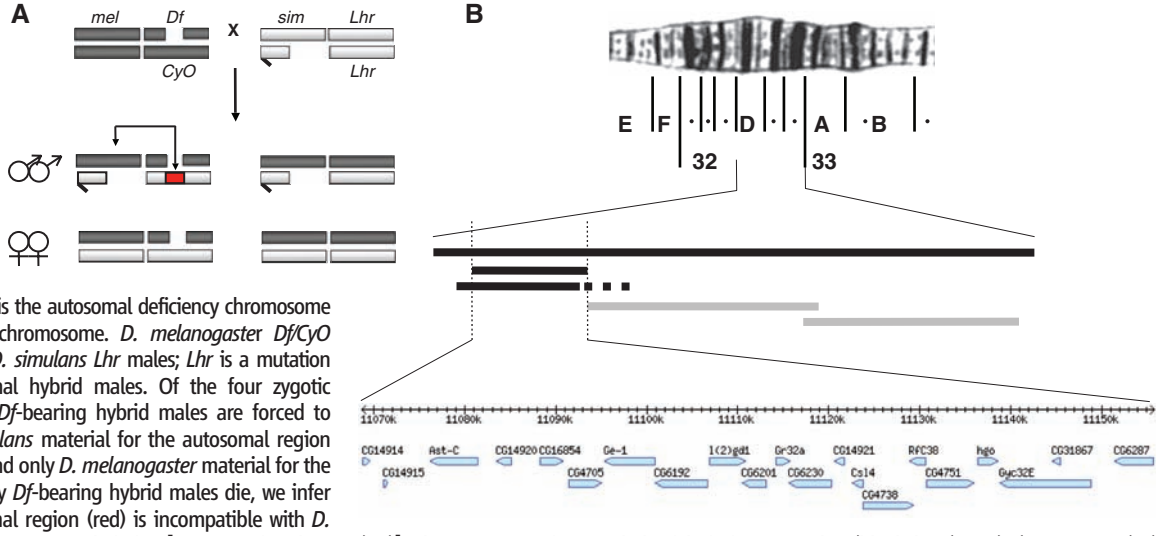


Fig. 1. (**A**) Deficiency mapping of recessive X-autosome hybrid-lethal interactions in F_1 hybrids between *D. melanogaster* (*mel*, gray) and *D. simulans* (*sim*, white). The sex chromosomes (left, Y chromosome with hook) and one representative pair of autosomes (right) are shown, where *Df* is the autosomal deficiency chromosome and *CyO* is the balancer chromosome. *D. melanogaster Df/CyO* females were crossed with *D. simulans Lhr* males; *Lhr* is a mutation that rescues ordinarily lethal hybrid males. Of the four zygotic genotypes produced, only *Df*-bearing hybrid males are forced to develop using only *D. simulans* material for the autosomal region exposed by the deficiency and only *D. melanogaster* material for the X chromosome. Thus, if only *Df*-bearing hybrid males die, we infer that a *D. simulans* autosomal region (red) is incompatible with *D. melanogaster* X chromosome, causing lethality [for more details, see (*14*)]. The hybrid-lethal interactions identified in this screen would normally afflict advanced-generation (for example, F_2) but not F_1 hybrids because both partners in these X-autosome incompatibilities are recessive. (**B**) Fine-scale mapping results in region 32D on chromosome arm 2L. Three deficiencies fail to complement hybrid lethality in males (black bars): *Df(2L)FCK-20*, *Df(2L)Exel6027*, and *Df(2L)BSC36*; two deficiencies complement hybrid lethality in males (gray bars): *Df(2L)Exel6028* and *Df(2L)Exel6029*. Together, these deficiencies implicate one of 21 genes in hybrid lethality that maps to a 89.6-kb region in the cytological interval 32D02;32D05.

[1]Department of Biology, University of Rochester, Rochester, NY 14627, USA. [2]Radcliffe Institute for Advanced Study, Harvard University, Cambridge, MA 02138, USA.

y/Y^{sim}*; PBac{RB}RfC38*^{e00704}*/sim* hybrid females are lethal (Table 1, line 7; Fisher's exact test, $P < 10^{-29}$). These results show, respectively, that *Nup160*^{sim} causes hybrid lethality via an incompatible epistatic interaction with one or more factors on the *D. melanogaster* X chromosome and that *Nup160*^{sim}-dependent hybrid lethality occurs in both sexes.

To study the evolution of *Nup160*, we surveyed DNA sequence polymorphism and divergence within and between *D. melanogaster* and *D. simulans* from 12 isofemale lines of each species collected from Zimbabwe (*15*). We tested the neutral mutation-drift model of protein evolution with the McDonald-Kreitman (MK) test (*18*). The MK test rejects the neutral hypothesis for *Nup160* and instead reveals a highly significant excess of fixed replacement differences between species (Table 3, line 1), consistent with recurrent adaptive protein evolution. To distinguish whether adaptive evolution at *Nup160* occurred during the history of one or both species, we polarized substitutions along lineages by using sequence data from the outgroup species, *D. yakuba*, and contrasted polymorphic and fixed mutations in the *D. melanogaster* and *D. simulans* lineages separately. These lineage-specific MK tests showed that *Nup160* has experienced adaptive evolution in both lineages (Table 3, lines 2 and 3).

Despite its history of recurrent positive selection, we found no evidence for a recent selective sweep at *Nup160* in either the *D. melanogaster* or *D. simulans* lineages. Selective sweeps should reduce local nucleotide diversity (*19*, *20*) and shift the distribution of allele frequencies toward an excess of both rare variants (*21*) and high-frequency derived variants (*22*). We found that mean silent nucleotide diversity, π, at *Nup160* is similar to other autosomal loci sampled from African populations of these species (*23*, *24*): π = 0.0160 in *D. melanogaster*, and π = 0.0335 in *D. simulans*. Moreover, coalescent simulations show that neither Tajima's *D* (*21*) nor Fay and Wu's *H* (*22*), two summaries of the allele frequency spectrum, deviate significantly from standard neutral equilibrium expectations in either species (*D. melanogaster*, $D = -0.511$ and $H = -14.273$; *D. simulans*, $D = 0.134$ and $H = 7.455$; $P \geq 0.124$ in all tests). These findings suggest that recurrent positive selection drove sequence evolution at *Nup160*, but that none of the selected substitutions occurred recently (that is, within the past ~0.1 N_e generations, where N_e is the effective population size).

Our current and previous (*11*) findings show that two rapidly evolving autosomal genes from *D. simulans*, *Nup160*^{sim} and *Nup96*^{sim}, are incompatible with one or more factors on the *D. melanogaster* X chromosome. NUP160 and NUP96 are members of the NUP107 subcomplex, a subset of interacting nucleoporins that together form a stable architectural component of the nuclear pore complex (NPC), the macromolecular structures that mediate all molecular traffic between the nucleus and cytoplasm (*25*). The structure, function, and interactions among particular nucleoporins seem largely conserved among eukaryotes (*26*). In addition to NPC functions, members of the NUP107 subcomplex also contribute to kinetochore function (*27*), cell cycle progression (*28*), and dosage compensation in *Drosophila* males (*29*). In *D. melanogaster*, three other *Nup107* subcomplex genes (including *Nup96*) have histories of adaptive evolution, similar to that observed here for *Nup160*, whereas in *D. simulans* six others have histories of adaptive evolution (*30*). The components of the dosage compensation complex (*31*, *32*) and its X chromosome–binding sites have also evolved by positive selection between *D. melanogaster* and *D. simulans* (*33*). However, it is unlikely that the rapid evolution of *Nup160*, and other members of the *Nup107* subcomplex, reflects coevolution with the dosage compensation pathway or that hybrid inviability involves a disruption of dosage compensation in hybrids. For one, the rapid evolution of dosage compensation is largely limited to the *D. melanogaster* lineage (*31*–*33*) and thus cannot explain the especially rapid evolution of *Nup160* and the other *Nup107* subcomplex genes in *D. simulans*. For another, *Nup160*^{sim} kills X^{mel}/Y^{sim} hybrid males and $X^{mel} \cdot X^{mel}/Y^{sim}$ hybrid females, excluding a loss-of-function disruption of dosage compensation—a male-specific problem—as the cause of hybrid lethality. Instead, the recurrent bouts of adaptation of *Nup107* subcomplex genes are most likely driven by selection arising from evolutionary conflict involving pathogens, retrotransposons, or meiotic drive elements (*30*, *34*).

As part of the NUP107 subcomplex, NUP160 and NUP96 interact directly (*25*), raising the possibility that *Nup160*^{sim} and *Nup96*^{sim} are incompatible with the same X-linked factor(s) from *D. melanogaster*. The *D. melanogaster* allele *Nup153*^{mel} is a strong candidate X-linked partner for both *Nup96*^{sim} and *Nup160*^{sim}: *Nup153* has a history of adaptive evolution (*30*), its protein interacts physically with the NUP107 subcomplex (*35*), and it is the only known interacting nucleoporin encoded on the X chromosome. However, it is not clear whether *Nup160* and *Nup96* are involved in two distinct two-locus hybrid incompatibilities with the *D. melanogaster* X or whether they are components of one complex hybrid incompatibility (that is, one involving more than two loci). Theory predicts that complex hybrid incompatibilities should evolve more readily than simple hybrid incompatibilities (*36*), and there is the suggestion that complex hybrid incompatibilities may be typical [for example, (*9*, *37*)]. Regardless of whether nucleoporin-based hybrid lethality has a simple or complex basis, our study suggests that the adaptive coevolution of a large multi-protein complex may have given rise to multiple hybrid-incompatibility genes. These findings suggest that divergent coevolution among the interacting partners of macromolecular complexes, particularly those prone to evolutionary conflicts, may drive the evolution of molecular incompatibilities that contribute to speciation.

Table 1. Interspecific complementation mapping of a hybrid lethal factor on *D. simulans* chromosome arm 2L.

D. melanogaster maternal genotype	Cytological breakpoints	Hybrid females		Hybrid males	
		–/sim	*CyO/sim*	*–/sim*	*CyO/sim*
Df(2L)FCK-20/CyO	32D1;32F1-3	471	504	0	88
Df(2L)Exel6027/CyO	32D2;32D5	338	296	1	261
Df(2L)BSC36/CyO	32D1;32F3	134	137	0	76
Df(2L)Exel6028/CyO	32D5;32E4	204	183	151	127
Df(2L)Exel6029/CyO	32E4;32F2	265	274	254	192
PBac{RB}RfC38^{e00704}*/CyO*	32D4	359	306	0	219
C(1)A, y; PBac{RB}RfC38^{e00704}*/CyO*	32D4	0	93	114	120

References and Notes

1. J. A. Coyne, H. A. Orr, *Speciation* (Sinauer, Sunderland, MA, 2004).
2. T. Dobzhansky, *Genetics and the Origin of Species* (Columbia Univ. Press, New York, 1937).

Table 2. Ubiquitous expression of a *D. melanogaster Nup160* transgene rescues hybrid lethality. Progeny counts are pooled results from four different *D. simulans Lhr P{UAS-Nup160*^{mel}*}* stocks (similar results hold across replicates). All progeny from these crosses inherit *Lhr* and *P{UAS-Nup160*^{mel}*}*, which for simplicity are not shown.

Line	Second chromosome genotype Third chromosome genotype	Hybrid autosomal genotypes			
		PBac{RB}RfC38^{e00704}*/sim; TM3, Ser/sim*	*CyO/sim; TM3, Ser/sim*	*PBac{RB}RfC38*^{e00704}*/sim; P{Tub-Gal4}/sim*	*CyO/sim; P{Tub-Gal4}/sim*
Hybrid males, X^{mel}/Y^{sim}		0	119	112	141
Hybrid females, X^{mel}/X^{sim}		191	204	224	219

Table 3. Recurrent adaptive evolution at the hybrid lethality gene *Nup160* in both *D. melanogaster* and *D. simulans*. R, replacement; S, synonymous.

	Polymorphic		Divergent				Fisher's exact *P* value
	R	S	R/S	R	S	R/S	
D. melanogaster–D. simulans pooled	27	154	0.175	58	64	0.906	9.3×10^{-10}
D. melanogaster lineage	10	48	0.208	19	33	0.576	0.0299
D. simulans lineage	11	85	0.129	26	21	1.238	7.1×10^{-8}

3. J. A. Coyne, H. A. Orr, *Evolution* **43**, 362 (1989).
4. M. Turelli, H. A. Orr, *Genetics* **140**, 389 (1995).
5. J. B. S. Haldane, *J. Genet.* **12**, 101 (1922).
6. J. A. Coyne, H. A. Orr, in *Speciation and Its Consequences*, D. Otte, J. Endler, Eds. (Sinauer, Sunderland, MA, 1989), pp. 180–207.
7. D. C. Presgraves, *Trends Genet.* **24**, 336 (2008).
8. D. A. Barbash, D. F. Siino, A. M. Tarone, J. Roote, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5302 (2003).
9. N. J. Brideau *et al.*, *Science* **314**, 1292 (2006).
10. J. P. Masly, C. D. Jones, M. A. F. Noor, J. Locke, H. A. Orr, *Science* **313**, 1448 (2006).
11. D. C. Presgraves, L. Balagopalan, S. M. Abmayr, H. A. Orr, *Nature* **423**, 715 (2003).
12. C.-T. Ting, S.-C. Tsaur, M.-L. Wu, C.-I. Wu, *Science* **282**, 1501 (1998).
13. J. Wittbrodt *et al.*, *Nature* **341**, 415 (1989).
14. D. C. Presgraves, *Genetics* **163**, 955 (2003).
15. Materials and methods are available as supporting material on *Science* Online.
16. S. T. Thibault *et al.*, *Nat. Genet.* **36**, 283 (2004).
17. M. Ashburner, K. G. Golic, R. S. Hawley, *Drosophila: A Laboratory Handbook* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 2, 2005).
18. J. H. McDonald, M. Kreitman, *Nature* **351**, 652 (1991).
19. N. L. Kaplan, R. R. Hudson, C. H. Langley, *Genetics* **123**, 887 (1989).
20. J. Maynard Smith, J. Haigh, *Genet. Res.* **23**, 23 (1974).
21. F. Tajima, *Genetics* **123**, 585 (1989).
22. J. C. Fay, C.-I. Wu, *Genetics* **155**, 1405 (2000).
23. P. Andolfatto, *Mol. Biol. Evol.* **18**, 279 (2001).
24. S. Hutter, H. Li, S. Beisswanger, D. DeLorenzo, W. Stephan, *Genetics* **177**, 469 (2007).
25. M. Suntharalingam, S. R. Wente, *Dev. Cell* **4**, 775 (2003).
26. E. Bapteste, R. L. Charlebois, D. MacLeod, C. Brochier, *Genome Biol.* **6**, R85 (2005).
27. M. Zuccolo *et al.*, *EMBO J.* **26**, 1853 (2007).
28. P. Chakraborty *et al.*, *Dev. Cell* **15**, 657 (2008).
29. S. Mendjan *et al.*, *Mol. Cell* **21**, 811 (2006).
30. D. C. Presgraves, W. Stephan, *Mol. Biol. Evol.* **24**, 306 (2007).
31. M. T. Levine, A. K. Holloway, U. Arshad, D. J. Begun, *Genetics* **177**, 1959 (2007).
32. M. A. Rodriguez, D. Vermaak, J. J. Bayes, H. S. Malik, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15412 (2007).
33. D. Bachtrog, *Genetics* **180**, 1123 (2008).
34. D. C. Presgraves, *Bioessays* **29**, 386 (2007).
35. S. K. Vasu *et al.*, *J. Cell Biol.* **155**, 339 (2001).
36. H. A. Orr, *Genetics* **139**, 1805 (1995).
37. D. A. Barbash, *Genetics* **176**, 543 (2007).

38. We thank the Drosophila Genome Resource Center for cloning vectors and the *D. melanogaster Nup160* cDNA; the Bloomington Stock Center and the Exelixis Stock Collection at Harvard Medical School for fly stocks; and V. Cattani, P. Gerard, C. Meiklejohn, A. Orr, A. Sweigart, and two anonymous reviewers for comments. This work was supported by funds to D.C.P. from NIH grant R01-GM079543, the University of Rochester, and the Radcliffe Institute for Advanced Study at Harvard University. Sequences have been deposited in GenBank under accession numbers FJ600378 to FJ600401.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5915/779/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 and S2
References

28 November 2008; accepted 6 January 2009
10.1126/science.1169123

Queen Ants Make Distinctive Sounds That Are Mimicked by a Butterfly Social Parasite

Francesca Barbero,[1,2] Jeremy A Thomas,[2,3]* Simona Bonelli,[1] Emilio Balletto,[1] Karsten Schönrogge[3]

Ants dominate terrestrial ecosystems through living in complex societies whose organization is maintained via sophisticated communication systems. The role of acoustics in information exchange may be underestimated. We show that *Myrmica schencki* queens generate distinctive sounds that elicit increased benevolent responses from workers, reinforcing their supreme social status. Although fiercely defended by workers, ant societies are infiltrated by specialist insects that exploit their resources. Sounds produced by pupae and larvae of the parasitic butterfly *Maculinea rebeli* mimic those of queen ants more closely than those of workers, enabling them to achieve high status within ant societies. We conclude that acoustical mimicry provides another route for infiltration for ~10,000 species of social parasites that cheat ant societies.

The main attribute that enables ants to dominate in most terrestrial ecosystems is their ability to live in complex societies whose cohesion is regulated by highly developed communication systems (*1*). Information is primarily transferred through exchanging distinctive semiochemicals, which, combined with physical contact, initiate and integrate different behaviors by colony members, providing a mechanism for caste determination and for recognizing and behaving appropriately to nestmates, non-kin ants, and potential intruders (*1*–*3*). In addition, the adults in 4 of the 11 subfamilies stridulate by scraping a plectrum located on an anterior segment of the abdomen (post-petiole) across a file (*pars stridens*) on the first segment of their gaster (*1*, *4*).

About 10,000 other species of invertebrates from 11 orders have evolved adaptations to infiltrate ant societies and feed as social parasites on the rich resources concentrated inside nests (*1*, *5*). Achieving this penetration generally includes the corruption through mimicry of their host's communication systems. The lycaenid butterfly *Maculinea rebeli* is among the better-understood examples. After briefly feeding on gentians, its final instar larvae are carried by *Myrmica schencki* workers (in Western Europe) into their nest, where the butterflies acquire ~98% of their ultimate biomass before pupating 11 to 23 months later (*5*, *6*). Inside the brood chambers, *M. rebeli* caterpillars beg like ant larvae and secrete semiochemicals that so closely mimic the surface hydrocarbons on *Myrmica schencki* larvae (*7*) that they are fed directly with regurgitations by the workers (*6*). However, neither begging nor chemical mimicry explains the high rank achieved by *M. rebeli* within its host's social hierarchy. For example, inert dummies painted with the surface pheromones of kin ant larvae or workers are retrieved in preference to dummies painted with *M. rebeli*'s mimetic allomones (*8*), yet living *M. rebeli* larvae are rescued in preference to ant larvae when a colony is disturbed (*9*). Furthermore, nurse workers kill and feed their own brood to the social parasite if food is scarce (*6*, *10*). We even observed *Myrmica schencki* queens treat *Maculinea rebeli* larvae or pupae like rivals (*11*) (fig. S1), whereas the workers regularly treat them like royalty (*6*). Lacking other cues, we speculated that the butterfly's elevated status might be achieved through mimicry of *Myrmica schencki* acoustics. Whereas *Myrmica* larvae are mute, distressed *Maculinea* larvae generate sounds that resemble the alarm stridulations of (adult) *Myrmica* workers (*12*). To be fully adaptive, the parasite would need to emulate the typical stridulation patterns of the most valued colony members, the queens. This, in

[1]Dipartimento di Biologia Animale e dell´Uomo Laboratorio di Zoologia, Turin, Università degli Studi di Torino, Via Accademia Albertina 13, 10123 Turin, Italy. [2]Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford,Wallingford, Oxfordshire, OX10 8BB, UK. [3]Department of Zoology, University of Oxford, Tinbergen Building, South Parks Road, Oxford, OX1 3PS, UK.

*To whom correspondence should be addressed. E-mail: jeremy.thomas@zoo.ox.ac.uk

turn, presupposes that queen stridulations differ from those of workers.

The substrate-borne (*13*) or airborne (*14*) calls of ants are generally considered to be a weakly developed means of communication (*1*), merely signaling alarm and the location of buried individuals, enhancing chemical recruitment, or (in *Pogonomyrmex*) indicating an end to mating (*1*, *3*, *15*–*17*). Two studies, however, suggested that different castes might produce distinctive signals: The major workers of *Atta cephalotes* make sounds that are more intense and carry further than those of their smaller nestmates (*16*), and the space between the ridges of the *pars stridens* of queens exceeds that of workers in four *Messor* species (*18*).

Here, we report that the stridulatory organ of *Myrmica schencki* shows a morphological distinction between workers and queens greater than those seen in *Messor*, with the queen possessing a 44% longer *pars stridens* ($P = 0.04$) and a 33% wider gap between the ridges ($P < 0.001$) (Fig. 1). We recorded the signals produced by unstressed workers and queens (*11*) [supporting online material (SOM) audio S1 and S2] and found that the sounds of *M. schencki* queens differ significantly from those of workers in their dominant frequency (Fig. 1, A and B) ($P = 0.014$) and overall acoustics (Fig. 2) ($P = 0.035$), although the pulse lengths and the pulse repetition frequencies (PRFs) of both castes are similar (Fig. 1). The first two attributes reflect the morphology of the stridulatory organs, whereas the PRF reflects the rhythm and speed with which the plectrum is scraped across the file.

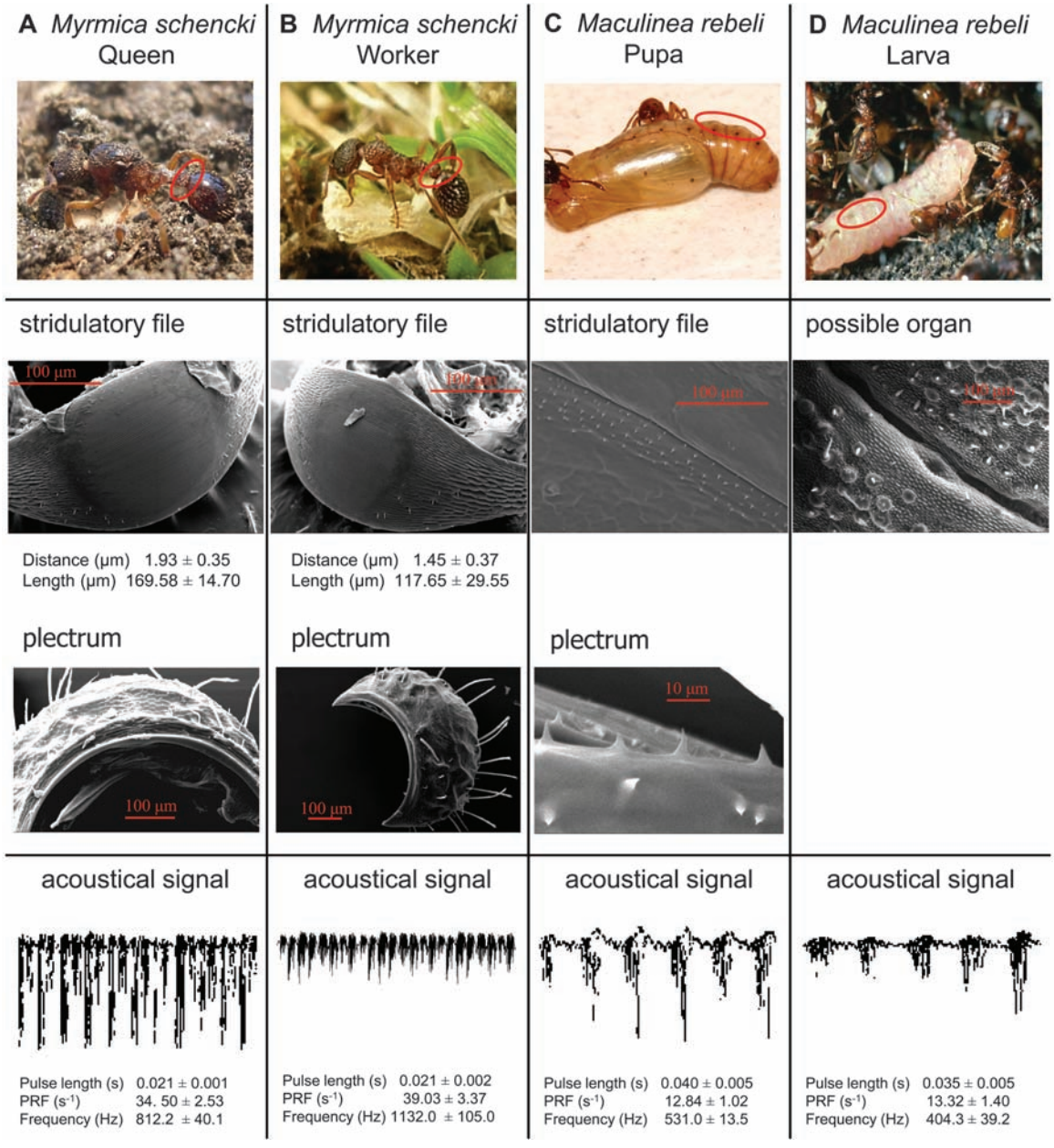


Fig. 1. Location, morphology, and sounds of the acoustical organs of (**A**) queens and (**B**) workers of the ant *Myrmica schencki* and (**C**) pupae and (**D**) larvae of its social parasite *Maculinea rebeli*. Measurements of the stridulatory file are the average distance between the parallel ribs and of length along the central line. Characteristic sound spectra are shown with measurements ± SE of the average pulse length (PL), PRF, and dominant frequency (DF). Linear mixed effect models show that pupal and larval sounds differ from sounds from queens and workers [likelihood ratio (LR) = 12.63, $P = 0.0004$ for PL and LR = 20.67, $P = 0.0005$ for PRF]; the DF of worker sounds differs from those of queens (LR = 10.16, $P = 0.0014$), pupae (LR = 16.65, $P = 0.0001$), and larvae (LR = 34.37, $P = 0.0002$); whereas the DF of queens is more similar to pupae (LR = 4.12, $P = 0.042$) than to larvae (LR = 13.43, $P = 0.0002$). Pupal and larval sounds do not differ significantly (LR = 0.81).

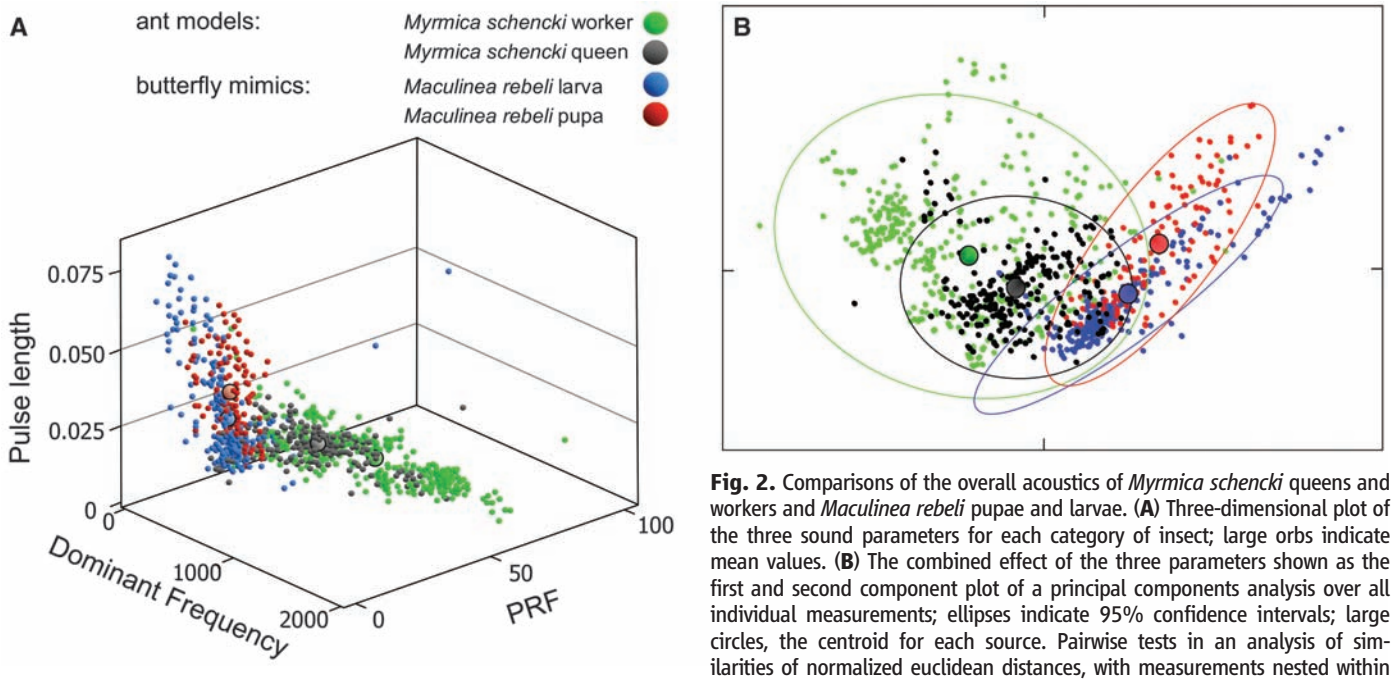


Fig. 2. Comparisons of the overall acoustics of *Myrmica schencki* queens and workers and *Maculinea rebeli* pupae and larvae. (**A**) Three-dimensional plot of the three sound parameters for each category of insect; large orbs indicate mean values. (**B**) The combined effect of the three parameters shown as the first and second component plot of a principal components analysis over all individual measurements; ellipses indicate 95% confidence intervals; large circles, the centroid for each source. Pairwise tests in an analysis of similarities of normalized euclidean distances, with measurements nested within individuals, showed that workers differ from queens ($R = 0.13$, $P = 0.035$), larvae ($R = 0.58$, $P < 0.001$), and pupae ($R = 0.64$, $P = 0.001$). Queens also differ from larvae ($R = 0.62$, $P < 0.001$) and pupae ($R = 0.90$, $P = 0.001$). There is no significant difference between larval and pupal sounds ($R = 0.02$). Mean larval and pupal calls are significantly closer ($P < 0.0001$ and $P < 0.001$, respectively) to queen sounds than to workers.

We played the sounds of both castes to undisturbed laboratory nests of *M. schencki* workers, together with controls consisting of a third miniature speaker that played computer-generated white noise and a fourth that was silent (*11*). No antagonistic or alarmed behavior was observed, except for a small repulsion of workers by white noise ($F_{5,66} = 4.33$, $P = 0.002$) (Fig. 3 and table S1). In contrast, the ant acoustics resulted in workers aggregating nonaggressively around the speaker, then tapping it with their antennae (antennation) or standing motionless on its surface in a posture similar to that adopted when they attend queens and objects of high value to their society (on-guard attendance) (Fig. 3 and table S1). The sounds of workers and queens elicited similar amounts of antennation, the main functions of which are to induce worker-worker recruitment, to smell nestmates, and to facilitate oral exchanges of food or pheromones (*1*). However, queen sounds induced significantly higher occurrences of on-guard attendance than did worker calls (posthoc test, least significant distance for workers–queens $P = 0.014$) (table S1), consistent with the exalted status and protection afforded to queens in the hierarchy of a colony (*1*). Combining these results, we suggest that acoustical communication within the vast subfamily Myrmicinae, to which *Messor* and *Myrmica schencki* belong, is more variable and conveys more social information within ant colonies than has previously been recognized. They also fulfill the strict adaptationist definition of biological communication, in which both the signal and response are adaptive (*19*).

We next recorded the sounds made by *Myrmica schencki*'s host-specific social parasite, *Maculinea*

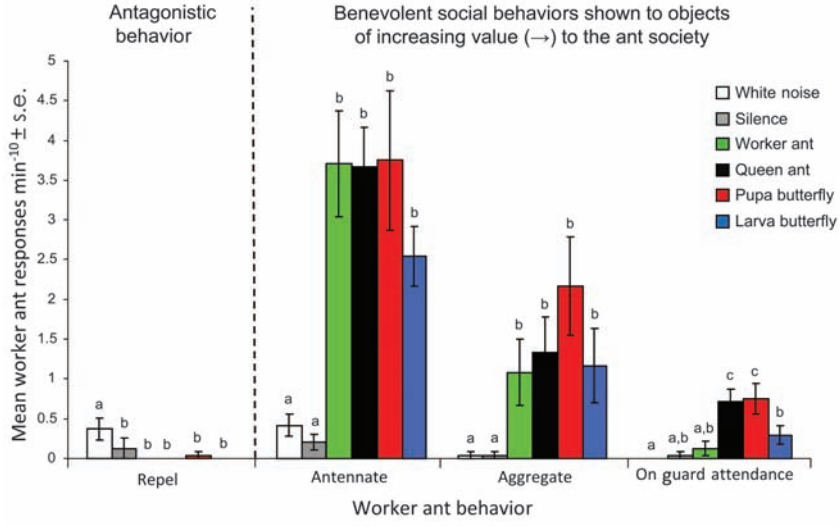


Fig. 3. Behavioral responses of *Myrmica schencki* colonies (10 workers) to sound recordings of *M. schencki* workers, *M. schencki* queens, *Maculinea rebeli* pupae, and *M. rebeli* larvae and to two controls (white noise and silence). Benevolent behaviors are in order of increasing respect, culminating in on-guard attendance. A significant overall difference in responses occurred within all four behaviors (repel $F_{5,66} = 4.33$, $P = 0.002$; aggregate $F_{5,66} = 4.82$, $P = 0.001$; antennate $F_{5,66} = 20.20$, $P < 0.001$; on-guard $F_{5,66} = 8.64$, $P < 0.001$). The letters above each column indicate pairwise posthoc tests of least significant differences ($P < 0.05$): The same letter indicates no significant difference within each type of behavior; different letters indicate significantly different responses (table S1).

rebeli, and, unlike an earlier study of *Maculinea* acoustics (*12*), we recorded pupae as well as larvae and did not stress the ants or butterflies (*11*). In contrast to the ants, pupal calls were generated by tooth-and-comb stridulatory organs (Fig. 1C and fig. S2) situated in the grooves between abdominal segments 5 and 8. We are unsure of the source of larval sounds, which may emanate from muscular contractions in the abdomen (*20*) or from sclerotized structures between segments 4 and 7 (Fig. 1D and fig. S2) that resemble the stridulatory organ of the mutualistic lycaenid *Arhopala madytus* (*21*). Despite the dissimilar sources, pupal and larval sounds resembled those made by the queens and workers of their host (Fig. 2A), but the similarity was 23% and

27% closer, respectively, to queen sounds than to those of workers (Fig. 2B) [mean normalized euclidean distances between individual butterflies' and ants' sounds are as follows: pupa-queen 2.47 ± (SE) 0.10, pupa-worker 3.03 ± 0.15, t = –3.14, df 87, $\text{distance}_{\text{pupa-queens}} < \text{distance}_{\text{pupa-workers}}$, $P < 0.001$; larva-queen 2.52 ± 0.11, larva-worker 3.21 ± 0.12, t = –4.32, df 237, $\text{distance}_{\text{larva-queens}} < \text{distance}_{\text{larva-workers}}$, $P < 0.001$]. The distributions in Fig. 2 also satisfy the concept that the perfect mimic should have maximal overlap with queen acoustics and minimal overlap with those of workers.

Playing recordings of *Maculinea* pupal calls to the same naïve cultures of *Myrmica schencki* workers resulted in enhanced benevolent responses similar to those elicited by queen ant sounds. We found no significant differences toward *Maculinea* pupal and *Myrmica* queen calls in any of the four behaviors scored, and pupal calls elicited six times more instances of royal on-guard attendance than occurred when worker sounds were played (Fig. 3 and table S1) ($P < 0.001$). Recordings of *M. rebeli* larvae induced lower worker responses and, despite eliciting 2.3 times more on-guard attendances than worker calls, did not differ significantly from responses toward worker sounds (Fig. 3 and table S1). We did not play *Maculinea* calls to queen ants but predict that they would provoke rivalry similar to that observed when live *Maculinea* pupae were artificially enclosed with *Myrmica schencki* queens (*11*) (fig. S1).

We suggest that regional host specificity in *Maculinea* populations is mediated first through chemical mimicry (*6*, *22*); but once the intruder is admitted and accepted as a member of a host society, it mimics adult ant acoustics (particularly queens) to advance its seniority toward the highest attainable position in the colony's hierarchy. Selection for accurate acoustical mimicry may have been stronger in pupae, which lack the main secretory organs of *M. rebeli* larvae and offer only weak rewards to tending workers.

The young stages of other *Maculinea* species make similar pulsed sounds to *M. rebeli* (*12*): All differ substantially from those of other studied Lycaenidae, most of which are commensals or mutualists or have no known relationship with ants (*12*, *23*–*27*). None of the latter mimics the acoustics of associated ants in obvious ways, although the sound of one strongly mutualistic species attracts workers (*23*–*26*). Thus, the use of acoustics to signal superior status to ants is unlikely to be a basal trait in the Lycaenidae, although we might expect it in *Phengaris*, the sister genus to *Maculinea*.

Beyond the Lycaenidae, ~10,000 species of ant social parasites may exist (*5*), particularly among other Lepidoptera, Coleoptera, Diptera, and inquiline ants (*1*, *6*). If acoustics plays the role that we suggest in reinforcing an ant's hierarchical status, it seems likely that this cue has evolved in other social parasites to infiltrate and exploit their societies.

References and Notes

1. B. Hölldobler, E. O. Wilson, *The Ants* (Springer, Berlin, 1990).
2. A. Lenoir, P. D'Ettorre, C. Errard, A. Hefetz, *Annu. Rev. Entomol.* **46**, 573 (2001).
3. B. Hölldobler, *J. Comp. Physiol.* **184**, 129 (1999).
4. E. Ruiz, M. H. Martínez, M. Martínez, J. M. Hernández, *Ann. Soc. Entomol. Fr.* **42**, 99 (2006).
5. J. A. Thomas, J. Settele, *Nature* **432**, 283 (2004).
6. J. A. Thomas, K. Schönrogge, G. W. Elmes, in *Insect Evolutionary Ecology*, M. D. Fellowes, G. Holloway, J. Rolff, Eds. (CABI, Wallingford, UK, 2005), pp. 479–518.
7. K. Schönrogge *et al.*, *J. Chem. Ecol.* **30**, 91 (2004).
8. T. Akino, J. J. Knapp, J. A. Thomas, G. W. Elmes, *Proc. R. Soc. London Ser. B* **266**, 1419 (1999).
9. J. A. Thomas, G. W. Elmes, J. C. Wardlaw, *Proc. R. Soc. London Ser. B* **265**, 1895 (1998).
10. G. W. Elmes, J. C. Wardlaw, K. Schönrogge, J. A. Thomas, *Entomol. Exp. Appl.* **110**, 53 (2004).
11. Materials and methods are available as supporting material on *Science* Online.
12. P. J. DeVries, R. B. Cocroft, J. A. Thomas, *Biol. J. Linn. Soc.* **49**, 229 (1993).
13. F. Roces, J. Tautz, *J. Acoust. Soc. Am.* **109**, 3080 (2001).
14. R. Hickling, R. L. Brown, *J. Acoust. Soc. Am.* **108**, 1920 (2000).
15. H. Markl, B. Hölldobler, *Behav. Ecol. Sociobiol.* **4**, 183 (1978).
16. H. Markl, *Z. Vgl. Physiol.* **60**, 103 (1968).
17. H. Markl, *Science* **149**, 1392 (1965).
18. D. A. Grasso, A. Mori, F. Le Moli, M. Giovannotti, A. Fanfani, *Ital. J. Zool.* **65**, 167 (1998).
19. T. C. Scott-Phillips, *J. Evol. Biol.* **21**, 387 (2008).
20. K. G. Schurian, K. Fiedler, *Nachr. Entomol. Vereins Apollo* **14**, 339 (1994).
21. C. J. Hill, *J. Aust. Entomol. Soc.* **32**, 283 (1993).
22. D. R. Nash, T. D. Als, R. Maile, G. R. Jones, J. J. Boomsma, *Science* **319**, 88 (2008).
23. P. J. DeVries, *Am. Mus. Nov.* **3025**, 1 (1991).
24. K. Fiedler, B. Hölldobler, P. Seufert, *Experientia* **52**, 14 (1996).
25. M. A. Travassos, N. E. Pierce, *Anim. Behav.* **60**, 13 (2000).
26. N. E. Pierce *et al.*, *Annu. Rev. Entomol.* **47**, 733 (2002).
27. J. C. Downey, A. C. Allyn, *Bull. Mus. Entomol.* **14**, 1 (1973).

28. We thank N. Elfferich and P. J. DeVries for introducing us to ant-butterfly acoustics; G. W. Elmes, J. C. Wardlaw, V. La Morgia, M. B. Bonsall, and referees for comments and advice; and M. Charles for designing the acoustical equipment.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5915/782/DC1
Materials and Methods
SOM Text
Figs. S1 and S2
Table S1
References
Audio S1 to S4

22 July 2008; accepted 28 November 2008
10.1126/science.1163583

Stability Predicts Genetic Diversity in the Brazilian Atlantic Forest Hotspot

Ana Carolina Carnaval,[1]* Michael J. Hickerson,[2] Célio F. B. Haddad,[3] Miguel T. Rodrigues,[4] Craig Moritz[1]

Biodiversity hotspots, representing regions with high species endemism and conservation threat, have been mapped globally. Yet, biodiversity distribution data from within hotspots are too sparse for effective conservation in the face of rapid environmental change. Using frogs as indicators, ecological niche models under paleoclimates, and simultaneous Bayesian analyses of multispecies molecular data, we compare alternative hypotheses of assemblage-scale response to late Quaternary climate change. This reveals a hotspot within the Brazilian Atlantic forest hotspot. We show that the southern Atlantic forest was climatically unstable relative to the central region, which served as a large climatic refugium for neotropical species in the late Pleistocene. This sets new priorities for conservation in Brazil and establishes a validated approach to biodiversity prediction in other understudied, species-rich regions.

Late Quaternary climate fluctuations helped to shape present-day diversity in temperate and boreal systems (*1*), providing a general context for understanding current patterns of endemism. In the tropics, Pleistocene refugia models have been dismissed because of conflicting evidence (*2*, *3*) or circularity in identifying putative refugia (*4*), but historical processes must be invoked to explain regions of high endemism (*5*, *6*). Recent studies from subtropical biomes have usefully employed post hoc palaeoclimate models of species and habitats to provide insights about processes shaping genetic and species diversity (*5*, *7*). Building on them, we first map the palaeodistribution of endemic species to identify temporally stable (refugial) and unstable (recently colonized) regions for species occurrence, which are then validated with multispecies molecular data. Going beyond the traditional species-by-species approach, the molecular analyses contrast the fit of assemblage-level data to the spatially explicit demographic scenarios suggested by the climate-based models.

We apply this approach to one of the world's most species-rich, yet notoriously endangered and understudied ecosystems: the Brazilian Atlantic

[1]Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720–3160, USA. [2]Biology Department, Queens College, City University of New York, Flushing, NY 11367, USA. [3]Departamento de Zoologia, Instituto de Biociências, UNESP, Rio Claro, SP 3526-4100, Brazil. [4]Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, SP 055008-090, Brazil.

*To whom correspondence should be addressed. E-mail: carnaval@berkeley.edu

rainforest. Originally extending for 1,300,000 km^2 along the Brazilian coast and reaching into Paraguay and Argentina, this biome has been reduced to less than 8% of its range (*8*). Today's fragments harbor one of the largest percentages of endemic species in the world, with many species and even genera of vertebrates still being described (*8*, *9*). Our ultimate goal is to pinpoint regions for inventory work and habitat protection before we lose a substantial fraction of described and undocumented diversity. The approach differs from previous methods by directly modeling historical processes, as opposed to observed biodiversity patterns (*10*), with the aim of informing conservation.

We use molecular genetic data from multiple, largely codistributed species to test whether spatial modeling of species-specific Late Quaternary refugia sheds light on historical processes and hence improves prediction of genetic endemism and diversity in tropical Brazil (*11*). We focus on three common species of tree frogs that are widely distributed along the Brazilian Atlantic forest: *Hypsiboas albomarginatus*, *H. semilineatus*, and *H. faber*. Given their life history traits, amphibians are useful indicators of environmental changes through time (*12*). Whereas *H. albomarginatus* and *H. semilineatus* occur in low and mid altitudes and are mostly restricted to the evergreen or semideciduous components of the Atlantic Forest in eastern Brazil, *H. faber* has a broader altitudinal range and also inhabits mixed and deciduous areas, occupying interior and coastal sites in the Atlantic Forest south to Paraguay and Argentina (figs. S1 and S2) (*13*). The comparative phylogeographic approach is a powerful test of assemblage-scale responses to former environmental change and thereby provides a means for critical assessment of the scenarios produced by modeling of species' distributions under palaeoclimates (*7*).

The palaeomodeling method intersects predicted species' distributions under current conditions and climatic extremes of the Late Quaternary (6000 years before present, or 6 kybp, and 21 kybp) to predict areas of stability (regions in which species are predicted to occupy irrespective of time period) and unstable areas (*7*, *14*). Because the stability maps raise specific hypotheses about regional differences in persistence and hence diversity, they lead to phylogeographic predictions for both individual species and assemblages (codistributed taxa; Fig. 1). Field sampling is driven by the model predictions to cover both predicted refugia and unstable (recently colonized) areas, particularly emphasizing previously undersampled areas. If the approach correctly predicts current patterns of biodiversity at the regional scale, species should consistently show (i) higher genetic diversity within and among populations in refugia relative to unstable areas, because of long-term persistence and population structure; (ii) genetic signature of population expansion in unstable areas, reflecting multispecies colonization from adjacent refugial regions after the Last Glacial Maximum (LGM, 21 kybp); (iii) absence of genetic patterns of isolation-by-distance in unstable areas, given that colonization has been too recent to permit restoration of equilibrium between migration and genetic drift (*15*); and (iv) strong phylogeographic structure between refugia, reflecting assemblage-wide, long-term population persistence in isolated areas.

Distribution models developed under current climatic conditions accurately predict distributions of each of the target species along the Atlantic rainforest domain [area-under-the-curve (AUC) values (*16*) 0.968, 0.989, and 0.994; maximum Kappa (*17*) 0.81, 0.925, and 0.94 in *H. albomarginatus*, *H. faber*, and *H. semilineatus*, respectively (fig. S2)]. Stability maps, depicting the intersection of distribution models for each taxon under current, 6 kybp, and 21 kybp climates, predict for all species a large central refugium throughout the Late Quaternary ("Bahia refugium") (Fig. 2). A second, much smaller refugium is predicted in the northeasternmost portion of the forest ("Pernambuco refugium"). In *H. faber*, a third, southeastern refugium of intermediate size is also predicted ("São Paulo refugium"). This is not surprising, given that this species occupies a broader environmental niche. In contrast to the central and northern regions, populations south of the Bahia or São Paulo refugia appear much less stable, despite the more extensive (preclearing) range of the forest in southern and southeastern Brazil. We hypothesize that these areas received a significant influx of migrants from adjacent, large refugial populations after the LGM. These palaeomodel results are congruent with the fossil pollen record, which documents a replacement of forests by grasslands in the southern Atlantic forest during the LGM (*14*, *18*) and suggests the occurrence of small forest refugia in the southernmost range of the putative Bahia refugium (*19*). The results also agree generally with forest models published previously (*14*), although the central refugium extends farther south in the frog-based models. Such differences are expected because the forest and its associated species may differ slightly in their climatic tolerances and realized niches. In *H. albomarginatus* and *H. faber*, the extension of the predicted São Paulo refugium westward into the neighboring Cerrado biome reflects model overprediction (fig. S2) (*14*).

Models of habitat stability through fluctuating climates correctly predict patterns of phylogeography in the Brazilian Atlantic rainforest (Fig. 2 and figs. S3 to S5). In all species, high levels of divergence and population structure are observed across refugia (Tamura-Nei corrected distances (*20*): 4 to 7% between Bahia and Pernambuco refugia, 1% between the nearby Bahia and São Paulo refugia in *H. faber*). Similarly, in all taxa there are multiple, divergent clades within the Bahia region, agreeing with model-based predictions of a large refugium in this area. In *H. faber*, divergent clades are also represented in the São Paulo region, matching predictions of a mid-sized refugium in this area. All taxa show low genetic diversity across the southernmost range of the forest, an area predicted to be less stable by the palaeomodels. Furthermore, mitochondrial DNA (mtDNA) lineages found in this region are shared with adjacent refugia (one in *H. albomarginatus* and *H. semilineatus*, two in *H. faber*).

Metrics of genetic diversity confirm the above patterns (Table 1). In *H. albomarginatus* and *H. semilineatus*, genetic diversity (*21*) is an order of magnitude larger in the central (Bahia) refugium relative to the less stable (southern) portion of the forest. Diversity of *H. faber* in this southern area is higher than the other species because of the

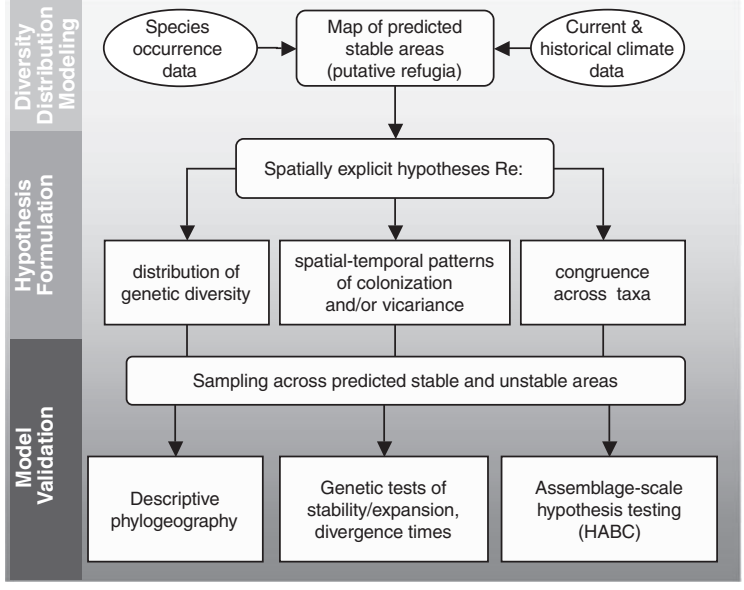


Fig. 1. Proposed method of biodiversity prediction. Three stages are involved: biodiversity distribution modeling (top), model-based hypothesis formulation (middle), hypothesis testing and model validation (bottom).

presence of two lineages that co-occur in the adjacent refugia. In all species, average net nucleotide differences across localities (*22*) reflects high geographic structure within refugia (2.6 to 6.2% divergence). In contrast, sites located outside (south of) the refugia are genetically more similar to each other, although to a lesser extent in *H. faber* (0.1 to 1.6%). Signatures of population expansion (*23*) are found in the unstable area for *H. albomarginatus* and *H. faber*, as well as in the Bahia refugium area for *H. faber* and *H. semilineatus*. The lack of signature of

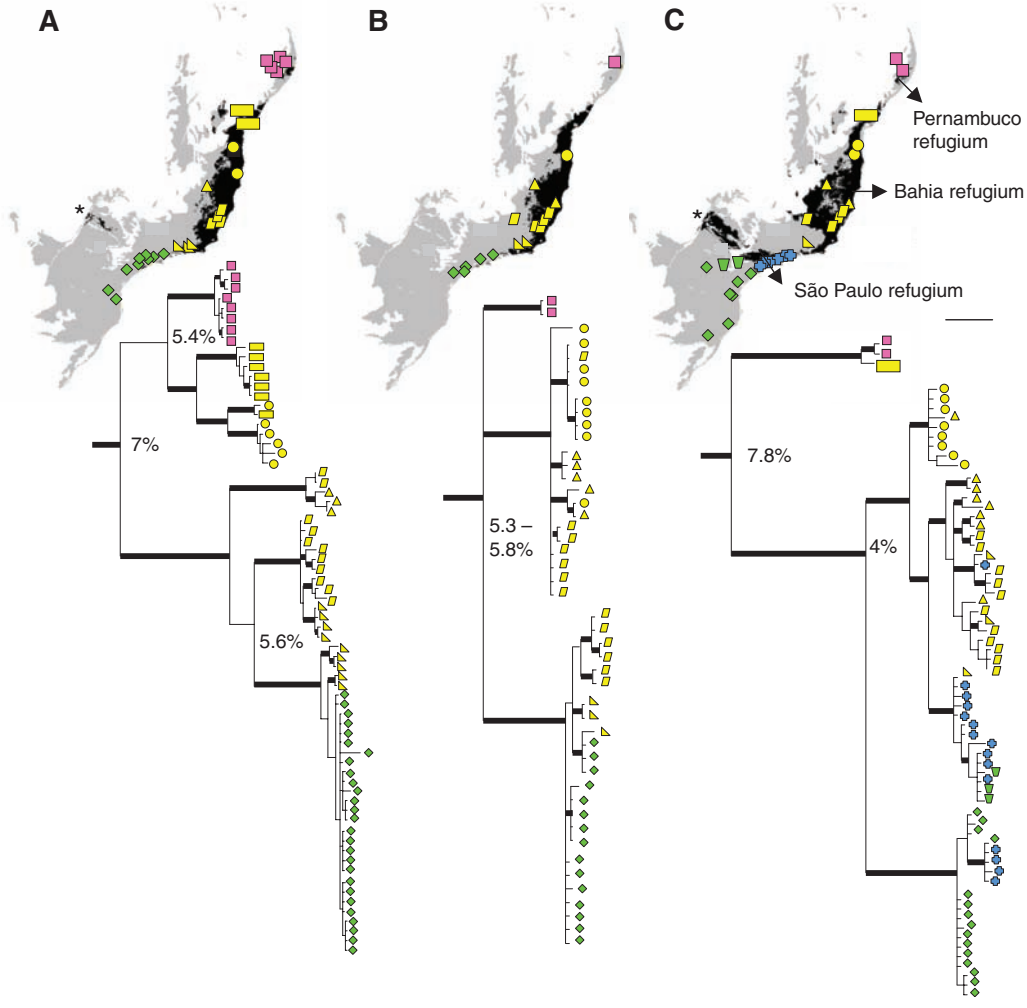


Fig. 2. Genetic diversity in putative refugial (stable) versus unstable areas in the Brazilian Atlantic rainforest. (**Top**) Species-specific stability maps; modeled refugia in black. (**A**) *H. albomarginatus*, (**B**) *H. semilineatus*, (**C**) *H. faber*. Note the absence of large stable regions in the southern portion of the forest (south of the Bahia and São Paulo refugia) relative to the central and northern areas. Asterisks denote refugia inferred beyond the current ranges of the target species. Symbols indicate localities sampled for molecular analysis. Scale bar, 400 km. (Bottom) The 50% majority-rule consensus Bayesian phylogenetic trees, rooted with sequences from the other two congeneric species studied (root not shown). Thick internodes denote clades with posterior probability greater than 90%. Percentages indicate Tamura-Nei corrected distances between clades (*20*).

Table 1. Population genetic summary metrics used in model validation. *n*, Sample size; *S*, number of segregating sites. The diversity parameter θ and mean D_a across localities are given per base pair (bp). *Hs* test (*23*) is used to detect population expansion. BA, Bahia; SP, São Paulo refugia. Because predicted refugia were often larger than predicted unstable (recently colonized) areas, *n*, *S*, θ, and average D_a values of the former were obtained not only from the total number of samples, but also from all possible combinations of spatially contiguous localities distributed within the geographic extension of the unstable area. Parentheses encompass minimum and maximum values from subsamples. *P* values in bold highlight statistical significance at 0.05 probability level.

Species	Area	*n* (min.; max.)	*S* (min.; max.)	θ (min.; max.)	Mean D_a (min.; max.)	*Hs* (*P* value)	Mantel's corr. coef. (*P* value)
H. albomarginatus (970 bp)	Stable (BA)	36 (13; 23)	207 (81; 155)	0.076 (0.034; 0.072)	0.062 (0.020; 0.082)	−20.546 (0.141)	0.499 **(0.001)**
	Unstable (south of BA)	27	22	0.003	0.001	−11.498 **(0.004)**	−0.140 (0.580)
H. semilineatus (718 bp)	Stable (BA)	28 (6; 13)	71 (14; 58)	0.031 (0.009; 0.034)	0.036 (0.007; 0.041)	−17.778 **(0.029)**	0.054 (0.460)
	Unstable (south of BA)	15	9	0.003	0.004	0.114 (0.357)	0.436 (0.248)
H. faber (771 bp)	Stable (BA)	28 (13; 23)	94 (42; 80)	0.018 (0.012; 0.022)	0.026 (0.001; 0.044)	−38.111 **(0.003)**	0.803 **(0.0003)**
	Stable (SP)	15	48	0.023	0.028	−5.981 (0.115)	0.305 (0.221)
	Unstable (south of SP)	18	40	0.015	0.016	−13.255 **(0.014)**	0.0001 (0.456)

population expansion in the southernmost localities of *H. semilineatus* may reflect low statistical power because of the exceptionally low levels of diversity observed in this species. As predicted, isolation by distance is not observed in unstable regions, but is detected within refugial areas for *H. albomarginatus* and *H. faber*.

The hierarchical approximate Bayesian computation (HABC) method (*24*) allows us to use data from all three species at once to test for assemblage-wide responses to Late Quaternary climate change. These analyses support both model-driven hypotheses of (i) simultaneous, multispecies colonization of unstable areas from adjacent refugial populations since the LGM, as opposed to long-term persistence of populations in unstable areas, and (ii) assemblage-scale, long-term persistence of populations in isolated refugial areas, as opposed to post-LGM colonization of refugial regions.

To test for assemblage-wide colonization of predicted unstable areas, we group mtDNA sequences from the southernmost refugial sites [population 1 (Fig. 3A)] and from localities in unstable areas south of the refugium [population 2 (Fig. 3A)] to contrast two alternative historical models across the three codistributed species, while allowing the taxon-specific demographic parameters to vary. In H_1, the long-term persistence model, two contemporary populations split from an ancestral population prior to the LGM (120,000 to 1.2 million years before present, or Mybp, Fig. 3A). In H_2, the recent colonization model, population 2 is modeled as being colonized from refugial population 1 subsequent to the LGM (0 to 20 kybp; Fig. 3A). The results indicate that all three species colonized the southern (unstable) areas after the LGM ($Z_2 = 3$, the number of species evolved under H_2), even when allowing for postisolation migration (Fig. 3, B and C). When Bayes factor is used (*25*), there is strong support for recent colonization in all three species ($Z_2 = 3$) under the no-migration model [$B(Z_2 = 3, Z_2 < 3) = 35.16$], and moderate support under a postisolation migration model [$B(Z_2 = 3, Z_2 < 3) = 5.70$].

Using the same framework to test for long-term persistence of refugial populations, we compare mtDNA sequences between the predicted Pernambuco refugium [population 1 (Fig. 3A)] and adjacent (northern) populations from the Bahia refugium [population 2 (Fig. 3A)] to contrast alternative historical models H_1 and H_2. In this case, the HABC results infer long-term persistence of populations in isolated refugia for all three species ($Z_2 = 0$, i.e., $Z_1 = 3$), even when allowing for postisolation migration (Fig. 3, D and E). Using Bayes factor (*25*), we also detect evidence for stability in both areas under the no-migration model [$B(Z_2 = 0, Z_2 > 0) = 4.89$], as well as under a postisolation migration model [$B(Z_2 = 0, Z_2 > 0) = 4.84$].

Relative to nuclear loci, mtDNA data are more variable and readily collected and often provide key insights into biological response to environmental modification (*1*). Although single-locus inference can be imprecise in the face of coalescent variance and the possibility of selection (*26*), our method benefits from a multitaxon approach, while explicitly accounting for the stochasticity of a single-locus coalescent across taxa. Combining data sets from several codistributed groups into a single hierarchical Bayesian analysis allowed us to estimate congruence across species, while borrowing strength from the full comparative phylogeographic sample (*24*). This can translate into higher analytical power and be more informative than qualitative comparisons of species-specific analyses. By capturing the historical signal that emerges from larger, combined multispecies molecular data sets, HABC will offer the possibility of looking at patterns of historical community assembly in codistributed nonmodel organisms for which barcode-type DNA sequence information (e.g., mtDNA data) can be feasibly collected.

Collectively, the results identify the central region as a hotspot within the Atlantic rainforest hotspot and a refuge for biodiversity during climatic extremes of the Late Pleistocene. This is not to say that southern areas entirely lacked forested habitats in the late Pleistocene: The existence of species and genera endemic to the southern forests (*27*), as well as some palaeoecological and genetic evidence (*28*), offer evidence to the contrary. Rather, the phylogeographically validated palaeomodels presented here show that the central region had much higher stability relative to the south. Forest lizards (*14*, *29*) and birds (*30*) also show high diversity in the central portion of the biome relative to southern areas, and provide evidence for population expansion in southern regions. This reassures us that the processes uncovered by the amphibian data may be generalized to and help to explain patterns of

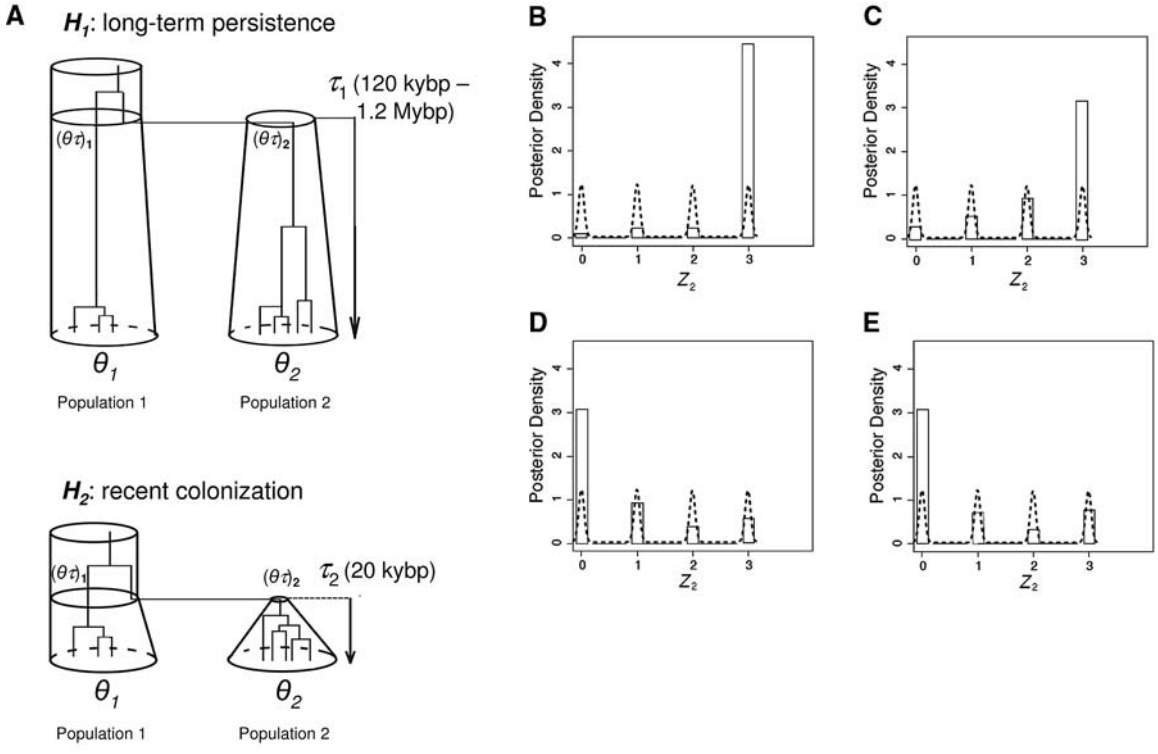


Fig. 3. HABC analyses. (**A**) Simulated models H_1 (long-term persistence) and H_2 (recent colonization). In both cases, each species was modeled as two contemporary populations with mutation-drift parameters θ_1 and θ_2 that split from an ancestral population at a time τ in the past. Ancestral population sizes are represented by $(\theta_\tau)_1$ and $(\theta_\tau)_2$; ybp, years before present. (**B** to **E**) Hyperposterior (bars) and hyperprior (dashed) densities of Z_2 (number of species evolved under H_2) given data from three codistributed frog species. (B) and (C) Models of refugial sites (population 1) and unstable, southern areas (population 2). (D) and (E) Models of Pernambuco refugium (population 1) and Bahia refugium (population 2). (B) and (D) Postisolation migration not included in model; (C) and (E) postisolation migration included in model.

diversity in other, much more distantly related groups of Atlantic forest endemics.

Because collection efforts, molecular studies, and conservation priorities have been heavily biased toward southern and southeastern Brazil (*8*, *9*, *31*), we predict that genetic diversity and narrow endemism in the central corridor of the biome have been substantially underestimated. This is serious, given the higher rate of deforestation in this region relative to the more extensive forests in São Paulo and southern Brazil (*9*, *31*). Not only could much unique diversity be lost, but ongoing habitat destruction could quickly erase the signature of the historical processes that led to it, preventing a full understanding of the mechanisms underlying local endemism and, therefore, impeding more effective conservation measures.

At a broader level, the congruence between model-based demographic hypotheses and joint, multispecies analyses of mtDNA diversity shows that palaeoclimatic niche models and assemblage-scale molecular genetic analyses can be used to forecast spatial patterns of diversity in poorly explored, highly threatened ecosystems. In a world of ever-accelerating environmental changes, this approach can help to guide research and conservation in other global hotspots or similarly complex tropical ecosystems.

References and Notes

1. G. Hewitt, *Nature* **405**, 907 (2000).
2. F. E. Mayle, D. J. Beerling, W. D. Gosling, M. B. Bush, *Philos. Trans. R. Soc. London B Biol. Sci.* **359**, 499 (2004).
3. E. P. Lessa, J. A. Cook, J. L. Patton, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 10331 (2003).
4. B. W. Nelson, C. A. C. Ferreira, M. F. da Silva, M. L. Kawasaki, *Nature* **345**, 714 (1990).
5. C. H. Graham, C. Moritz, S. E. Williams, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 632 (2006).
6. W. Jetz, C. Rahbek, R. K. Colwell, *Ecol. Lett.* **7**, 1180 (2004).
7. A. Hugall, C. Moritz, A. Moussalli, J. Stanisic, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6112 (2002).
8. L. P. C. Morellato, C. F. B. Haddad, *Biotropica* **32**, 786 (2000).
9. M. T. Rodrigues, *Conserv. Biol.* **19**, 659 (2005).
10. C. Kremen *et al.*, *Science* **320**, 222 (2008).
11. Materials and methods are available as supporting material on *Science* Online.
12. M. B. Araújo *et al.*, *Ecography* **31**, 8 (2008).
13. D. R. Frost, *Amphibian Species of the World: An Online Reference*, version 5.2, 15 July 2008 (American Museum of Natural History, New York, 2008); electronic database accessible at http://research.amnh.org/herpetology/amphibia/index.php.
14. A. C. Carnaval, C. Moritz, *J. Biogeogr.* **35**, 1187 (2008).
15. M. Slatkin, *Evolution* **47**, 264 (1993).
16. J. A. Hanley, B. J. McNeil, *Radiology* **143**, 29 (1982).
17. J. Cohen, *Ed. Psych. Meas.* **20**, 37 (1960).
18. H. Behling, R. R. B. Negrelle, *Quat. Res.* **56**, 383 (2001).
19. H. Behling, H. W. Arz, J. Pätzold, G. Wefer, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **179**, 227 (2002).
20. K. Tamura, M. Nei, *Mol. Biol. Evol.* **9**, 678 (1993).
21. F. Tajima, *Genetics* **143**, 1457 (1996).
22. M. Nei, *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987).
23. J. C. Fay, C. I. Wu, *Genetics* **155**, 1405 (2000).
24. M. J. Hickerson, C. P. Meyer, *BMC Evol. Biol.* **8**, 322 (2008).
25. R. E. Kass, A. E. Raftery, *J. Am. Stat. Assoc.* **90**, 773 (1995).
26. J. W. O. Ballard, M. Kreitman, *Genetics* **138**, 757 (1994).
27. C. F. B. Haddad, L. F. Toledo, C. P. A. Prado, *Anfibios da Mata Atlântica* (*Atlantic Forest Amphibians*) (Editora Neotropica, São Paulo, Brazil, 2008).
28. M.-P. Ledru *et al.*, *Divers. Distrib.* **13**, 761 (2007).
29. K. C. M. Pellegrino *et al.*, *Biol. J. Linn. Soc. London* **85**, 13 (2005).
30. G. S. Cabanne, F. M. d'Horta, E. H. R. Sari, F. R. Santos, C. Y. Miyaki, *Mol. Phylogenet. Evol.* **49**, 760 (2008).
31. J. M. C. da Silva, M. Tabarelli, *Nature* **404**, 72 (2000).

32. We thank U. Caramaschi and H. Zaher for providing access to collections MNRJ and MZUSP; O. Peixoto, M. Gomes, A. Muri, R. Kautskyi, S. Lima, E. Santos, J. S. Filho, J. V. Filho, G. Barros, J. Queiroz, R. Araújo, L. Japp, H. Japp, J. Giovanelli, J. Alexandrino, L. Toledo, O. Araújo, G. Egito, J. Zina, D. Loebmann, D. Pavan, R. Amaro, V. Verdade, F. Curcio, M. Dixo, and J. Cassimiro for field work assistance; W. Monahan and R. Hijmans for discussions about the modeling work; L. Smith and D. Turong for DNA-sequencing assistance; R. Pereira, R. Damasceno, S. Rovito, J. Kolbe, S. Singhal, R. Puschendorf, and A. Pounds for discussions about earlier versions of the manuscript. Funding was provided by the NSF (awards DBI 0512013 to A.C.C., DEB 0743648 to M.J.H., DEB 416250 and DEB 0817035 to C.M.), Fundação de Amparo à Pesquisa do Estado de São Paulo and Conselho Nacional de Desenvolvimento Científico e Tecnológico (grants to C.F.B.H. and M.T.R.). Sequences are deposited in GenBank (FJ502639-FJ502822). A.C.C. and C.M. designed the study; A.C.C., C.F.B.H., and M.T.R. collected the data; A.C.C., C.M. and M.J.H. analyzed the data; A.C.C. wrote the paper. All authors discussed the results and commented on earlier versions of the manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5915/785/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 and S2
References

8 October 2008; accepted 9 December 2008
10.1126/science.1166955

Orc1 Controls Centriole and Centrosome Copy Number in Human Cells

Adriana S. Hemerly,[1,2] Supriya G. Prasanth,[1]* Khalid Siddiqui,[1]† Bruce Stillman[1]‡

Centrosomes, each containing a pair of centrioles, organize microtubules in animal cells, particularly during mitosis. DNA and centrosomes are normally duplicated once before cell division to maintain optimal genome integrity. We report a new role for the Orc1 protein, a subunit of the origin recognition complex (ORC) that is a key component of the DNA replication licensing machinery, in controlling centriole and centrosome copy number in human cells, independent of its role in DNA replication. Cyclin A promotes Orc1 localization to centrosomes where Orc1 prevents Cyclin E-dependent reduplication of both centrioles and centrosomes in a single cell division cycle. The data suggest that Orc1 is a regulator of centriole and centrosome reduplication as well as the initiation of DNA replication.

The assembly of a bipolar, microtubule spindle during mitosis is essential for accurate chromosome segregation. In animal cells, spindle formation is organized by centrosomes, organelles that contain a pair of centrioles surrounded by pericentriolar material (PCM) that need to be duplicated exactly once every cell division cycle, in coordination with DNA replication to maintain genome stability (*1*). Licensing DNA for replication is a critical regulatory step involving the origin recognition complex (ORC), the first component for assembly of a pre–replicative complex (pre-RC) at each origin (*2*). Accumulated evidence supports roles for ORC subunits in addition to licensing DNA replication (*3*). In particular, human Orc2 subunit localizes to centrosomes, and depletion of Orc2 and Orc3 causes centrosome amplification in mitosis (*4*).

Several regulators of the DNA licensing machinery have been reported to be involved in the control of both DNA and centriole duplication (*5*). Both Cyclin E and Cyclin A, as well as Cdk2 activity, are well-known positive regulators of DNA replication and also promote centrosome duplication (or reduplication) (*6*–*11*). Depletion of the DNA replication licensing inhibitor Geminin causes reduplication of both DNA and centrosomes in human cells (*12*).

In a screen using small interfering RNA (siRNA) for human ORC proteins with roles in centrosome biology, we found that depletion of the largest human ORC subunit, HsOrc1, leads to excess centrosomes (fig. S1 and Fig. 1A). Orc1 siRNA-treated U2OS cells were analyzed for centrosome defects by dual-color indirect immunofluorescence (IF) using antibodies to centrin 2 (stains centrioles) and antibodies to γ-tubulin (stains centrosomes). Seventy-two hours after siRNA treatment, 39.77 ± 3.5% of cells transfected with Orc1-1 siRNA and 25.53 ± 0.3% of

[1]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor 11724, NY, USA. [2]Instituto de Bioquímica Médica, UFRJ, 21941-590, Rio de Janeiro, Brazil.

*Present address: Department of Cell and Developmental Biology, 601 South Goodwin Avenue, University of Illinois at Urbana–Champaign, Champaign, IL 61801, USA.
†Present address: Clare Hall Laboratories, Cancer Research UK, South Mimms, Herts, EN6 3LD, UK.
‡To whom correspondence should be addressed. E-mail: stillman@cshl.edu

cells transfected with Orc1-2 siRNA showed multiple centrosomes and centrioles, in comparison with 2.4 ± 1.2% of control cells (fig. S1C). In Orc1-1 treated cells, 33.29 ± 2.6% of the multiple centrosomes were separated rather than linked, and the reduplication occurred in different human cell lines (figs. S1 and S2).

To study the kinetics of centriole and centrosome reduplication after Orc1 depletion, U2OS cells were synchronized in mitosis by nocodazole treatment and released into the next cycle. Centriole and centrosome numbers were assessed 12 hours, 36 hours, and 60 hours after nocodazole release (Fig. 1B and fig. S3). Disorganized disengagement of duplicated centrioles (2Pairs Dis.) was observed 12 hours after release when most of the cells were in G1 or early S phase (Fig. 1B and fig S3, centrin 2). This was followed by an increase in the percentage of cells showing centriole reduplication (>2Pairs). The complete centrosome structure, as evidenced by γ-tubulin staining, was observed later, about 60 hours after release (γ-tubulin). The very early effect of Orc1 depletion on centriole disengagement and reduplication suggests that Orc1 might have a more direct effect on these processes, whereas complete centrosome formation follows centriole reduplication. Similar dynamics of centriole reduplication after Orc1 depletion were observed in HeLa cells (fig. S4). It has been proposed that centriole engagement blocks reduplication (*13*). In Orc1-depleted cells, however, the duplicated centrioles were found disengaged and underwent reduplication, suggesting that Orc1 may block reduplication by participating in processes of centriole engagement of the newly duplicated centrioles.

Different mechanisms for centrosome amplification have already been reported. Apparent multiple centrosomes can arise by centrosome fragmentation, centriole splitting that occurs during early mitosis after DNA damage (*14*), or authentic reduplication. Analysis of multiple parameters—including DNA content (fig. S1B), cell cycle markers (fig. S5), centriolin staining of mature centrosomes (figs. S6 and S7), and the numbers of centrioles per PCM (Fig. 1A)—all indicate that Orc1 depletion causes bona fide centriole and centrosome reduplication in U2OS cells (see supporting material).

To address whether overexpression of Orc1 would affect centrosome reduplication caused by prolonged S-phase arrest in the presence of hydroxyurea (HU) (*9*, *15*), U2OS cells were transiently transfected with YFP-Orc1^{WT}, YFP-Orc2, or a vector carrying only yellow fluorescent protein (YFP) in the presence of HU (Fig. 2, A and D, and fig. S8). Centrosome reduplication after HU treatment, but not normal duplication, was substantially inhibited in YFP-Orc1^{WT} cells, whereas YFP or YFP-Orc2 did not have any effect, suggesting a role for Orc1 as an inhibitor of centrosome reduplication.

YFP-Orc1 associated with centrosomes in addition to its nuclear localization in ~35% of transfected cells (fig. S9A and movie S1). Furthermore, endogenous Orc1 and other ORC subunits cofractionated with purified centrosomes (fig. S9B). Transiently expressed Orc1 fused to the pericentrin-AKAP450 centrosomal targeting (PACT) domain, a conserved motif that targets proteins to centrosomes (*16*), resulted in inhibition of centrosome reduplication in HU-treated U2OS cells (Fig. 2, B and D), suggesting that Orc1 controls centrosome copy number when located at centrosomes.

Centrosome reduplication was inhibited by the full-length and N-terminal region of Orc1 (amino acids 1 to 500; Orc1^{WT}.Nterm.Flag) (Fig. 2C), the latter of which is deficient in ORC assembly (*17*). Both Orc1 versions also inhibited centrosome reduplication when expressed after the application of the HU block, indicating that the inhibition was not caused by a G1 arrest in these cells (fig. S10). Thus, Orc1 control of centrosome copy number is separate from its DNA replication role and does not require assembly of ORC, but such assembly may facilitate a more efficient block.

Orc1 depletion in synchronized U2OS cells did not trigger early responses of DNA damage checkpoint pathways that are known to affect centriole and centrosome copy number (*18*, *19*). Activation of Chk1 by phosphorylation on serine 317 was not detected, and p27 and p53 pathways were activated well after centriole amplification was initiated (Fig. 1B and fig. S11).

Several reports have shown that Cyclin E overexpression in mammalian cells is linked to genome and centrosome abnormalities (*20*, *21*) and that centrosome reduplication requires Cdk2 activity (*7*–*9*, *11*). Elevated levels of Cyclin E were identified as early as 12 hours after release

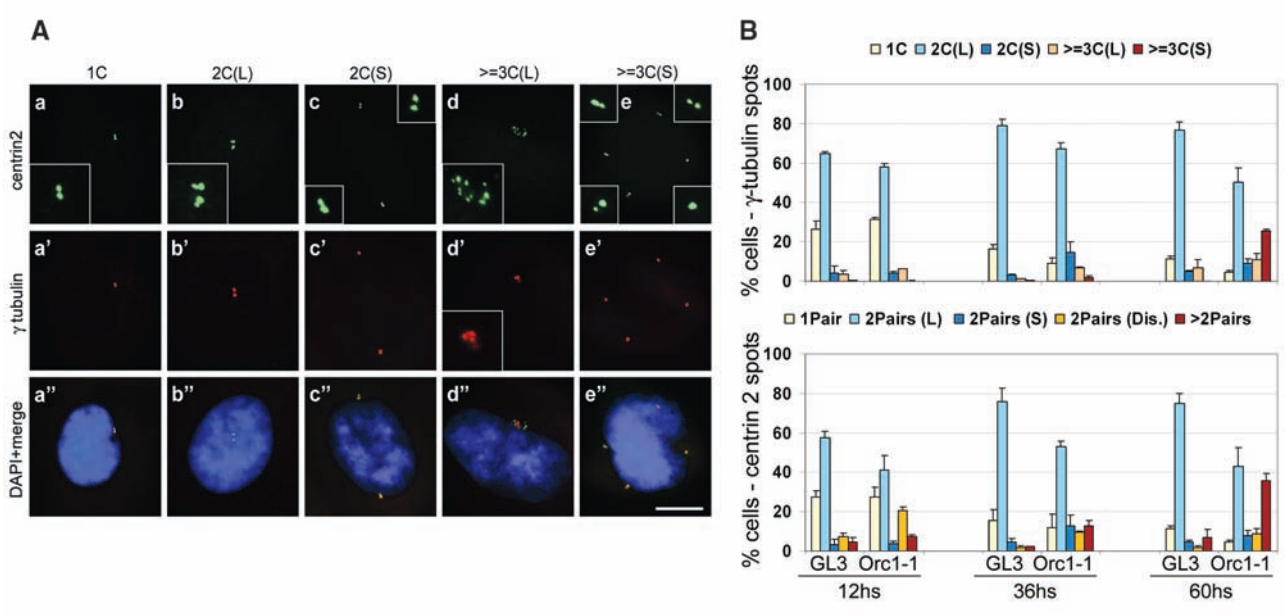


Fig. 1. Orc1 depletion causes centriole and centrosome reduplication. (**A**) Human U2OS cells treated for 72 hours with Orc1-1 siRNA duplex were coimmunostained for centrioles with antibodies to centrin 2 (green, a to e) and centrosomes with antibodies to γ-tubulin (red, a′ to e′). Insets are higher magnification. DNA was stained with 4′,6′-diamidino-2-phenylindole (DAPI) (blue), and merged images are shown in a″ to e″. Scale bar, 10 μm. (**B**) Quantification of centrosome and centriole numbers by γ-tubulin (top) and centrin 2 (bottom) immunostaining. Cells were harvested at 12, 36, or 60 hours after nocodazole release. Error bars, 1 SE. 1C, one centrosome; 2C(L), two centrosomes linked; 2C(S), two centrosomes separated; >=3C(L), three or more centrosomes linked; >=3C(S), three or more centrosomes separated. 1Pair, one centriole pair; 2Pairs (L), two centriole pairs linked; 2Pairs (S), two centriole pairs separated; 2Pairs (Dis.), two centriole pairs disorganized and disengaged; >=2Pairs, more than two centriole pairs.

of synchronized Orc1-depleted U2OS cells in Cyclin A–negative cells (Fig. 3, A and B, and fig. S5, A and B) and continued to increase. We thus hypothesized that Orc1 control of centriole and centrosome reduplication involved Cyclin E and Cdk2. The Cdk2 inhibitor roscovitine suppressed centriole and centrosome reduplication in Orc1-depleted cells (fig. S12). In a second experiment, simultaneous depletion of Orc1 and Cyclin E blocked centriole and centrosome reduplication, whereas normal duplicated centriole and centrosome numbers were observed (Fig. 3C and fig. S13A). Cyclin E siRNA did not change the number of cells with a 4C DNA content after Orc1-1 treatment (fig. S13B), which implies that the block of centrosome overduplication in Orc1-1–depleted cells by Cyclin E siRNA was not due to a compete G1 arrest. In contrast, Cyclin A siRNA did not abolish centrosome reduplication in Orc1-depleted cells (Fig. 3C).

We tested the ability of elevated levels of cyclins to cause centrosome reduplication in HU-treated cells expressing YFP.Orc1^{WT} (Fig. 4A). Coexpression of Cyclin E with YFP.Orc1^{WT} blocked YFP.Orc1^{WT} inhibition of HU-induced centrosome reduplication, whereas neither Cyclin A nor Cyclin B had any effect (Fig. 4, A and B). HU-treated cells coexpressing YFP.Orc1^{WT} with Cyclin A showed a cytoplasmic pool of over-expressed YFP.Orc1^{WT} in addition to the nuclear localization and an enhancement of YFP-Orc1 located at centrosomes (Fig. 4B, b to b″) (The number of YFP-Orc1–positive centrosomes in the presence of Cyclin A increased six-fold compared with YFP-Orc1 alone). Coexpression of Cyclin E with YFP.Orc1^{WT}.PACT also eliminated the ability of YFP.Orc1^{WT}.PACT to block centrosome reduplication, whereas Cyclin A or Cyclin B had no effect (Fig. 4, A and B). Altogether, the results suggest that Orc1 might directly regulate centriole and centrosome copy number in human cells and that the mechanism involves Cyclin E.

Orc1 coprecipitated with Cyclin E and Cyclin A, as previously shown (*22*), whereas no association was observed with Cyclin B (Fig. 4C). Association in a complex with Cyclin E was stronger than with Cyclin A; however, so far we have not been able to find evidence that Orc1 binds Cyclin E directly. Nevertheless, purified wild-type Orc1 inhibited Cyclin E-Cdk2 and Cyclin A-Cdk2 kinase activity using histone H1 as substrate (fig. S14B).

The Orc1 "RXL" mutant (K235A-L237A, Orc1^{A-A}) impairs the direct interaction with Cyclin A but does not affect Cyclin E binding (Fig. 4C and fig. S14A) and cannot block Cyclin A–associated kinase activity (fig. S14B). Although some inhibition was observed, Orc1^{A-A} and Orc1^{A-A}.Nterm did not block centrosome reduplication as efficiently as did wild-type Orc1, which suggests that Orc1-Cyclin A interaction is required for optimal centrosome copy number control (Fig. 4D). Orc1^{A-A}.PACT blocked centrosome reduplication to the same extent as Orc1^{WT}.PACT (Fig. 4E). Because Cyclin A function has been previously linked to the cytoplasmic localization of Orc1 (*23*), possibly Orc1^{A-A} and Orc1^{A-A}.Nterm mutants were not exported efficiently to centrosomes to block HU-induced reduplication, whereas Orc1^{A-A}.PACT was directly targeted to centrosomes. Accordingly, Cyclin A stimulation of Orc1 association with centrosomes was significantly reduced when coexpressed with the RXL Orc1^{A-A} mutant.

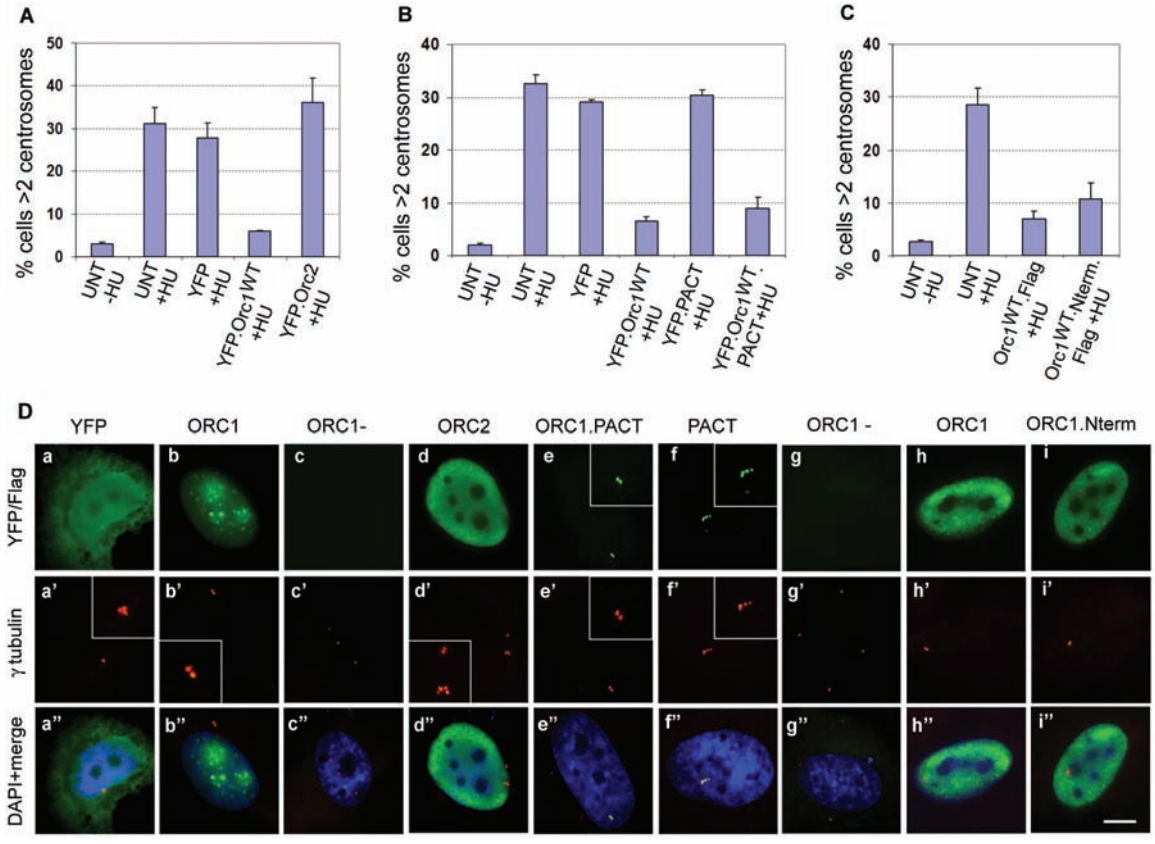


Fig. 2. Overexpression of Orc1 blocks centrosome reduplication. U2OS asynchronous cells were transfected with the indicated constructs and treated with 16 mM hydroxyurea (+HU). Untransfected cells were treated or not with HU (UNT +HU and UNT -HU, respectively). Cells were harvested at 68 hours after HU treatment. Centrosome numbers were scored by γ-tubulin immunostaining in YFP- or Flag-positive cells. (**A** to **C**) Quantification of multiple centrosomes in HU-arrested cells. Error bars, 1 SE. (**D**) Cells were immunostained for centrosomes with antibody to γ-tubulin (red, a′ to i′); YFP expression was immunostained with antibody to green fluorescent protein (GFP) in (A) (green, a to d) or visualized directly in (B) (e and f); and Flag expression was immunostained with antibody to Flag in (C) (g to i). DNA was stained with DAPI (blue), and merged images are shown in a″ to i″. Scale bar, 10 μm. Insets are higher magnification.

Altogether, the data indicate that both Cyclin E and Cyclin A play a role in Orc1 mediated control of centrosome numbers.

ORC in human cells is dynamically assembled and disassembled during the cell cycle and participates in many aspects of chromosome duplication and segregation, including functions at kinetochores, centromeres, and heterochromatin (*3*). Orc2 is localized to centrosomes, but its depletion affects centrosome copy number in mitosis in addition to controlling DNA replication, centromere activity, and chromosome structure (*4*). In contrast, Orc1 controls Cyclin E–dependent centrosome copy number by directly preventing reduplication of centrioles immediately after centrioles duplicate upon commitment to cell division. Without Orc1, duplicated centrioles are disengaged and extensive Cyclin E–dependent centriole

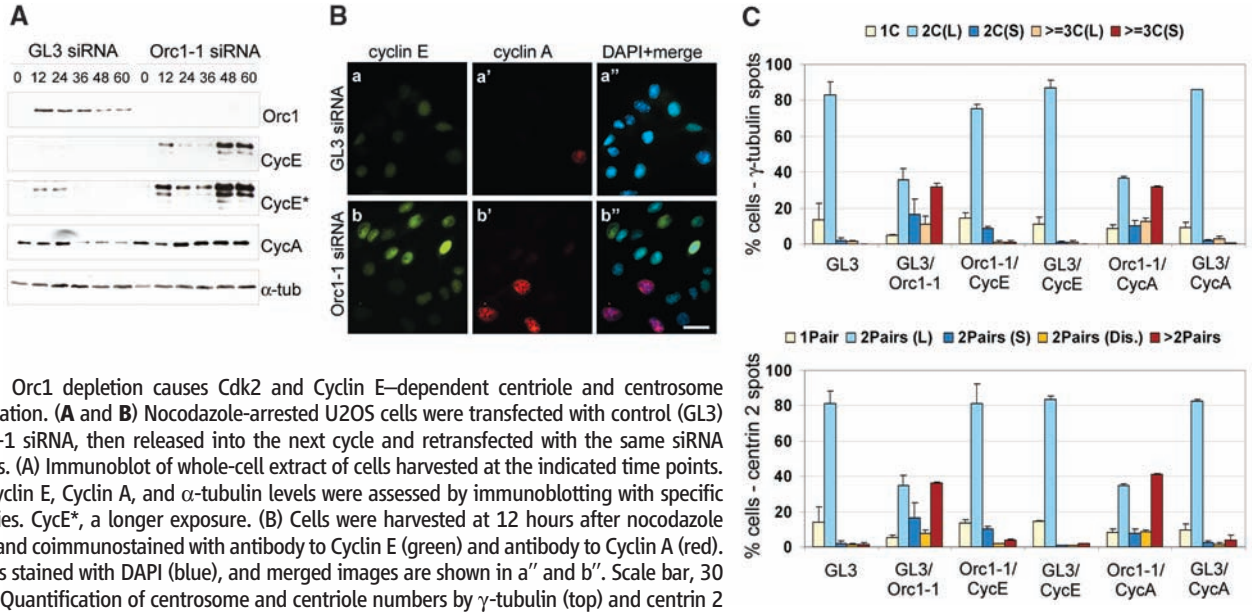


Fig. 3. Orc1 depletion causes Cdk2 and Cyclin E–dependent centriole and centrosome reduplication. (**A** and **B**) Nocodazole-arrested U2OS cells were transfected with control (GL3) or Orc1-1 siRNA, then released into the next cycle and retransfected with the same siRNA duplexes. (A) Immunoblot of whole-cell extract of cells harvested at the indicated time points. Orc1, Cyclin E, Cyclin A, and α-tubulin levels were assessed by immunoblotting with specific antibodies. CycE*, a longer exposure. (B) Cells were harvested at 12 hours after nocodazole release and coimmunostained with antibody to Cyclin E (green) and antibody to Cyclin A (red). DNA was stained with DAPI (blue), and merged images are shown in a″ and b″. Scale bar, 30 μm. (**C**) Quantification of centrosome and centriole numbers by γ-tubulin (top) and centrin 2 (bottom) immunostaining, respectively, of asynchronous U2OS cells transfected with the indicated siRNAs. Cells were analyzed 72 hours after the first transfection. Error bars, 1 SE.

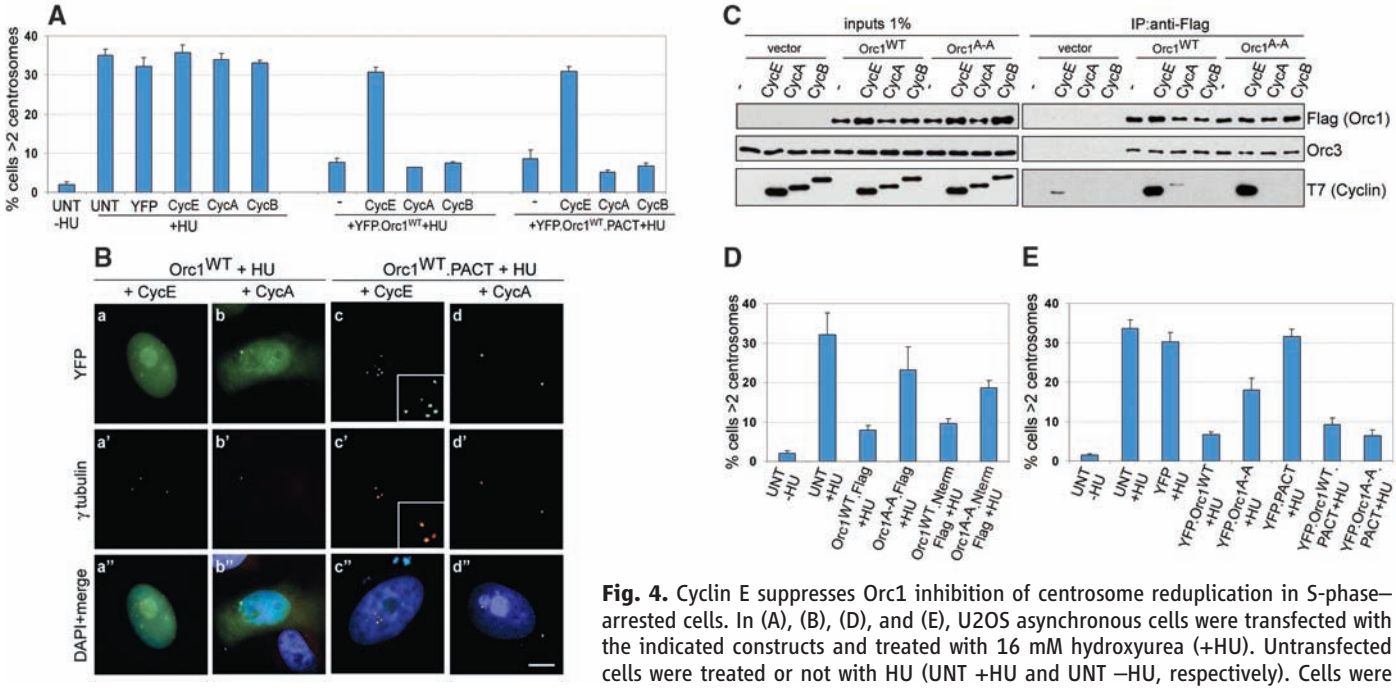


Fig. 4. Cyclin E suppresses Orc1 inhibition of centrosome reduplication in S-phase–arrested cells. In (A), (B), (D), and (E), U2OS asynchronous cells were transfected with the indicated constructs and treated with 16 mM hydroxyurea (+HU). Untransfected cells were treated or not with HU (UNT +HU and UNT –HU, respectively). Cells were harvested at 68 hours after transfection and HU treatment. (**A**) Quantification of multiple centrosomes in transfected HU-arrested cells. Centrosome numbers were scored by γ-tubulin immunostaining in YFP-positive cells. Error bars, 1 SE. (**B**) Transfected, HU-treated cells were immunostained for centrosomes with antibody to γ-tubulin (red, a′ to d′), and YFP expression was visualized directly (green, a to d). DNA was stained with DAPI (blue), and merged images are shown in a″ to d″. Scale bar, 10 μm. (**C**) Immunoprecipitation with antibody to Flag from whole-cell extract from HEK293 cells transiently coexpressing the indicated constructs and immunoblotting with the indicated antibodies. Vector, full-length Orc1^{WT}.Flag (wild type) or full-length Orc1^{A-A}.Flag (mutant) were transiently transfected alone (–) or with either T7-Cyclin E, T7-Cyclin A, or T7-Cyclin. Right panel: 10% of the immunoprecipitates. (**D** and **E**) Quantification of multiple centrosomes in transfected HU-arrested cells. Centrosome numbers were scored by γ-tubulin immunostaining in Flag-positive cells in (D) and by γ-tubulin immunostaining in YFP-positive cells directly visualized in (E). Error bars, 1 SE.

reduplication occurs. In the absence of Orc1 and Cyclin E, centrosome reduplication does not occur; however, centrioles duplicate, which suggests that such duplication is Cyclin A–dependent, consistent with previous studies (*24*). Similarly, Cyclin E, but not Cyclin A, overrides the Orc1 inhibition of centrosome reduplication.

Cyclin E is required for centrosome duplication and reduplication and coprecipitates with Orc1 (*7*, *11*, *25*–*27*). We have not been able to detect direct binding between Cyclin E and Orc1, which suggests that a mediator protein may facilitate the binding. The mediator may be another DNA replication protein implicated in centrosome biology (*12*, *28*). Furthermore, depletion of Orc1 increases Cyclin E protein levels in cells that may contribute to centriole reduplication. Both proteins are degraded by SKP1/cullin/F-box protein (SCF), which localizes to centrosomes and has also been implicated in the control of centrosome duplication (*29*, *30*).

There is a short time during the cell division cycle after centrioles have duplicated when Cyclin E is still present in cells (late G1 and early S phase). During this time, if Cyclin E activity is not checked, centriole reduplication might occur. We suggest that a mechanism involving Cyclin A–dependent localization of Orc1 to centrosomes prevents Cyclin E–dependent centriole reduplication during this time window (fig. S15). In this manner, Orc1 plays a dual role in coordinating DNA replication and centrosome copy number.

References and Notes

1. E. A. Nigg, *Trends Cell Biol.* **17**, 215 (2007).
2. J. J. Blow, A. Dutta, *Nat. Rev. Mol. Cell Biol.* **6**, 476 (2005).
3. T. Sasaki, D. M. Gilbert, *Curr. Opin. Cell Biol.* **19**, 337 (2007).
4. S. G. Prasanth, K. V. Prasanth, K. Siddiqui, D. L. Spector, B. Stillman, *EMBO J.* **23**, 2651 (2004).
5. M. Bettencourt-Dias, D. M. Glover, *Nat. Rev. Mol. Cell Biol.* **8**, 451 (2007).
6. D. Coverley, H. Laman, R. A. Laskey, *Nat. Cell Biol.* **4**, 523 (2002).
7. K. R. Lacey, P. K. Jackson, T. Stearns, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2817 (1999).
8. Y. Matsumoto, K. Hayashi, E. Nishida, *Curr. Biol.* **9**, 429 (1999).
9. P. Meraldi, J. Lukas, A. M. Fry, J. Bartek, E. A. Nigg, *Nat. Cell Biol.* **1**, 88 (1999).
10. L. De Boer *et al.*, *Oncogene* **27**, 4261 (2008).
11. E. H. Hinchcliffe, C. Li, E. A. Thompson, J. L. Maller, G. Sluder, *Science* **283**, 851 (1999).
12. K. E. K. Tachibana, M. A. Gonzalez, G. Guarguaglini, E. A. Nigg, R. A. Laskey, *EMBO Rep.* **6**, 1052 (2005).
13. M. F. B. Tsou, T. Stearns, *Curr. Opin. Cell Biol.* **18**, 74 (2006).
14. H. M. J. Hut *et al.*, *Mol. Biol. Cell* **14**, 1993 (2003).
15. C. Wong, T. Stearns, *Nat. Cell Biol.* **5**, 539 (2003).
16. A. K. Gillingham, S. Munro, *EMBO Rep.* **1**, 524 (2000).
17. K. Siddiqui, B. Stillman, *J. Biol. Chem.* **282**, 32370 (2007).
18. E. Bourke *et al.*, *EMBO Rep.* **8**, 603 (2007).
19. E. Sugihara *et al.*, *Cancer Res.* **66**, 4020 (2006).
20. C. H. Spruck, K. A. Won, S. I. Reed, *Nature* **401**, 297 (1999).
21. H. I. Saavedra *et al.*, *Cancer Cell* **3**, 333 (2003).
22. J. Mendez *et al.*, *Mol. Cell* **9**, 481 (2002).
23. H. Laman, G. Peters, N. Jones, *Exp. Cell Res.* **271**, 230 (2001).
24. K. Hanashiro, M. Kanai, Y. Geng, P. Sicinski, K. Fukasawa, *Oncogene* **27**, 5288 (2008).
25. M. F. B. Tsou, T. Stearns, *Nature* **442**, 947 (2006).
26. Y. Matsumoto, J. L. Maller, *Science* **306**, 885 (2004).
27. A. Duensing *et al.*, *Oncogene* **25**, 2943 (2006).
28. R. L. Ferguson, J. L. Maller, *J. Cell Sci.* **121**, 3224 (2008).
29. E. Freed *et al.*, *Genes Dev.* **13**, 2242 (1999).
30. K. Nakayama *et al.*, *EMBO J.* **19**, 2069 (2000).

31. We thank D. Spector for critical reading of the manuscript, P. Wendel for technical support, and J. Salisbury and S. Doxsey for providing antibodies to centrin 2 and centriolin, respectively. This work was supported by a grant from the National Institutes of Health (CA13106), a fellowship from CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) to A.S.H., and a Special Fellow award from the Leukemia and Lymphoma Society to S.G.P.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5915/789/DC1
Materials and Methods
Figs. S1 to S15
References
Movie S1

2 October 2008; accepted 9 December 2008
10.1126/science.1166745

Function of Mitochondrial Stat3 in Cellular Respiration

Joanna Wegrzyn,[1,2]* Ramesh Potla,[3]* Yong-Joon Chwae,[1] Naresh B. V. Sepuri,[4] Qifang Zhang,[1] Thomas Koeck,[5] Marta Derecka,[1,6] Karol Szczepanek,[1,6] Magdalena Szelag,[1,2] Agnieszka Gornicka,[1,7] Akira Moh,[8] Shadi Moghaddas,[9] Qun Chen,[9] Santha Bobbili,[1] Joanna Cichy,[6] Jozef Dulak,[2] Darren P. Baker,[10] Alan Wolfman,[11] Dennis Stuehr,[3,5] Medhat O. Hassan,[12] Xin-Yuan Fu,[8] Narayan Avadhani,[13] Jennifer I. Drake,[14] Paul Fawcett,[14] Edward J. Lesnefsky,[9,15] Andrew C. Larner[1]†

Cytokines such as interleukin-6 induce tyrosine and serine phosphorylation of Stat3 that results in activation of Stat3-responsive genes. We provide evidence that Stat3 is present in the mitochondria of cultured cells and primary tissues, including the liver and heart. In Stat3$^{-/-}$ cells, the activities of complexes I and II of the electron transport chain (ETC) were significantly decreased. We identified Stat3 mutants that selectively restored the protein's function as a transcription factor or its functions within the ETC. In mice that do not express Stat3 in the heart, there were also selective defects in the activities of complexes I and II of the ETC. These data indicate that Stat3 is required for optimal function of the ETC, which may allow it to orchestrate responses to cellular homeostasis.

Binding of cytokines to their cell surface receptors activates the Jak protein tyrosine kinases that phosphorylate conserved tyrosine residues on Stat proteins (*1*). Tyrosine-phosphorylated Stats translocate to the nucleus and bind to the promoters of early response genes. Several of the Stats are also phosphorylated on a conserved serine residue.

GRIM-19 was identified as a protein involved in interferon-β (IFNβ)– and retinoic acid–stimulated apoptosis of breast cancer cells (*2*). GRIM-19 inhibits Stat3 transcriptional activity, and these proteins directly interact (*3*, *4*). GRIM-19 was also purified as a component of complex I of the electron transport chain (ETC) (*5*). GRIM-19$^{-/-}$ mice are embryonically lethal, and in GRIM19$^{-/-}$ embryonic stem cells, assembly of complex I (of which there are 46 known components) is defective. The absence of GRIM-19 also influences the assembly and function of other complexes of the ETC (*6*). These observations indicate that Stat3 and GRIM-19 might colocalize in the mitochondria.

To address this possibility, we used SDS–polyacrylamide gel electrophoresis (SDS-PAGE) to separate isolated mitochondria and cytoplasm from the livers and hearts of mice, and we used Western blotting to detect Stat3 (Fig. 1A) (*7*). Stat3 was present in mitochondria and cytoplasmic fractions. The amount of Stat3 in the mitochondria (mStat3) was about one-tenth that in the cytosol. The blots were also probed for GRIM-19, tubulin as a cytoplasmic marker, calreticulin as

[1]Department of Biochemistry and Molecular Biology and Massey Cancer Center, Virginia Commonwealth University, Richmond, VA 23298, USA. [2]Department of Medical Biotechnology, Faculty of Biochemistry, Biophysics, and Biotechnology, Jagiellonian University, Krakow, Poland. [3]Department of Biology, Cleveland State University, Cleveland, OH 44114, USA. [4]Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad 500 046, India. [5]Department of Pathobiology, Lerner Research Institute, The Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195, USA. [6]Department of Immunology, Faculty of Biochemistry, Biophysics, and Biotechnology, Jagiellonian University, Krakow, Poland. [7]Department of Cell Biochemistry, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland. [8]Department of Microbiology and Immunology, Indiana University School of Medicine, Indianapolis, IN 46202, USA. [9]Division of Cardiology, Department of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA. [10]BiogenIdec, 14 Cambridge Center, Cambridge, MA 02142, USA. [11]Department of Cell Biology, Lerner Research Institute, The Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195, USA. [12]Pathology and Laboratory Medicine Service, Louis Stokes Cleveland Department of Veterans Affairs Medical Center, Cleveland, OH 44106, USA. [13]Department of Animal Biology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [14]Department of Internal Medicine, Virginia Commonwealth University, Richmond, VA 23298, USA. [15]Medical Service, Louis Stokes Cleveland Department of Veterans Affairs Medical Center, Cleveland, OH 44106, USA.

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

an endoplasmic reticulum marker, and porin as a mitochondria marker. In most experiments, we detected little or no tubulin or calreticulin in the mitochondrial fraction. Although GRIM-19 is reported to be located in the nucleus and cytosol, we only observed the protein in mitochondria (*3*, *4*). Similar results were obtained from mitochondria isolated from the mouse brain, kidney, and primary splenocytes, as well as several other mouse and human cell lines. We have yet to find a tissue or cell line that does not contain Stat3 in the mitochondrial fraction.

To confirm that mStat3 did not represent contamination from the cytoplasm, we immunobloted proteins from increasing amounts of purified mitochondria for Stat3, tubulin, and cytochrome c (Fig. 1B). The ratio of Stat3 to tubulin in the cytoplasm was 0.9 in the heart and 1.0 in the liver. The ratio of Stat3 to tubulin in the heart mitochondria ranged from 13 to 19 and in the liver mitochondria ranged from 4 to 9. If the Stat3 detected in the mitochondria was due to cytosolic contamination, then the ratio of Stat3 to tubulin should have been ~1.

To determine whether mStat3 was on the outer membrane, we incubated liver mitochondria with proteinase K in the presence or absence of triton X-100. Stat3 and GRIM-19 (an inner membrane protein) were resistant to proteinase K, whereas Bcl-2 and porin, both of which are localized on the outer membrane, were degraded (Fig. 1C). When proteinase K was added to mitochondria in the presence of triton X-100, all of the proteins were degraded. These results indicate that Stat3 does not adhere to the outer membrane of the mitochondria in a potentially nonspecific manner and is probably located in either the intermembrane space, the inner mitochondrial membrane, or the matrix.

We detected Stat3 in immunoprecipitates of mitochondrial extracts prepared from the liver with the use of a monoclonal antibody that captures all components of complex I of the ETC (Fig. 2) (*8*). We focused on complex I because GRIM-19 is a component of complex I, and Stat3 has been shown to interact with GRIM-19 (*3*, *4*). As a control, identical samples were incubated with isotype-matched antibody under the same conditions. To ensure antibody specificity, the blots were also probed for known components of complex I. Immunoprecipitates of complex I contained Stat3, whereas no immunoreactive Stat3 or any other proteins were present in the immunoprecipitates using the immunoglobulin G (IgG)–matched control. As expected, NDUFA9 [NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 9, where NADH is the reduced form of nicotinamide adenine dinucleotide] and GRIM-19 (both components of complex I) were immunoprecipitated with the antibody. The complex I antibody did not immunoprecipitate complex III core protein 2, complex IV subunit I, or the 19-kD adenosine triphosphate (ATP) synthase, a component of complex V (5th, 6th, and 7th panels in Fig. 2). The complex II Fp subunit did immunoprecipitate with the complex I antibody (4th panel in Fig. 2). These results indicate the presence of Stat3 in complex I and possibly II of the ETC. Alternatively, there could be an interaction of complexes I and II of the ETC.

To examine whether mitochondrial Stat3 expression might modulate oxidative phosphorylation, we performed polarographic studies with the use of intact mitochondria isolated from $Stat3^{+/+}$ and $Stat3^{-/-}$ pro-B cells. This assay evaluates the integrated function of the ETC coupled to ATP synthesis, membrane transport, dehydrogenase activities, and the structural integrity of the mitochondria. Glutamate was used to drive complex I–dependent electron transport. Succinate is the specific substrate that we used as the electron donor for complex II. We were not able to accumulate a sufficient amount of mitochondria to measure complex III activity. We measured complex IV activity using N,N,N′,N′-tetramethyl-*p*-phenylenediamine (TMPD) with L-ascorbate.

Maximal state 3 oxidation rates with adenosine diphosphate (ADP) (2 mM) were reduced by 70% in $Stat3^{-/-}$ cells with glutamate as a substrate for complex I and by 50% in cells with succinate as a substrates for complex II (Table 1). Oxidation rates for TMPD/ascorbate (complex IV) were not substantially different, localizing defects to complexes I, II, and/or III of the ETC.

Oxidase activities were measured in mitochondria permeabilized by freezing and thawing (*9*). The oxidase studies bypass potential defects in substrate carriers to exclude defects of substrate transport as causes of the decreased oxidation rates. NADH oxidase activity requires complexes I, III, and IV, and duroquinol oxidase (DHQ) requires complexes III and IV. Consistent with the respiration rates obtained with intact mitochondria, NADH oxidase activity was reduced by 65% in mitochondrial preparations from $Stat3^{-/-}$

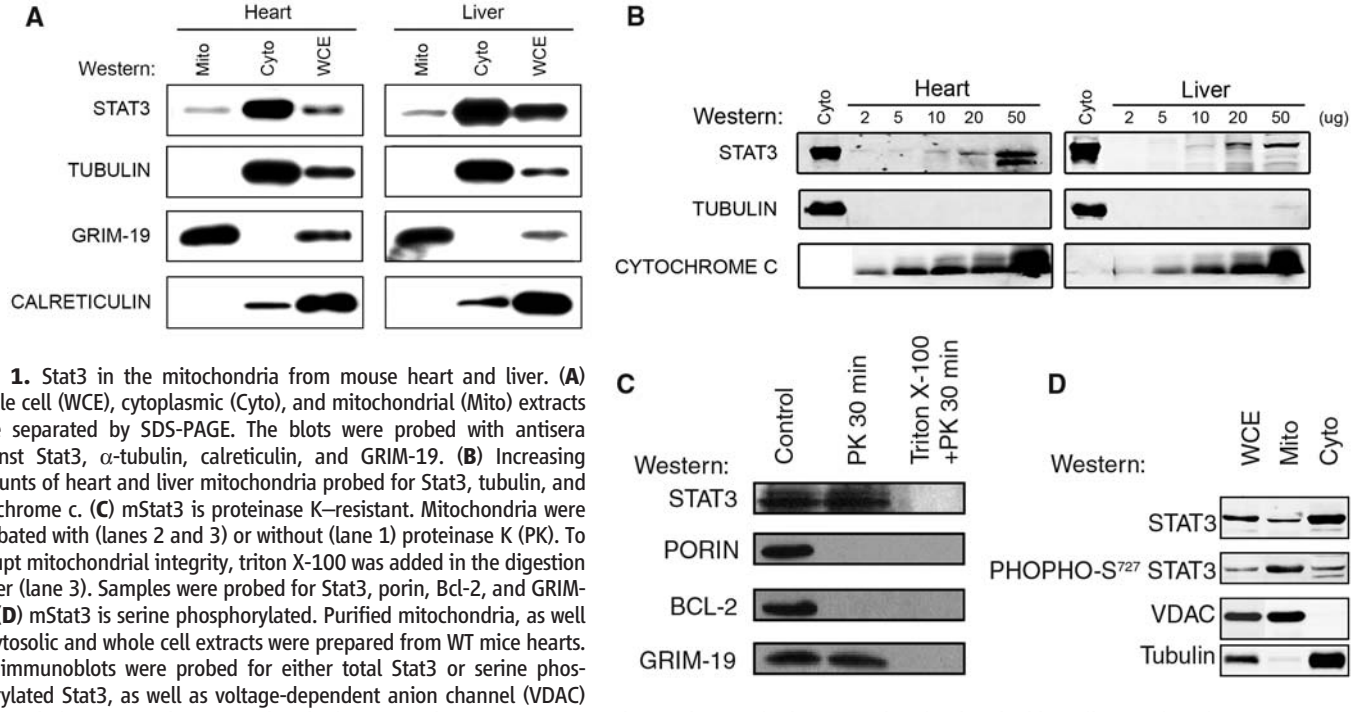


Fig. 1. Stat3 in the mitochondria from mouse heart and liver. (**A**) Whole cell (WCE), cytoplasmic (Cyto), and mitochondrial (Mito) extracts were separated by SDS-PAGE. The blots were probed with antisera against Stat3, α-tubulin, calreticulin, and GRIM-19. (**B**) Increasing amounts of heart and liver mitochondria probed for Stat3, tubulin, and cytochrome c. (**C**) mStat3 is proteinase K–resistant. Mitochondria were incubated with (lanes 2 and 3) or without (lane 1) proteinase K (PK). To disrupt mitochondrial integrity, triton X-100 was added in the digestion buffer (lane 3). Samples were probed for Stat3, porin, Bcl-2, and GRIM-19. (**D**) mStat3 is serine phosphorylated. Purified mitochondria, as well as cytosolic and whole cell extracts were prepared from WT mice hearts. The immunoblots were probed for either total Stat3 or serine phosphorylated Stat3, as well as voltage-dependent anion channel (VDAC) and α-tubulin as controls for mitochondrial purity. A fluorescent conjugated secondary antibody was used to develop the blots allowing the relative amount of total and serine phosphorylated Stat3 to be measured. The ratio of total Stat3 to serine phosphorylated Stat3 in whole cell extracts was 2.5, in cytosolic extracts was 2.3, and in mitochondria extracts was 0.3.

pro-B cells as compared with that in wild-type (WT) pro-B cells (Fig. 3A). In contrast to that of NADH oxidase, the activities of DHQ (Fig. 3A) were similar in Stat3$^{-/-}$ and WT pro-B cells.

These results reveal a defect at complexes I and II of the ETC in mitochondria from Stat3$^{-/-}$ pro-B cells. To further localize the ETC defects, we measured enzyme activities of the electron transport chain complexes (ETC assays) in solubilized mitochondria. Consistent with the polarographic studies, the defects of complexes I were decreased by 40% in Stat3$^{-/-}$ pro-B cells as compared with WT pro-B cells (Fig. 3B). Complex II activity in Stat3$^{-/-}$ cells was also decreased by ~85%. Complex III and IV activities for both WT and Stat3$^{-/-}$ pro-B cells were not affected (Fig. 3B).

To determine whether the deficiency in oxygen consumption in Stat3$^{-/-}$ cells was due to a decrease in mitochondrial content, we measured the activity of the mitochondrial enzyme citrate synthase in WT and Stat3$^{-/-}$ pro-B cells. This enzyme is an exclusive marker of the mitochondrial matrix, and its activity was similar in Stat3$^{+/+}$ and Stat3$^{-/-}$ cells (fig. S1A). We also measured amounts of porin, cytochrome c, and several components of ETC complexes including NDUFA9, NDUFS3 [NADH dehydrogenase (ubiquinone) Fe-S protein 3], complex II subunit Fp (70 kD), complex II subunit I p (30 kD), and COX-I (cytochrome c oxidase-I) in WT and Stat3$^{-/-}$ cells by immunoblotting. All of these proteins were present in similar amounts in both cell types (fig. S1B).

The amount of mitochondrial DNA–encoded genes was similar between WT and Stat3$^{-/-}$ pro-B cells (fig. S1C), as were the concentrations of mitochondrial-encoded RNAs (fig. S1D). These results suggest that mitochondrial content is not affected in pro-B cells that lack Stat3.

To determine whether alterations in ETC were directly due to a loss of Stat3 expression, we reconstituted Stat3a into Stat3$^{-/-}$ B cells with a retrovirus expressing Stat3a and a downstream green fluorescent protein cDNA to allow identification of

Table 1. Maximal rates of state 3 respiration in mitochondria isolated from WT and Stat3$^{-/-}$ pro-B cells with glutamate, succinate, or TMPD/L-ascorbate as substrates. Results are expressed as nAtom O/min/mg of mitochondrial protein and are presented as mean ± SD. Significant differences between Stat3$^{-/-}$ and WT pro-B cells are shown in the table as the corresponding *P* value. *N* denotes the number of independent experiments.

Electron transport chain complexes required	WT 2 mM ADP	Stat3$^{-/-}$ 2 mM ADP
Complexes I, III, IV 20 mM glutamate	52 ± 28 (*N* = 4)	15 ± 9 (*N* = 4) *P* = 0.049
Complexes II, III, IV 20 mM succinate	129 ± 21 (*N* = 4)	64 ± 15 (*N* = 4) *P* = 0.005
Complex IV 1mM TMPD + 20 mM L-ascorbate	521 ± 92 (*N* = 2)	468 ± 40 (*N* = 2)

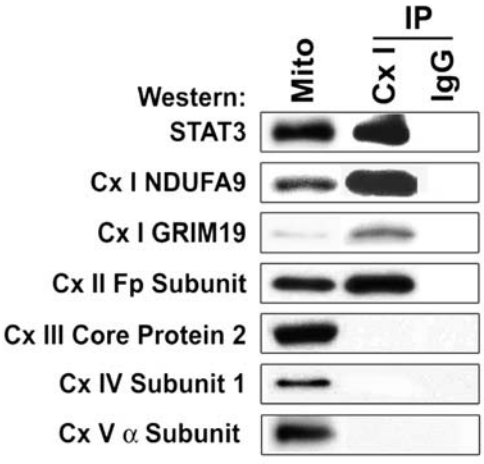


Fig. 2. Stat3 in complex I immunoprecipitates (IP). Antibody to complex I (CxI) or a nonspecific isotype-matched IgG were incubated with liver mitochondrial extracts. Immunoprecipitates were resolved on SDS-PAGE and probed for Stat3, GRIM-19, NDUFA9, Cx II Fp subunit, Cx III core protein 2, Cx IV subunit I, and Cx V α subunit.

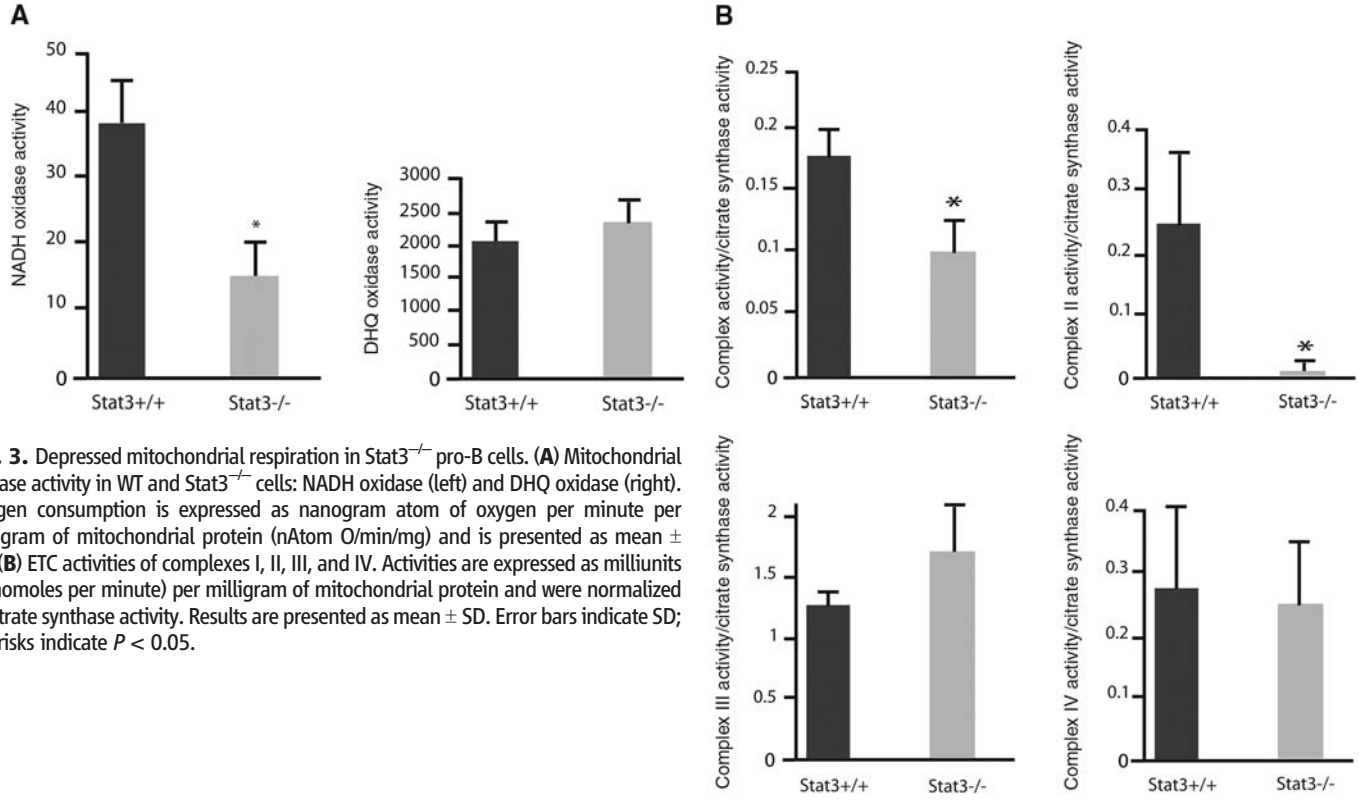


Fig. 3. Depressed mitochondrial respiration in Stat3$^{-/-}$ pro-B cells. (**A**) Mitochondrial oxidase activity in WT and Stat3$^{-/-}$ cells: NADH oxidase (left) and DHQ oxidase (right). Oxygen consumption is expressed as nanogram atom of oxygen per minute per milligram of mitochondrial protein (nAtom O/min/mg) and is presented as mean ± SD. (**B**) ETC activities of complexes I, II, III, and IV. Activities are expressed as milliunits (nanomoles per minute) per milligram of mitochondrial protein and were normalized to citrate synthase activity. Results are presented as mean ± SD. Error bars indicate SD; asterisks indicate $P < 0.05$.

infected cells (*10*). Expression of Stat3a restored complex I and II activities (see Table 2, legend). Immunoblots of isolated mitochondria and cytosol from the reconstituted cells confirmed that Stat3a was expressed in both cellular fractions (fig. S2A).

To determine whether Stat3, expressed only in mitochondria, restored respiration in Stat3$^{-/-}$ cells, we constructed a Stat3 that contains the mitochondrial targeting sequence of cytochrome c oxidase subunit VIII (MLS) fused to the N terminus of the protein (MLS Stat3). This sequence places the protein of interest in the inner mitochondrial membrane. Fractionation of MLS Stat3–expressing cells demonstrated that Stat3 was present only the mitochondria (fig. S2B). Cells expressing MLS Stat3 had about the same amount of Stat3 in the mitochondria as did WT cells, but they had less total Stat3 than WT cells. MLS Stat3 cells (Table 2, row 3) showed similar activities of complexes I and II of the ETC as Stat3$^{+/+}$ or Stat3$^{-/-}$ cells that expressed Stat3a (Table 2, row 1).

To determine whether tyrosine and serine phosphorylation and DNA binding activity of Stat3 are important for regulation of the ETC, MLS Stat3 that contained mutations in its DNA binding domain (*11*), Tyr^{705} (Table 2, rows 4 and 7), or Ser^{727} and Tyr^{705} (Table 2, row 5) was expressed in Stat3$^{-/-}$ cells. Mutation in either Tyr^{705} or the DNA binding domain (Table 2, rows 4 and 7) restored activity of complexes I and II (Table 2). Because MLS Stat3 E434A/E435A (*12*) cannot form a dimer, it is possible that Stat3 exerts its actions in the mitochondria as a monomer (*11*). Consistent with this observation, Stat3$^{-/-}$ cells expressing a constitutively active Stat3 (Stat3a CA) that forms a dimer without being tyrosine phosphorylated (*13*) did not restore mitochondrial respiration (Table 2, row 8).

Expression of MLS Stat3 with serine 727 mutated to an alanine and tyrosine 705 mutated to a phenylalanine (row 5) also did not restore activity of complex I or II. To further characterize the role of Ser^{727}, we mutated Tyr^{705} to a phenylalanine and Ser^{727} to either an alanine (Y705F/S727A) (row 5) or an aspartic acid (Y705F/S727D) (row 6). The former functions as a dominant negative and the latter as a mimetic of a constitutively serine-phosphorylated Stat3. MLS Stat3Y705F/S727D reconstituted complex I and II activities (Table 2, row 6), whereas MLS Y705F/S727A (Table 2, row 5) was ineffective, another indication that the actions of this protein in the mitochondria are distinct from its activity as a transcription factor. Interestingly, the relative concentration of serine-phosphorylated Stat3 was highly enriched in mouse mitochondria as compared with that present in the cytoplasm (Fig. 1D).

To confirm that MLS Stat3 and variants did not have transcriptional activity, we compared the expression of *SOCS3*, a Stat3-dependent gene, in cells incubated with or without IFNβ (*14*). Real-time polymerase chain reaction demonstrated that incubation of WT pro-B cells with IFNβ induced an eightfold increase in SOCS3 mRNA (fig. S3). However, Stat3$^{-/-}$ cells or those cells expressing MLS Stat3 or MLS Stat3 with a mutation in its DNA binding domain showed no increase in SOCS3 expression (fig. S3). Furthermore, cells that expressed constitutively active Stat3 in which respiration was not restored displayed robust IFNβ-stimulated expression of SOCS3 RNA. We obtained similar results when we examined IFNβ–induced expression of adenosine deaminase 1 RNA in pro-B cells (fig. S3). These findings are another independent support of the concept that MLS Stat3 controls the activity of the electron transport chain, and this effect of Stat3 is unrelated to its actions as a transcription factor.

To examine the role of Stat3 in mitochondrial respiration in vivo, we examined respiration in Stat3$^{-/-}$ mouse hearts. Mice that do not express Stat3 in cardiomyocytes develop cardiac inflammation with fibrosis, have dilated cardiomyopathy, and die prematurely of congestive heart failure (*15*). Female mice that do not express Stat3 in cardiomyocytes also develop postpartum cardiomyopathy, which is also seen in

Table 2. Activities of complexes I and II of the respiratory chain in mouse Stat3$^{-/-}$ pro-B cells reconstituted with Stat3 mutants. Complex I and II activity in Stat3$^{-/-}$ pro-B cells and different mutants of Stat 3 expressed in terms of percentage values, considering the activity of complexes I and II being 100% in Stat3$^{-/-}$ pro-B cells reconstituted with WT Stat3a. (Stat3$^{-/-}$ cells reconstituted with Stat3a showed complex I activity of 111 ± 9 and complex II activity of 74 ± 29, as compared with that seen in WT pro-B cells.) Results are presented as mean ± SD. Significant differences with *P* values of <0.05 were observed between Stat3$^{-/-}$ +Stat3a and Stat3$^{-/-}$ cells reconstituted with Stat3 mutants.

	Cell type	Complex I	Complex II
1	Stat3$^{-/-}$ + Stat3a	100 (*N* = 7)	100 (*N* = 3)
2	Stat3$^{-/-}$ + MSCV	44 ± 26 (*N* = 7)*	46 ± 26 (*N* = 3)*
3	Stat3$^{-/-}$ + MLS Stat3a	82 ± 20 (*N* = 6)	114 ± 42 (*N* = 3)
4	Stat3$^{-/-}$ + MLS Stat3a Y705F	83 ± 34 (*N* = 4)	124 ± 75 (*N* = 3)
5	Stat3$^{-/-}$ + MLS Stat3a Y705F/S727A	28 ± 19 (*N* = 4)*	39 ± 30 (*N* = 3)*
6	Stat3$^{-/-}$ + MLS Stat3a Y705F/S727D	98 ± 20 (*N* = 3)	115 ± 80 (*N* = 3)
7	Stat3$^{-/-}$ + MLS Stat3a E434A/E435A	90 ± 6 (*N* = 3)	126 ± 49 (*N* = 3)
8	Stat3$^{-/-}$ + Stat3CA	61 ± 4 (*N* = 3)*	16 ± 5 (*N* = 3)*

*$P < 0.05$

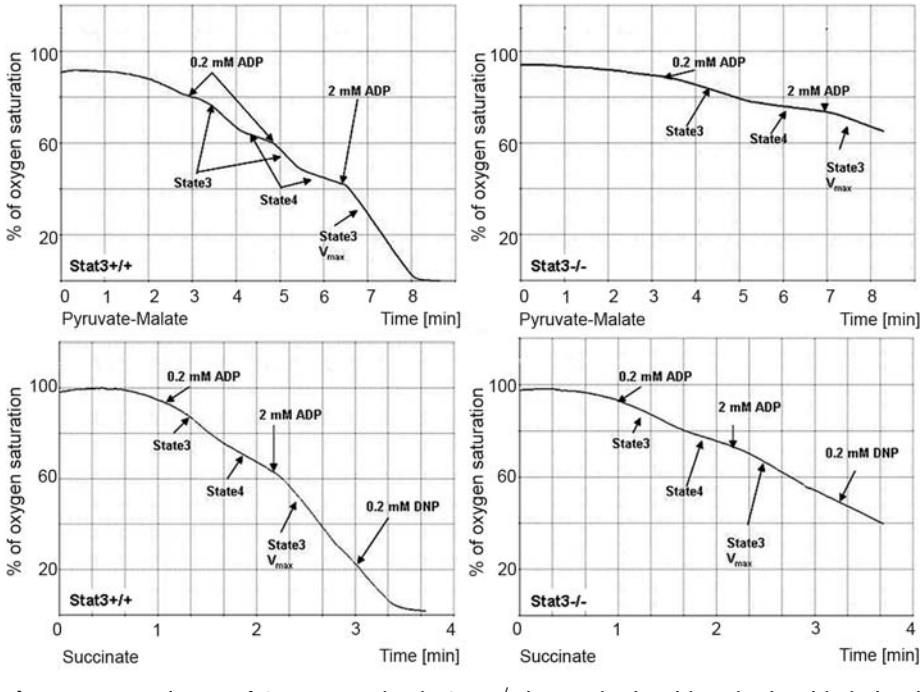


Fig. 4. Decreased rates of O_2 consumption in Stat3$^{-/-}$ heart mitochondria. Mitochondria isolated from littermates of WT (left) and Stat3$^{-/-}$ (right) hearts were incubated in an oxygen sensor chamber, and O_2 consumption (*y* axis) as a function of incubation time (*x* axis) was recorded. In the upper panels, the mitochondria were incubated with pyruvate/malate (complex I substrates), and they were incubated with succinate (complex II substrate) in the lower panels. Different concentrations of ADP were added to the mitochondria to measure state 3 respiration (0.2 mM ADP) or state 3 V_{max} rates of respiration (2 mM ADP). State 4 rates of respiration were calculated when 0.2mM ADP was depleted from the reaction.

humans with reduced Stat3 expression in the myocardium (*16*).

Mitochondria from the hearts of Stat3$^{flox/flox/cre}$ (Stat3$^{-/-}$) and Stat3$^{flox/flox}$ (wild type) mice were assayed for complex I and II activity. The mice used for these studies were 8 weeks old, at which time the hearts are normal as analyzed by histology and various physiological parameters (*15*). We used pyruvate and malate as complex I substrates (upper panels) and succinate as a complex II substrate (lower panels) to measure O_2 consumption in WT and Stat3$^{-/-}$ heart mitochondria (Fig. 4). Rates of O_2 consumption were decreased in Stat3$^{-/-}$ mitochondria with both substrates. Maximal rates of state 3 respiration (V_{max}) stimulated by the addition of 2 mM ADP were blunted in Stat3$^{-/-}$ mitochondria. Pyruvate and malate V_{max} in state 3 was decreased by ~30% and succinate state 3 V_{max} by ~60% (table S1). ETC assays confirmed that Stat3$^{-/-}$ heart mitochondria had defects in complexes I and II but had normal activity of complex III (fig. S4).

The results presented here reveal a role of Stat3 as a modulator of mitochondrial respiration. The most likely mechanism by which Stat3 exerts its actions is not as a transcription factor that regulates nuclear gene expression, but rather through its localization in the mitochondria. The effects of Stat3 potentially represent a general mechanism by which this protein can influence cell metabolism. These studies are consistent with the concept of multidirectional communication between the mitochondria and the nucleus and changes in homeostasis detected at the plasma membrane. The precise mechanism by which Stat3 regulates complexes I and II remains to be determined.

References and Notes

1. S. J. Baker, S. G. Rane, E. P. Reddy, *Oncogene* **26**, 6724 (2007).
2. J. E. Angell, D. J. Lindner, P. S. Shapiro, E. R. Hofmann, D. V. Kalvakolanu, *J. Biol. Chem.* **275**, 33416 (2000).
3. C. Lufei *et al.*, *EMBO J.* **22**, 1325 (2003).
4. J. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9342 (2003).
5. I. M. Fearnley *et al.*, *J. Biol. Chem.* **276**, 38345 (2001).
6. G. Huang *et al.*, *Mol. Cell. Biol.* **24**, 8447 (2004).
7. Materials and methods are available as supporting material on *Science* Online.
8. J. Murray *et al.*, *J. Biol. Chem.* **278**, 13619 (2003).
9. S. Krahenbuhl, M. Chang, E. P. Brass, C. L. Hoppel, *J. Biol. Chem.* **266**, 20998 (1991).
10. A. M. Gamero *et al.*, *J. Biol. Chem.* **281**, 16238 (2006).
11. C. M. Horvath, Z. Wen, J. E. Darnell Jr., *Genes Dev.* **9**, 984 (1995).
12. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
13. J. F. Bromberg *et al.*, *Cell* **98**, 295 (1999).
14. L. Zhang *et al.*, *Mol. Cell. Biochem.* **288**, 179 (2006).
15. J. J. Jacoby *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12929 (2003).
16. D. Hilfiker-Kleiner *et al.*, *Cell* **128**, 589 (2007).

17. This work was supported by NIH grants CA098924 (A.C.L.) and PO1AG15885 and the Office of Research and Development, Medical Research Service Department of Veterans Affairs (E.J.L.). We thank I. Scheffler for his many suggestions. DHQ was a kind gift from C. Hoppel. This manuscript is dedicated to Joseph Larner in honor of 60 years of scientific achievement and enthusiasm.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1164551/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References

12 August 2008; accepted 5 December 2008
Published online 8 January 2009;
10.1126/science.1164551
Include this information when citing this paper.

Platelets Kill Intraerythrocytic Malarial Parasites and Mediate Survival to Infection

Brendan J. McMorran,[1]* Vikki M. Marshall,[2]† Carolyn de Graaf,[2,3] Karen E. Drysdale,[1] Meriam Shabbar,[1] Gordon K. Smyth,[2] Jason E. Corbin,[2] Warren S. Alexander,[2] Simon J. Foote[1]

Platelets play a critical role in the pathogenesis of malarial infections by encouraging the sequestration of infected red blood cells within the cerebral vasculature. But platelets also have well-established roles in innate protection against microbial infections. We found that purified human platelets killed *Plasmodium falciparum* parasites cultured in red blood cells. Inhibition of platelet function by aspirin and other platelet inhibitors abrogated the lethal effect human platelets exert on *P. falciparum* parasites. Likewise, platelet-deficient and aspirin-treated mice were more susceptible to death during erythrocytic infection with *Plasmodium chabaudi*. Both mouse and human platelets bind malarial-infected red cells and kill the parasite within. These results indicate a protective function for platelets in the early stages of erythrocytic infection distinct from their role in cerebral malaria.

Adaptive immunity to malarial infection accumulates slowly over the lifetime of an individual living in an endemic region. Cross-immunity between isolates is low, and every new infection requires the development of a virtually novel immune response. Therefore, innate mechanisms that limit parasite growth within the red blood cell (RBC) are extremely important in determining survival, especially during the first few years of life, because clinical severity correlates closely with parasite mass. Several known mutations affect the RBC and decrease the severity of infection through either impediment of parasite entry or development within the cell (*1*). However, we know little about other non-RBC mechanisms that may influence the course of infection before the onset of effective adaptive immune responses.

Thrombocytopenia is a common clinical accompaniment of *Plasmodium falciparum* (*2*–*4*), *Plasmodium vivax* (*5*, *6*), *and Plasmodium chabaudi* (*7*) malarial infections. Low platelet concentrations correlate with increased parasite density (*8*, *9*) and poor outcome (*10*). Platelets bind preferentially to infected RBCs (*11*, *12*) and have been postulated to play a role in the pathogenesis of malarial infection, either positively (*13*, *14*) or negatively (*15*, *16*). Given the role platelets play in other infectious diseases (*17*), we explored their influence on the course of a malarial infection in the mouse model, *P. chabaudi*, and in vitro using the cultured human malarial parasite, *P. falciparum*.

The megakaryocyte growth and differentiation factor, C-mpl, is the receptor for thrombopoietin and is encoded by the *Mpl* gene. Homozygous disruption of the *Mpl* gene on a C57BL/6 background (Mpl$^{-/-}$) results in mice with only one-tenth as many circulating platelets as their undisrupted counterparts have (*18*). Platelet-deficient Mpl$^{-/-}$ mice were significantly more susceptible to death from *P. chabaudi* than were isogenic C57BL/6 mice ($P < 0.0001$, Mantel-Cox log-rank test on survival data), irrespective of well-known sex differences. In females, 5% of C57BL/6 versus 50% of Mpl$^{-/-}$ mice died. Twenty percent of the male C57BL/6 mice and all the male Mpl$^{-/-}$ mice died (Fig. 1, A and B).

Cohort analysis of groups of mice showed that platelet numbers decreased concomitantly with the appearance of parasites in the peripheral circulation (Fig. 1, E and F). Platelet numbers dropped to 10 to 20% of normal values in the C57BL/6 mice, with a smaller percentage drop in the Mpl$^{-/-}$ mice. Platelet numbers reached a nadir before the onset of severe disease symptoms and anemia. No evidence of excess bleeding was seen in the Mpl$^{-/-}$ mice during autopsy. Similar erythropoietic kinetics occurred between Mpl$^{-/-}$ and C57BL/6 mice during both malarial challenge

[1]The Menzies Research Institute, University of Tasmania, Private Bag 23, Hobart, Tasmania 7000, Australia. [2]The Walter and Eliza Hall Institute of Medical Research, 1G Royal Pde, Parkville, Victoria 3050, Australia. [3]Department of Medical Biology, University of Melbourne, Parkville, Victoria 3010, Australia.

*To whom correspondence should be addressed. E-mail: brendan.mcmorran@utas.edu.au

†Present address: Peter MacCallum Cancer Centre Research Division, Locked Bag 1, A'Beckett Street, Victoria 8006, Australia.

and experimentally induced hemolysis (figs. S1 and S2), excluding differential erythropoiesis as a cause for the different susceptibilities. Nor were differences in immunological responses, including production of lymphocytes and inflammatory cytokines, observed between the mouse strains during infection (fig. S3).

Acetyl salicylic acid (aspirin) inhibits cyclooxygenases I and II, resulting in a decrease in thromboxane A2 and decreased platelet activation as seen by lower platelet aggregation (*19*). Mice were given 25 mg of aspirin per kilogram of body weight intraperitoneally daily from the day before *P. chabaudi* infection until day 10 of infection and monitored for survival. C57BL/6 and $Mpl^{-/-}$ mice of both sexes treated with aspirin were significantly more susceptible than placebo-treated mice (Fig. 1, C and D) (Cox proportional hazard analysis, $P = 0.0009$). Thus, pharmacological inhibition of platelet function increases susceptibility to malarial infection in mice, although other indirect effects of aspirin cannot be excluded.

Synchronized cultures of *P. falciparum* trophozoites were grown in human RBCs and incubated with increasing concentrations of purified human platelets. A concentration-dependent increase in the inhibition of parasite growth was observed (Fig. 2A). Growth inhibition (measured after 44 hours of incubation) peaked at 65% with 25 to 75 million platelets/ml. Most viable parasites observed at this time were mature trophozoites and shizonts. Reductions in mature shizonts at 5 hours and new rings at 21 to 29 hours (fig. S4) suggest that platelets were inhibiting developmental maturation of the parasites.

Using the *P. falciparum* model, we investigated the dependence of parasite growth inhibition on platelet function by first incubating platelets with specific platelet antagonists. Maintenance of elevated intracellular adenosine 3′,5′-monophosphate and guanosine 3′,5′-monophosphate concentrations are key mechanisms in suppressing platelet activation (*20*, *21*). The addition of the adenylyl cyclase activator prostaglandin E1 completely abrogated the platelet growth-inhibitory activity. Exposure of platelets to the guanylyl cyclase activator, NO (donated by sodium nitroprusside), also suppressed platelet-mediated effects on parasite growth (Fig. 2B). To test for the requirement of adenosine 5′-diphosphate (ADP), an important physiological regulator of platelet activation [reviewed in (*20*)], we incubated the platelet–*P. falciparum* cultures with adenosine diphosphatase (ADPase) (*22*) to remove extracellular ADP from the culture medium. Unimpeded parasite growth in the presence of platelets and ADPase was observed. ADP-dependent platelet activation is mediated via two metabotrophic puronergic receptors, $P2Y_1$ and $P2Y_{12}$. Antagonism of $P2Y_1$, but not $P2Y_{12}$, abolished the parasite killing activity (Fig. 2B). Thus, *P. falciparum* growth is inhibited by platelets in a concentration-dependent and activation-dependent manner that requires ADP and activation through the $P2Y_1$ receptor.

As with *P. chabaudi* in the mouse, inhibition of cyclooxygenase by aspirin also prevented the inhibition of parasite growth by platelets (Fig. 2B).

To approximate a clinical effect, a volunteer platelet-donor took aspirin orally twice daily (4.2 mg/kg) for a week before donating blood for platelet

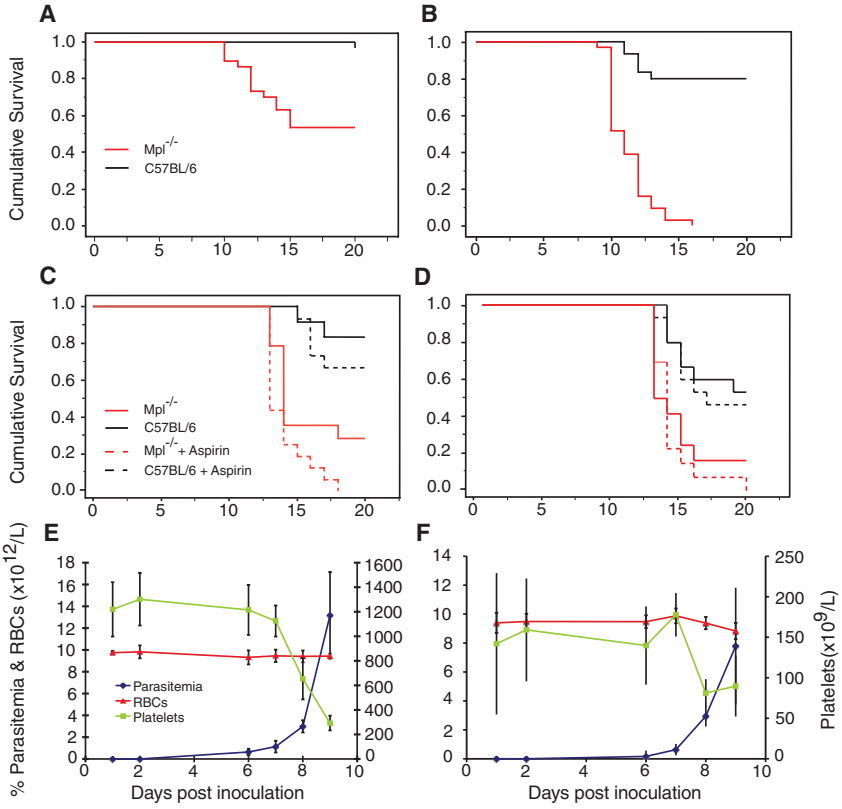


Fig. 1. Survival data and parasitological and hematological kinetics of *P. chabaudi* infection in C57BL/6 and $Mpl^{-/-}$ mice. (**A** and **B**) Survival curves for $Mpl^{-/-}$ (red) and C57BL/6 (black) mice infected with *P. chabaudi* and monitored for survival. (**C** and **D**) Effect of aspirin or vehicle-only treatment on survival of $Mpl^{-/-}$ and C57BL/6 mice. Given that gender contributes a significant effect, the females (A and C) and males (B and D) have been separated. (**E** and **F**) Parasite and hematological parameters in C57BL/6 (E) and $Mpl^{-/-}$(F) mice. Parasitemias (percentage of infected RBCs), RBC counts ($\times 10^{12}$/liter), and platelet counts ($\times 10^9$/liter).

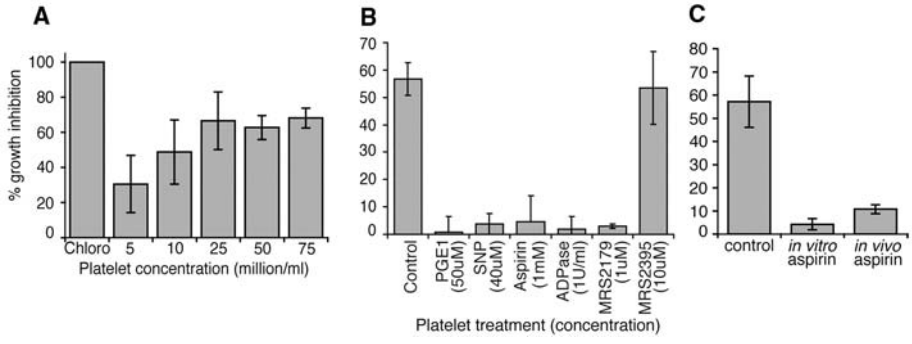


Fig. 2. Platelet-mediated growth inhibition of *P. falciparum* cultured in human RBCs. (**A**) Parasite growth inhibition (relative to Tyrode's buffer–treated cultures) with increasing concentrations of purified platelets. Chloro, parasites incubated with 640 nM chloroquine. (**B**) Parasite growth inhibition with platelets (50 million/ml) preincubated in platelet activation inhibitors at the indicated concentrations. Control, untreated platelets; PGE1, prostaglandin E1; SNP, sodium nitroprusside; MRS2179, $P2Y_1$ receptor antagonist; MRS2395, $P2Y_{12}$ receptor antagonist. (**C**) Inhibition of parasite growth in the presence of platelets (50 million/ml) preincubated with aspirin before addition to culture (in vitro aspirin) or untreated (control) compared with platelets taken from the same donor on separate occasions after treatment with aspirin (4.2 mg/kg taken orally twice daily for 1 week before platelet collection) (in vivo aspirin). All experiments were performed with washed platelets. Drugs used alone did not affect parasite growth. Bars for graphs represent the means of at least three independent experiments (A and B) and two independent experiments (C). Error bars indicate ±1 SD.

purification. These platelets were unable to inhibit growth of *P. falciparum* parasites (Fig. 2C), implying that aspirin is potentially harmful in malarial infections.

In mice infected with *P. chabaudi* for 8 days, with parasitemias at ~10 to 15%, platelets were preferentially bound to infected RBCs. As shown with Giemsa-stained thin smears and light microscopy (Figs. 3A, i and ii), the percentage parasitemia was twice as high when only platelet-bound RBCs were counted compared with all RBCs (Fig. 3B). CD41 (a platelet surface antigen) on the surface of RBCs served as a marker for platelet binding. Three times as many infected RBCs stained for CD41 as did uninfected RBCs (Fig. 3C). Similar results were obtained from cultured *P. falciparum*, where twice the numbers of infected RBCs were bound by human platelets (counted on Giemsa-stained thin smears) (Fig. 3A, iii and iv, and 3D) and twice the number of infected RBCs were also CD41 positive (Fig. 3E). Therefore, platelets selectively bind infected RBCs.

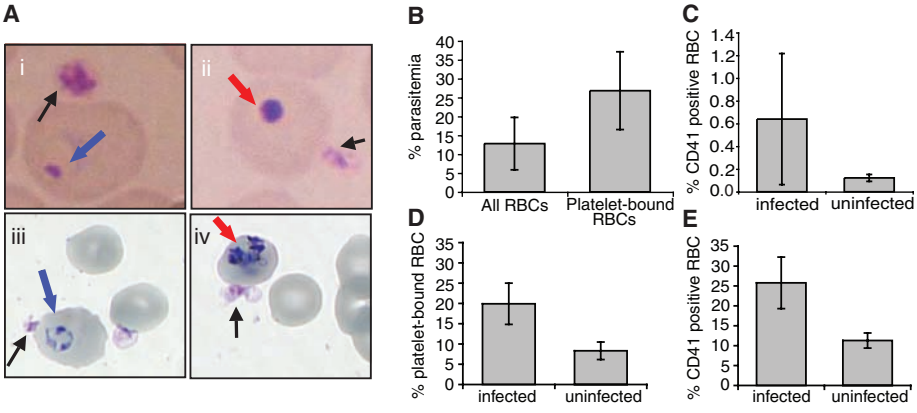


Fig. 3. Infected RBCs are preferentially bound by platelets. (**A**) Giemsa-stained thin blood smears showing individual RBC-bound platelets (black arrows) [(i) and (ii), mouse; (iii) and (iv), human]. Some parasites look dysmorphic (red arrows) compared with normal parasites (blue arrows) (**B**) The percentage parasitemia (counted on Giemsa-stained thin smears) of all mouse RBCs compared to only those with platelets bound ($P < 0.01$). (**C**) The percentage CD41 staining of thiazole orange (TO) positive (infected) versus TO-negative (uninfected) mouse RBCs ($P < 0.05$) (**D**) Percentage of *P. falciparum*–infected and uninfected human RBCs bound by platelets counted on Giemsa-stained thin smears ($P < 0.05$). (**E**) Percentage of infected (TO-positive) and uninfected (TO-negative) human RBCs that stained positively for CD41 ($P < 0.01$). (B) and (C) refer to the in vivo mouse model (day 8 of infection) and (D) and (E) to the in vitro human falciparum culture model (after 24 hours of incubation with platelets). Bars for graphs represent mean of three independent experiments. Error bars represent ±1 SD.

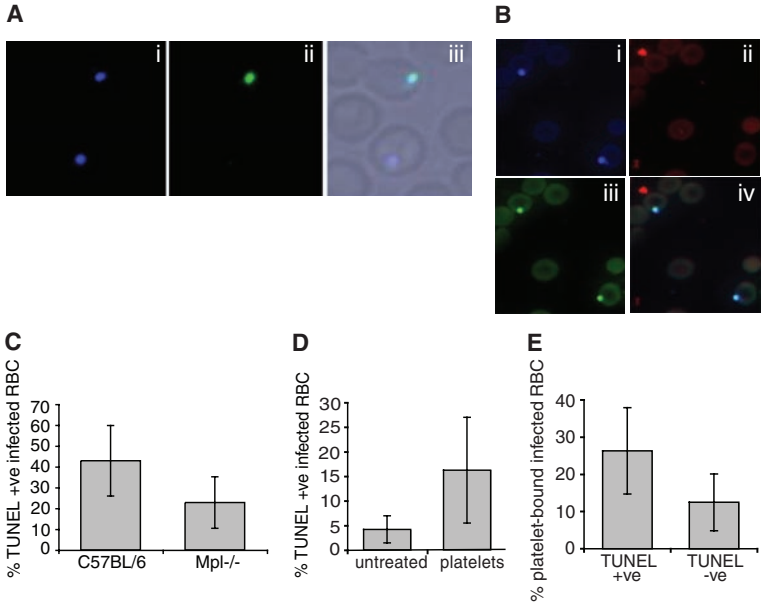


Fig. 4. Evidence that intraerythrocytic parasites are killed by platelets. (**A**) TUNEL staining of a *P. chabaudi*–infected C57BL/6 thin blood smear (day 7 of infection). (i) Hoechst staining of parasite DNA. (ii) TUNEL staining of parasite DNA, and (iii) both TUNEL and Hoechst staining superimposed on dark-field microscopy showing intraerythrocytic parasites. (**B**) Immunofluorescent images of *P. falciparum*-parasitized human RBCs cocultured with platelets for 24 hours. Separate images of cells stained with (i) Hoechst (all parasites), (ii) anti-CD41 (platelets), and (iii) TUNEL (dead parasites); (iv) merged image of (i), (ii), and (iii) (anti-CD41, red; TUNEL, green; Hoechst, blue). (**C**) The percentage of all *P. chabaudi* parasites staining positive with TUNEL in C57BL/6 and Mpl$^{-/-}$ mice ($P < 0.05$). (**D**) Percentage of TUNEL-positive parasites in *P. falciparum* cultures incubated with platelets (50 million/ml) and without platelets for 24 hours ($P < 0.05$). (**E**) Percentage of *P. falciparum*–parasitized RBCs bound by platelets that stained positively or negatively with TUNEL stain ($P < 0.05$). Bars for graphs represent the mean of at least three independent experiments with error bars representing ±1 SD.

The presence of dead parasites in the peripheral circulation was investigated by adapting the terminal deoxynucleotidyl transferase deoxyuridine triphosphate nick end labeling (TUNEL) technique to thin blood smears. TUNEL stains nuclei containing sheared or degraded DNA, an accompaniment of apoptosis or necrosis. As a positive control, all extraerythrocytic parasites, which commonly occur late in *P. chabaudi* infections, stained positive for TUNEL (fig. S5). In mice with lower parasitemias (day 7), TUNEL staining of intraerythrocytic parasites was evident (Fig. 4A), indicative of intraerythrocytic parasite death.

Twice as many dead intraerythrocytic parasites were seen in C57BL/6 mice as in Mpl$^{-/-}$ mice (Fig. 4C). Although no differences in total parasitemia between Mpl$^{-/-}$ and C57BL/6 were detected at this time point (most TUNEL-stained parasites are morphologically indistinguishable from live parasites), parasitemias measured at later time points were significantly higher in Mpl$^{-/-}$ than in C57BL/6 mice (fig. S6). This suggests that platelet-mediated parasite killing influences parasite numbers past the platelet nadir, resulting in higher parasitemias in Mpl$^{-/-}$ mice, and explains their increased susceptibility.

TUNEL-stained intracellular parasites were also observed in the *P. falciparum* cultures (Fig. 4B). The frequency of stained parasites was more than three times as high in platelet-treated cultures (Fig. 4D). Significantly higher proportions of TUNEL-stained than nonstained infected RBCs were bound by human platelets (by staining with antibodies to CD41) (Fig. 4E). The greater numbers of dead intraerythrocytic parasites in the presence of normal numbers of platelet in vivo and in vitro implies that platelets directly kill malarial parasites.

Increased mortality in platelet-deficient mice and direct killing of *P. falciparum* by human platelets indicate that platelets are important in controlling malarial infection. Thrombocytopenia occurs early in infection in both humans and mice, before severe forms of the disease develop (*3*, *6*). Therefore, the major protective effect of platelets will be early in infection. Modulation of parasite growth by platelets will buffer the parasite growth rate and allow the engagement of the adaptive immune response. This finding is in addition to the well-established role for platelets in cerebral

disease (*15*, *23*, *24*). We have also shown that inhibition of platelet activation abrogates the protective effect, which could explain the deleterious effect aspirin may have on malarial outcome (*25*).

References and Notes

1. G. Min-Oo, P. Gros, *Cell. Microbiol.* **7**, 753 (2005).
2. A. D. Adedapo, C. O. Falade, R. T. Kotila, G. O. Ademowo, *J. Vector Borne Dis.* **44**, 266 (2007).
3. U. Hellgren *et al.*, *Bull. World Health Organ.* **67**, 197 (1989).
4. Z. A. Jeremiah, E. K. Uko, *Platelets* **18**, 469 (2007).
5. A. Kakar, S. Bhoi, V. Prakash, S. Kakar, *Diagn. Microbiol. Infect. Dis.* **35**, 243 (1999).
6. M. D. Oh *et al.*, *Am. J. Trop. Med. Hyg.* **65**, 143 (2001).
7. F. J. DeGraves, H. W. Cox, *J. Parasitol.* **69**, 262 (1983).
8. R. D. Horstmann, M. Dietrich, U. Bienzle, H. Rasche, *Blut* **42**, 157 (1981).
9. S. Ladhani, B. Lowe, A. O. Cole, K. Kowuondo, C. R. Newton, *Br. J. Haematol.* **119**, 839 (2002).
10. P. Gerardin *et al.*, *Am. J. Trop. Med. Hyg.* **66**, 686 (2002).
11. K. Chotivanich *et al.*, *J. Infect. Dis.* **189**, 1052 (2004).
12. A. Pain *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1805 (2001).
13. F. Peyron, B. Polack, D. Lamotte, L. Kolodie, P. Ambroise-Thomas, *Parasitology* **99**, 317 (1989).
14. B. Polack, F. Delolme, F. Peyron, *Haemostasis* **27**, 278 (1997).
15. G. E. Grau *et al.*, *J. Infect. Dis.* **187**, 461 (2003).
16. J. Lou *et al.*, *Am. J. Pathol.* **151**, 1397 (1997).
17. M. R. Yeaman, A. S. Bayer, in *Platelets*, A. D. Michelson, Ed. (Academic Press, Burlington, MA, 2007), pp. 727–755.
18. W. S. Alexander, A. W. Roberts, N. A. Nicola, R. Li, D. Metcalf, *Blood* **87**, 2162 (1996).
19. H. J. Weiss, L. M. Aledort, S. Kochwa, *J. Clin. Invest.* **47**, 2169 (1968).
20. L. F. Brass, T. J. Stalker, L. Zhu, D. S. Woulfe, in *Platelets* A. D. Michelson, Ed. (Academic Press, Burlington, MA, 2007), pp. 319–346.
21. S. Rex, J. E. Freedman, in *Platelets*, A. D. Michelson, Ed. (Academic Press, Burlington, MA, 2007), pp. 251–280.
22. M. Cattaneo, in *Platelets*, A. D. Michelson, Ed. (Academic Press, Burlington, MA, 2007), pp. 201–220.
23. G. E. Grau *et al.*, *Eur. Cytokine Netw.* **4**, 415 (1993).
24. S. C. Wassmer *et al.*, *J. Immunol.* **176**, 1180 (2006).
25. M. English *et al.*, *Lancet* **347**, 1736 (1996).

26. We thank M. Cozens for flow cytometry support; D. Senyschen, G. Panoschi, and F. Rodda for technical support; the Australian Red Cross Blood Service for providing purified red blood cells; S. Jackson for advice; R. Anders for providing the *P. falciparum* parasites; and C. Flowers for manuscript preparation. Funding support was from the National Health and Medical Research Council of Australia (Program Grants 490037 and 461219), Australian Cancer Research Foundation, and Howard Hughes Medical Institute. Statistical Analysis. *P* values were calculated by means of two-tailed *t* tests, except where specifically mentioned. Ethics approval for animal experiments was received from the Royal Melbourne Hospital, Melbourne, Australia (2002.053), and University of Tasmania (A0008702). Ethics approval for the platelet donations was received from Human Research Ethics Committee (Tasmania) Network (H0009004).

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5915/797/DC1
Methods
Figs. S1 to S6
References

23 September 2008; accepted 4 December 2008
10.1126/science.1166296

Changes in Cortical Dopamine D1 Receptor Binding Associated with Cognitive Training

Fiona McNab,[1] Andrea Varrone,[2] Lars Farde,[2,3] Aurelija Jucaite,[2,3] Paulina Bystritsky,[1] Hans Forssberg,[1] Torkel Klingberg[1]*

Working memory is a key function for human cognition, dependent on adequate dopamine neurotransmission. Here we show that the training of working memory, which improves working memory capacity, is associated with changes in the density of cortical dopamine D1 receptors. Fourteen hours of training over 5 weeks was associated with changes in both prefrontal and parietal D1 binding potential. This plasticity of the dopamine D1 receptor system demonstrates a reciprocal interplay between mental activity and brain biochemistry in vivo.

Working memory (WM) is the ability to retain information for short periods of time and is important for a wide range of cognitive functions (*1*, *2*). Reduced WM capacity is associated with neurological and psychiatric disorders (*3*, *4*) as well as normal aging (*5*). Several of these conditions are also associated with impaired dopamine transmission (*6*, *7*). Intensive training on WM tasks can improve WM capacity (*8*–*12*) and reduce cognitively related clinical symptoms (*10*). Training-related improvements in WM have been associated with an increase in brain activity in parietal and frontal regions linked to WM (*9*), but the biochemical underpinnings of cognitive training are unknown.

Dopaminergic neurotransmission has a central role in WM performance (*13*–*16*), and cortical dopamine release has been observed in humans during the performance of WM tasks (*17*). In nonhuman primates, locally applied D1 agonists, as well as antagonists, affect both performance and the neuronal firing patterns of prefrontal neurons when information is kept in WM (*18*, *19*). The effects seem to be dose-dependent (*15*, *16*), with evidence of an optimal level, so that either too much or too little stimulation of D1 receptors results in reduced WM performance or tuning of prefrontal activity (*18*–*21*).

The availability of dopamine can lead to the translocation of dopamine D1 receptors from the cytosol to the plasma membrane (*22*), and down-regulation of striatal dopamine D2 receptors has been shown to occur after 7 days of motor

[1]Neuropediatric Unit, Department of Woman and Child Health, Stockholm Brain Institute, Karolinska Institutet, Stockholm, Sweden. [2]Department of Clinical Neuroscience, Psychiatry Section, Stockholm Brain Institute, Karolinska Institutet, Stockholm, Sweden. [3]AstraZeneca Research and Development, Södertälje, Sweden.

*To whom correspondence should be addressed. E-mail: torkel.klingberg@ki.se

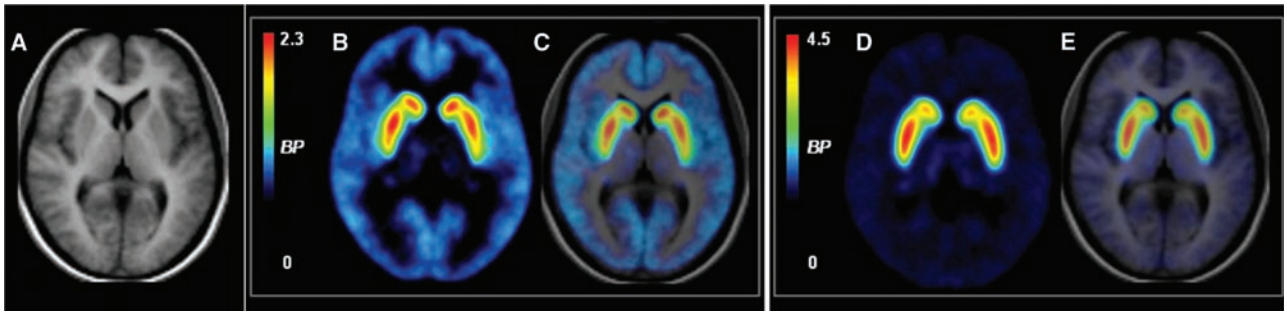


Fig. 1. Maps of baseline D1 and D2 BP, averaged across 13 human volunteers. (**A**) The averaged MRI, normalized to MNI space. (**B**) D1 BP, measured with [^{11}C]SCH23390, averaged across participants (the bar shows absolute D1 BP). (**C**) Overlay of (B) on (A). (**D**) D2 BP, measured with [^{11}C]Raclopride, averaged across participants (the bar shows absolute D2 BP). (**E**) Overlay of (D) on (A).

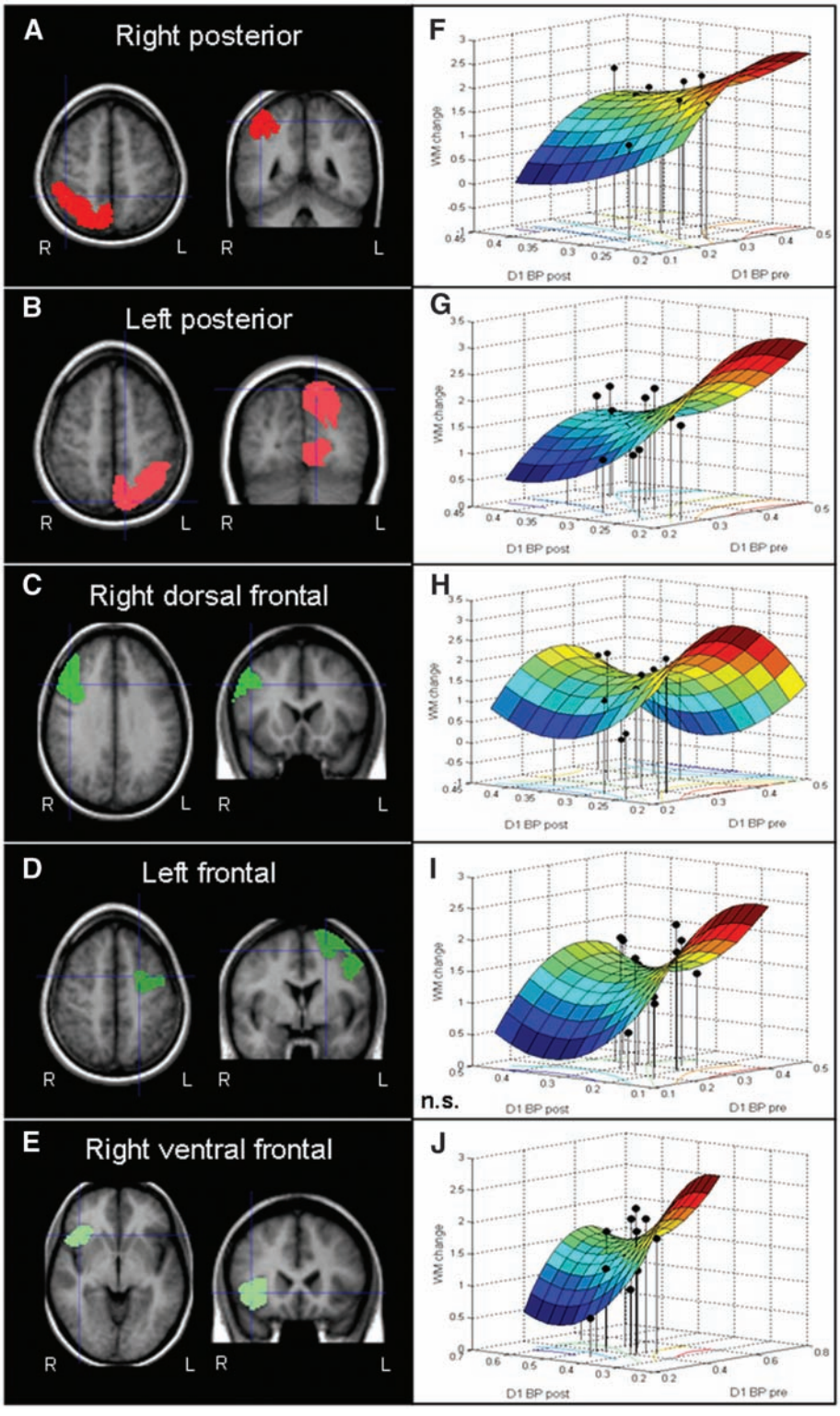


Fig. 2. (**A** to **E**) The five posterior (red) and frontal (green) cortical ROIs, identified from the fMRI results (the contrast of activity recorded during the WM task minus activity recorded during a control task) and used to constrain the analysis of D1 BP. The blue lines indicate where the axial and coronal planes intersect. (**F** to **J**) The application of the quadratic model for the analysis of change in WM capacity for each of the ROIs. The *x* axis shows D1 BP before training, the *y* axis the D1 BP after training, and the *z* axis the improvement in WM. The colored surface represents predicted values, with warmer colors representing higher values on the *z* axis. The black circles represent the observed data. This model predicted change in WM capacity in all except the left frontal ROI [right posterior: $F(2,10) = 4.87$, $P = 0.033$; left posterior: $F(2,10) = 4.82$, $P = 0.034$; right dorsal frontal: $F(2,10) = 7.19$, $P = 0.012$; left frontal: $F(2,10) = 0.71$, $P = 0.515$; right ventral frontal: $F(2,10) = 5.38$, $P = 0.026$].

training in developing rats (*23*). However, the regulation of dopamine receptors as a result of cognitive training has not been studied. We thus investigated the possibility that up- or down-regulation of cortical D1 receptors and subcortical D2 receptors is associated with intensive mental activity during cognitive training.

We used a previously described method of WM training in which participants perform WM tasks with a difficulty level close to their individual capacity limit for about 35 min per day over a period of 5 weeks (*8*–*10*). Thirteen volunteers (healthy males 20 to 28 years old) performed the 5-week WM training. Five computer-based WM tests (three visuospatial and two verbal) were used to measure each participant's WM capacity before and after training, and they showed a significant improvement of overall WM capacity (paired *t* test, $t = 11.1$, $P < 0.001$). The binding potential (BP) of D1 and D2 receptors was measured with positron emission tomography (PET) while the participants were resting, before and after training, using the radioligands [^{11}C]SCH23390 and [^{11}C]Raclopride, respectively (Fig. 1).

To identify brain regions implicated in WM, we conducted functional magnetic resonance imaging (fMRI) on each individual. By comparing activity during a WM task to that during a control task, we identified regions specifically linked to WM ($P < 0.05$, false discovery rate corrected). This resulted in five regions of interest (ROIs) (Fig. 2, A to E), which were used to constrain the analysis of the D1 BP as follows: (i) A right posterior ROI, which included regions of the right parietal, temporal, and occipital cortices; (ii) a left posterior ROI, which included regions of the left parietal, temporal, and occipital cortices; (iii) a right dorsolateral prefrontal ROI, which included the right middle frontal gyrus and right superior frontal gyrus; (iv) a left frontal ROI, which included the left middle frontal gyrus; and (v) a right ventrolateral prefrontal ROI, which included the right inferior frontal gyrus. For calculation of D2 BP, bilateral caudate and putamen ROIs were defined anatomically. Although WM activity in the basal ganglia was not identified from the fMRI data in the present study, these regions have previously been associated with WM (*11*, *24*) and are known to have a high density of D1 and D2 receptors (Fig. 1D). Based on suggestions of an inverted u-shape relationship between DA levels and performance, we analyzed the outcome using both linear ($WM = \alpha + \beta_1 BP$) and quadratic ($WM = \alpha + \beta_1 BP + \beta_2 BP^2$) regression models (where α is the intercept and β_1 and β_2 are the regression coefficients).

First we averaged baseline D1 BP across the five cortical ROIs and averaged baseline D2 BP in the four subcortical ROIs, then analyzed the relationship with overall WM capacity before training. There was no significant association for either D1 or D2 BP (D1: linear $r^2 = 0.09$, $P = 0.33$; quadratic $r^2 = 0.34$, $P = 0.12$ for the whole

model and $P = 0.08$ for the second-order term, β_2, suggesting a trend for an inverted-u shape; D2: linear $r^2 = 0.11$, $P = 0.26$; quadratic $r^2 = 0.12$, $P = 0.53$). Next we investigated the effect of training. For each participant, D1 BP change was averaged across the five ROIs. The change in WM capacity could be explained by both the linear model (negative correlation, $P = 0.016$) and the quadratic model ($P = 0.001$). However, the quadratic model $\{WM_2 - WM_1 = [\alpha + \beta_1 BP_2 + \beta_2(BP_2)^2] - [\alpha + \beta_1 BP_1 + \beta_2(BP_1)^2]$, where WM_1 and WM_2 represent WM capacity before and after training, respectively; and BP_1 and BP_2 represent BP before and after training, respectively$\}$ predicted a larger amount of variance ($r^2 = 0.75$) as compared to the linear model ($r^2 = 0.42$; r^2 of change between models = 0.33, $P = 0.005$). The quadratic model was then fitted for each region individually and described the data at a statistically significant level ($P < 0.05$) for the right ventrolateral frontal, right dorsolateral frontal, and both posterior ROIs (Fig. 2, F to J). For D2 BP, the average change across all ROIs was not related to the change in WM capacity (linear model: $r^2 = 0.02$, $P = 0.67$; quadratic model: $r^2 = 0.08$, $P = 0.66$).

These findings show that training-related changes in WM capacity are associated with changes in D1 BP. The binding of [^{11}C]SCH23390 has been shown to be insensitive to the immediate effect of drugs changing endogenous dopamine concentration and may thus serve as an index for the density of available D1 receptors (*25*). Although the relation between performance and dopamine BP is probably nonlinear, our data (Fig. 2, F to J) generally showed that, within the measured range, a negative correlation dominated for all regions, with larger decreases in D1 BP being associated with larger improvements in WM. This is consistent with the finding that low doses of a D1 antagonist enhance the delay activity of prefrontal neurons during the performance of WM tasks (*18*, *19*). An association between a decrease in BP and an increase in WM is also consistent with the negative correlation observed between WM capacity and D1 binding in individuals with schizophrenia (*26*).

The underlying mechanisms responsible for the plasticity of receptor densities are not known. One possibility is that other transmitters influence the trafficking of dopamine receptors; for example, it has been shown that the activation of *N*-methyl-D-aspartate receptors affects dopamine signaling by recruiting D1 receptors from the interior of the cell to the plasma membrane (*27*). Another interpretation is that the changes reflect long-term adjustment of the concentration of D1 receptors in response to a prolonged increase in the level of endogenous dopamine during WM training.

More generally, the present results demonstrate a high level of plasticity of the neuronal system defined by cortical D1 receptors in human volunteers. The findings were specific because the D2 system did not show any relation to WM changes. The training-induced changes emphasize the reciprocal interplay between behavior and the underlying brain biochemistry and should be relevant for studies of neuropsychiatric disorders as well as correlational studies between behavior and biochemical markers.

References and Notes

1. H.-M. Süß, K. Oberauer, W. W. Wittmann, O. Wilhelm, R. Schulze, *Intelligence* **30**, 261 (2002).
2. A. Baddeley, *Nat. Rev. Neurosci.* **4**, 829 (2003).
3. P. S. Goldman-Rakic, *J. Neuropsychiatry Clin. Neurosci.* **6**, 348 (1994).
4. F. X. Castellanos, R. Tannock, *Nat. Rev. Neurosci.* **3**, 617 (2002).
5. K. L. Bopp, P. Verhaeghen, *J. Gerontol. B Psychol. Sci. Soc. Sci.* **60**, 223 (2005).
6. J. M. Swanson *et al.*, *Neuropsychol. Rev.* **17**, 39 (2007).
7. M. Laruelle, *Q. J. Nucl. Med.* **42**, 211 (1998).
8. T. Klingberg, H. Forssberg, H. Westerberg, *J. Clin. Exp. Neuropsychol.* **24**, 781 (2002).
9. P. Olesen, H. Westerberg, T. Klingberg, *Nat. Neurosci.* **7**, 75 (2004).
10. T. Klingberg *et al.*, *J. Am. Acad. Child Adolesc. Psychiatry* **44**, 177 (2005).
11. S. M. Jaeggi, M. Buschkuehl, J. Jonides, W. J. Perrig, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6829 (2008).
12. E. Dahlin, A. S. Neely, A. Larsson, L. Bäckman, L. Nyberg, *Science* **320**, 1510 (2008).
13. T. J. Brozoski, R. M. Brown, H. E. Rosvold, P. S. Goldman, *Science* **205**, 929 (1979).
14. U. Müller, D. Y. von Cramon, S. Pollmann, *J. Neurosci.* **18**, 2720 (1998).
15. T. Sawaguchi, P. S. Goldman-Rakic, *Science* **251**, 947 (1991).
16. T. Sawaguchi, P. S. Goldman-Rakic, *J. Neurophysiol.* **71**, 515 (1994).
17. S. Aalto, A. Brück, M. Laine, K. Någren, J. O. Rinne, *J. Neurosci.* **25**, 2471 (2005).
18. G. V. Williams, P. S. Goldman-Rakic, *Nature* **376**, 572 (1995).
19. S. Vijayraghavan, M. Wang, S. G. Birnbaum, G. V. Williams, A. F. Arnsten, *Nat. Neurosci.* **10**, 376 (2007).
20. M. S. Lidow, G. V. Williams, P. S. Goldman-Rakic, *Trends Pharmacol. Sci.* **19**, 136 (1998).
21. J. X. Cai, A. F. Arnsten, *J. Pharmacol. Exp. Ther.* **283**, 183 (1997).
22. H. Brismar, M. Asghar, R. M. Carey, P. Greengard, A. Aperia, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5573 (1998).
23. M. Soiza-Reilly, M. Fossati, G. R. Ibarra, J. M. Azcurra, *Brain Res.* **1004**, 217 (2004).
24. B. R. Postle, M. D'Esposito, *J. Cogn. Neurosci.* **11**, 585 (1999).
25. N. Guo *et al.*, *Neuropsychopharmacology* **28**, 1703 (2003).
26. A. Abi-Dargham *et al.*, *J. Neurosci.* **22**, 3708 (2002).
27. L. Scott *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1661 (2002).

28. The authors thank C. Halldin, K. Kolaas, G. Lei, J. Macoveanu, and members of the Karolinska Institutet PET center for assistance in the PET measurements. This work was supported by the Swedish Foundation for Strategic Research, Swedish Research Council, VINNOVA (the Swedish Governmental Agency for Innovation Systems), and Knut and Alice Wallenberg Foundation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5915/800/DC1
Methods
Table S1
References

18 September 2008; accepted 10 December 2008
10.1126/science.1166102

Axon Regeneration Requires a Conserved MAP Kinase Pathway

Marc Hammarlund,[1,2]*† Paola Nix,[1]† Linda Hauth,[1] Erik M. Jorgensen,[1,2] Michael Bastiani[1]‡

Regeneration of injured neurons can restore function, but most neurons regenerate poorly or not at all. The failure to regenerate in some cases is due to a lack of activation of cell-intrinsic regeneration pathways. These pathways might be targeted for the development of therapies that can restore neuron function after injury or disease. Here, we show that the DLK-1 mitogen-activated protein (MAP) kinase pathway is essential for regeneration in *Caenorhabditis elegans* motor neurons. Loss of this pathway eliminates regeneration, whereas activating it improves regeneration. Further, these proteins also regulate the later step of growth cone migration. We conclude that after axon injury, activation of this MAP kinase cascade is required to switch the mature neuron from an aplastic state to a state capable of growth.

Severed neurons can regenerate. After axons are cut, neurons can extend a new growth cone from the axon stump and can attempt to regrow a normal process. Most invertebrate neurons are able to regenerate, as are neurons in the mammalian peripheral nervous system. By contrast, neurons in the mammalian central nervous system have limited regenerative capability (*1*). Regeneration is thought to be initiated by signals arising from the injury, including calcium spikes and the retrograde transport and nuclear import of regeneration factors (*2*). These mechanisms lead to increased cyclic adenosine monophosphate (cAMP) levels, local and somatic protein synthesis, and changes in gene transcription that, in turn, promote remodeling of the cytoskeleton and plasma membrane at the site of injury. The ability of specific neurons to regenerate is deter-

[1]Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, UT 84112–0840, USA. [2]Howard Hughes Medical Institute, University of Utah, 257 South 1400 East, Salt Lake City, UT 84112–0840, USA.

*Present address: Department of Genetics and Program in Cellular Neuroscience, Neurodegeneration and Repair, Yale University School of Medicine, 295 Congress Avenue, New Haven, CT 06510, USA.
†These authors contributed equally to this work.
‡To whom correspondence should be addressed. E-mail: bastiani@bioscience.utah.edu

mined in part by the balance between proregeneration signals and cellular pathways that inhibit regeneration. For example, regeneration in the mammalian CNS is inhibited by extrinsic signals from myelin and chondroitin sulfate proteoglycans; these signals activate pathways in the damaged neuron that prevent regrowth (*3*). But CNS regeneration can be achieved even in the presence of inhibitory signals. A conditioning lesion to a peripheral process results in increased regeneration of the CNS branch of dorsal root ganglion neurons, presumably by triggering injury signals that result in an overall increase in regenerative potential (*4*). Thus, intrinsic regeneration signals can influence regenerative success, and these signaling processes represent potential targets for therapies to enhance regeneration.

We identified the mitogen-activated protein (MAP) kinase kinase kinase (MAPKKK) *dlk-1* as essential for regeneration in the course of a large screen for genes required for regeneration. This screen was conducted in a β-spectrin mutant (*unc-70*) background (*5*). Neurons in β-spectrin mutant nematodes break because of mechanical strain induced by locomotion. The γ-aminobutyric acid (GABA)–releasing motor neurons respond to breaks by regenerating toward their targets in the dorsal cord (*6*). Axon guidance during regeneration is imperfect, resulting in axons in mature animals with branching and other abnormalities (Fig. 1). We found that RNA interference of *dlk-1* eliminates regeneration in *unc-70* mutant animals. Neither *unc-70* (*6*) nor *dlk-1* (*7*) is essential for axon outgrowth during development of the GABA neurons (Fig. 1) (*8*). The *unc-70 dlk-1* synthetic phenotype for axon morphology suggests that *dlk-1* may function specifically in regeneration.

To demonstrate that *dlk-1* functions in regeneration independently of *unc-70*, we used laser axotomy to trigger regeneration. The GABA motor neurons can regenerate after laser axotomy (*9*). We used a pulsed 440-nm laser to cut axons (*10*). In larval stage 4 (L4) wild-type animals, 70% of severed axons initiated growth cones within 24 hours after axotomy (Fig. 2 and table S1). But when axons were cut in L4 stage *dlk-1(ju476)* null mutants, growth cones were never observed. These severed neurons appeared healthy after surgery; both the stump of the remaining axon and its cell body showed no decrease in green fluorescent protein (GFP) expression or other signs of injury. Nevertheless, these neurons failed to regenerate. To test whether regeneration was merely delayed, we monitored some severed axons for 5 days; regeneration still was not observed. Thus, *dlk-1* is essential for axon regeneration after spontaneous breaks and after laser surgery but is dispensable for axon outgrowth during development.

Mosaic experiments demonstrate that the DLK-1 protein acts in the damaged cell rather than in the surrounding tissue. To determine whether DLK-1 acts cell-autonomously to promote regeneration, we expressed *dlk-1* under the GABA-specific promoter P*unc-47* in the *dlk-1* null background. Neurons were severed by laser surgery, and regeneration was assayed after 18 to 24 hours. We found that expressing DLK-1 in the GABA neurons restored regeneration to *dlk-1* null mutants (Fig. 2C). Further, mutations that affect DLK-1 levels also affect regeneration (Fig. 2D). In *Caenorhabditis elegans*, DLK-1 levels are negatively regulated by RPM-1 (*7*). Overexpression of RPM-1 reduced regeneration after surgery to levels similar to those of *dlk-1* loss-of-function mutants. Conversely, initiation of regeneration was enhanced in *rpm-1* mutant animals. Initiation of regeneration was also enhanced in animals lacking FSN-1, an F-box protein that functions with RPM-1 to promote DLK-1 degradation (*11*). However, loss of GLO-1 (Rab) or GLO-4 (Rab guanine nucleotide exchange factor), both of which mediate ubiquitin-independent functions of RPM-1 (*11*), did not have strong effects on regeneration. These results confirm that changes in DLK-1 protein abundance can determine regenerative ability.

Although regeneration is age-dependent, *dlk-1* is required at all stages, and overexpression of DLK-1 can rescue some age-dependent decline (Fig. 2E). We analyzed regeneration at various developmental stages. We found that regeneration declines significantly with age, and only a few axons in old adults regenerated (*10*). Despite these differences, *dlk-1* is required for all regeneration, even in very young animals. Because growth cones in the very young animals had not yet reached their targets, the dependence on *dlk-1* is not correlated with target contact or synaptogenesis. Further, *dlk-1* is required for regeneration of both presynaptic and postsynaptic processes (*8*).

To mediate regeneration, DLK-1 is required at the time of injury (Fig. 2F). We expressed the DLK-1 protein at different times using the heat shock promoter *Phsp-16.2*. DLK-1 expression at the time of injury was sufficient for regeneration. Applying heat shock hours before or hours after surgery resulted in less regeneration, and when heat shock was applied either 11 hours before or 48 hours after surgery, little or no regeneration was observed. This effect was independent of age, because surgery in L2 stage larvae failed to elicit regeneration when heat shock was applied 48 hours later. Thus, DLK-1 must function within a short temporal window near the time of injury to mediate regeneration, rather than establishing a permissive state for regeneration during development. These data suggest that DLK-1 signaling must coincide with other proregeneration signals, such as calpain activation (*12*) or cAMP elevation (*13*–*15*), for regeneration to occur.

DLK-1 is required for growth cone formation rather than the earlier step of filopodial extension. We used time-lapse microscopy to monitor morphological changes in axons after surgery. We

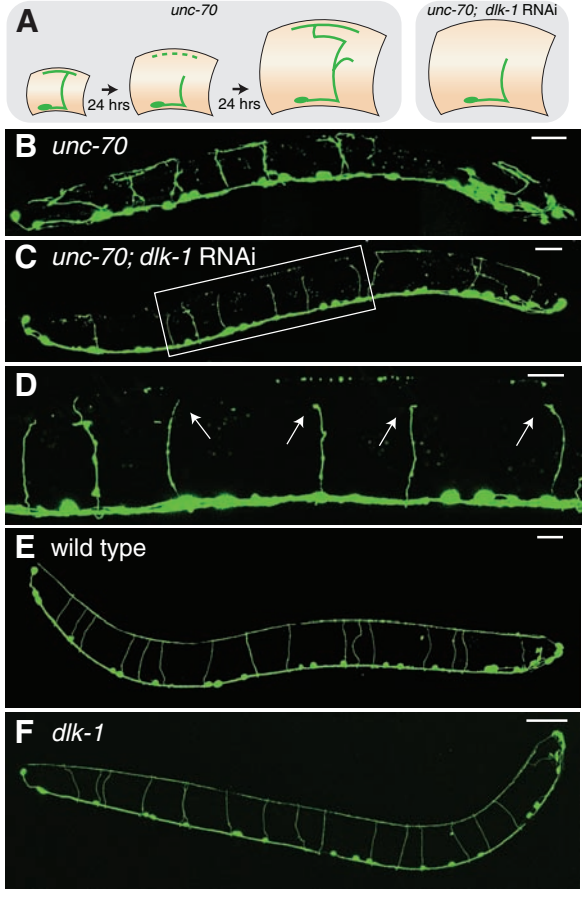


Fig. 1. *dlk-1* is required for axon regeneration in β-spectrin mutant nematodes. (**A**) Cartoon showing development of axon morphology in control β-spectrin mutant (left) and in β-spectrin mutant lacking hypothetical regeneration gene, e.g., *dlk-1* (right). (**B** and **C**) GABA neurons in representative L4 stage β-spectrin mutants (*unc-70*), expressing green fluorescent protein, under control conditions or after *dlk-1* RNA interference. Scale bars, 20 μm. (**D**) High-magnification view of boxed region in (C). Arrows indicate inert axon stumps. Scale bar, 10 μm. (**E** and **F**) GABA neurons in representative L4 stage wild-type and *dlk-1* mutant animals. Scale bars, 20 μm.

found that in wild-type animals, newly severed axons repeatedly extend short, transient filopodia from the axon stump (Fig. 3 and movie S1). The first filopodium appears with an average delay of more than 3 hours. In animals that successfully initiate regeneration, a single filopodium eventually persists and is transformed into a growth cone. Growth cone formation in wild-type animals occurs with an average delay of 7 hours after surgery. In *dlk-1* mutants, transient filopodia appear at approximately the same time and the same rate as in the wild type. However, growth cones were never observed in these mutants. These data demonstrate that DLK-1 is required to transform exploratory filopodia into growth cones.

Increased expression of DLK-1 in wild-type animals accelerates the formation of growth cones and improves migration success. We overexpressed DLK-1 in GABA neurons and found that time of growth cone initiation by axon stumps was advanced relative to the wild type (Fig. 3 and movie S1). Also, more axons initiated growth cones (Fig. 4). In addition to these effects on growth cone initiation, DLK-1 overexpression improved growth cone performance. In wild-type animals, regenerating growth cones often have a branched, dystrophic morphology. Dystrophic growth cones migrate poorly, and most never reach the dorsal nerve cord in 24 hours (Fig. 3). These dystrophic growth cones resemble dystrophic growth cones observed in failed regeneration in the mammalian CNS (*16*). By contrast, regenerating growth cones in neurons that overexpress DLK-1 have a compact shape, similar to growth cones observed during initial axon development (*17*). These compact growth cones were much more likely to reach the dorsal nerve cord. Growth cone migration during regeneration in *C. elegans* shares some genetic requirements with developmental axon guidance, including components of the netrin and slit signaling pathways (*18*) and the ephrin pathway (*10*). It is possible that DLK-1 overexpression improves regeneration by affecting the response of the growth cone to such signals. Thus, DLK-1 acts at two steps of regeneration: It is required for growth cone formation, and it also controls growth cone morphology and behavior.

DLK-1 functions in a MAP kinase signaling cascade that also includes the MAP kinase kinase (MAPKK) MKK-4, and the p38 MAP kinase PMK-3 (*7*). We tested whether this entire MAP kinase signaling module functions in regeneration by examining null mutants in *mkk-4* and *pmk-3*. Like *dlk-1*, neither of these mutants has appreciable defects in axon outgrowth during development. But after axotomy, both mutant strains fail to initiate regeneration (Fig. 4A). These data suggest that MKK-4 and PMK-3 are the downstream targets of DLK-1 for regeneration. Inhibition of p38 also reduces regeneration of cultured vertebrate neurons (*19*), which suggests that the function of p38 MAP kinases in regeneration is conserved. Do other MAP kinase cascades also contribute to regeneration? We tested a sampling of *C. elegans* MAP kinase components and found that mutations in these genes did not eliminate regeneration (Fig. 4B and table S1). Initiation of regeneration was not affected by loss of the MAPKKK *nsy-1* or its target MAPKK *sek-1*. Loss of the MAPKK *jkk-1* also did not affect regeneration. By contrast, loss of the MAPKKK *mlk-1* reduced initiation of regeneration (although some regeneration still occurred), as did loss of its downstream target *mek-1*. MLK-1 and MEK-1 are thought to activate a second *C. elegans* p38 MAP kinase, PMK-1 (*20*), which suggests that multiple p38 family members contribute to regeneration. (Because null mutations in *pmk-1* are lethal, we were unable to test its function directly.) Loss of the MAP kinase *jnk-1* increased initiation of regeneration. Thus, whereas the DLK-1/MKK-4/PMK-3 MAP kinase cascade is required to initiate regeneration, other MAP kinase pathways also regulate this process. Consistent with these data, mutations in *mkk-4* or *pmk-3* did not eliminate the stimulation of regeneration by DLK-1 overexpression, which suggests that cross-talk between MAP kinase modules may contribute to regeneration (Fig. 4C). However, the modest phenotype of other MAP kinase mutants and the inability of DLK-1 overexpression to bypass the requirement for *mkk-4* and *pmk-3* suggest that the DLK-1/MKK-4/PMK-3 module is the major MAP kinase pathway for axon regeneration.

What stimulates DLK-1 function when an axon breaks? In the simplest model, the axon break interrupts trafficking of DLK-1. Local DLK-1 accumulation in the injured neuron then leads to homodimerization and activation, followed by activation of the downstream targets MKK-4 and PMK-3 (Fig. 4D). Alternatively,

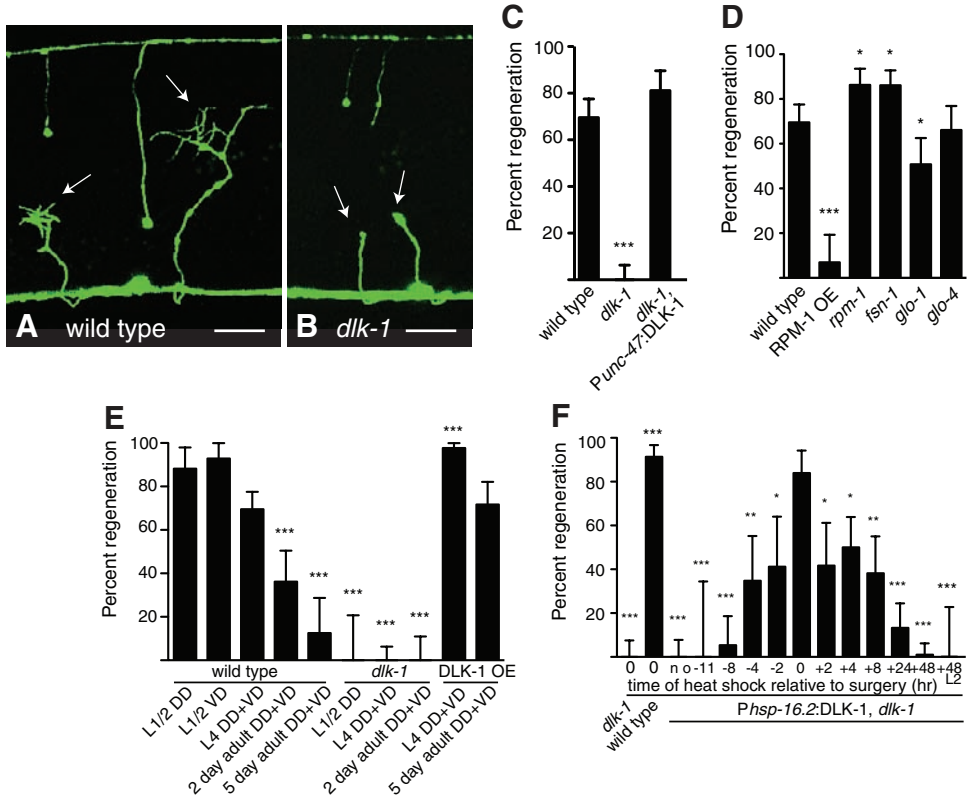
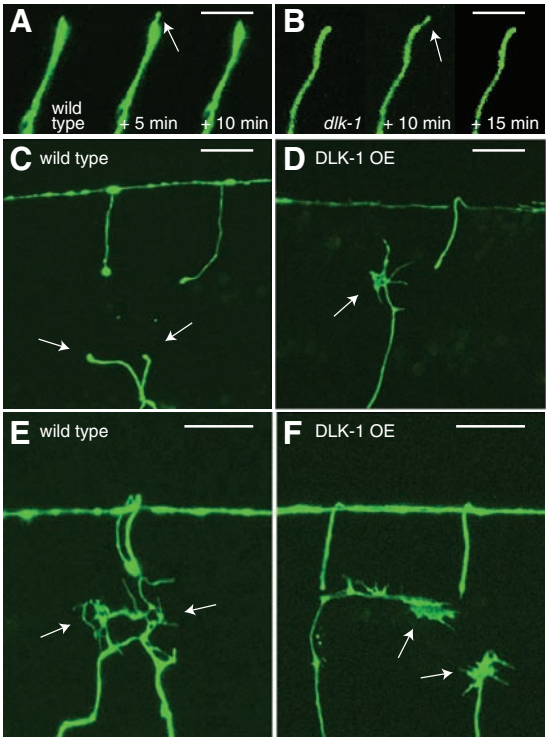


Fig. 2. *dlk-1* is required in severed axons for growth cone initiation. (**A**) Regenerating axons 18 to 20 hours after laser surgery in a wild-type animal. Both severed axons have generated a growth cone (arrows). Scale bar, 10 μm. (**B**) Axons in *dlk-1* mutants fail to generate growth cones 18 to 24 hours after surgery. Scale bar, 10 μm. (**C**) DLK-1 acts cell intrinsically to mediate regeneration. (**D**) RPM-1 controls DLK-1 activity in axon regeneration. (**E**) Regeneration requires *dlk-1* at all ages and overexpression of DLK-1 rescues age-associated decline. DD and VD are motor neurons of different lineages (*8*). (**F**) DLK-1 acts at the time of injury to mediate regeneration. Time of heat shock relative to surgery is indicated in hours. No heat shock is indicated by "no." "L2" indicates surgery at the L2 stage. Axotomy at L4 stage unless indicated otherwise. (C) to (F) Percentage of axons that initiated regeneration and 95% confidence interval (CI). *P < 0.05, **P < 0.01, ***P < 0.001.

specific regulatory mechanisms activate DLK-1 after injury, such as scaffolding proteins like Jip1 (*21*); phosphatases such as PP1, PP2a, and calcineurin (*22*); or regulators of the proteasome (*23*). The strict requirement for the DLK-1 pathway in regeneration suggests that mature neurons have intrinsic barriers to growth that are not present during development. Once axons have reached their target and have begun synaptogenesis, termination of growth signals by mechanisms like RPM-1 may down-regulate growth to allow synapse maturation and to stabilize neuronal architecture. Indeed, mutations in RPM-1 or its homologs cause overgrowth of axons in worms (*24*), aberrant sprouting in *Drosophila* (*25*), and aberrant growth cone initiation on axon shafts in mouse (*26*). We found that *dlk-1* is required for regeneration even in neurons that are actively growing at the time of injury. Further, overexpressing DLK-1 partially prevented the loss of regeneration in old animals (Fig. 2E). Thus, barriers to growth are quickly erected in axons, and post-injury signaling via DLK-1, MKK-4, and PMK-3 is required to drive the neuron back to its pre-lapsarian, embryonic state.

How does the MAP kinase PMK-3 stimulate regeneration? The DLK-1 pathway is first required for growth cone formation about 7 hours after a break occurs—a process likely to be mediated by the polymerization of microtubules. Activated p38 MAP kinase regulates microtubule dynamics (*26*), and microtubule remodeling is required for growth cone initiation during regeneration (*27*). Further,



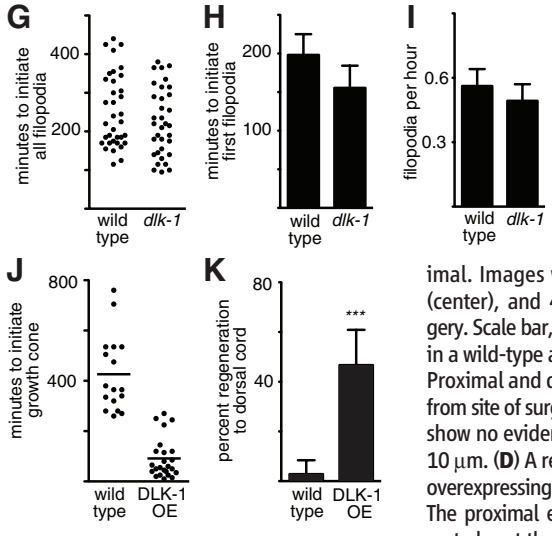


Fig. 3. *dlk-1* controls growth cone initiation and morphology during axon regeneration. (**A**) Transient filopodium in a wild-type animal. Images were taken at 165 (left), 170 (center), and 180 (right) minutes after surgery. Scale bar, 5 μm. (**B**) Transient filopodium in a *dlk-1* mutant animal. Images were taken at 475 (left), 480 (center), and 490 (right) minutes after surgery. Scale bar, 5 μm. (**C**) Representative axons in a wild-type animal 120 min after axotomy. Proximal and distal ends have retracted away from site of surgery, but proximal ends (arrows) show no evidence of regeneration. Scale bar, 10 μm. (**D**) A representative axon in an animal overexpressing DLK-1 120 min after axotomy. The proximal end (arrow) has already regenerated past the retracted distal end. Scale bar, 10 μm. (**E**) Representative growth cones in a wild-type animal. Although these axons successfully initiated regeneration, the growth cones (arrows) have a dystrophic morphology. Scale bar, 10 μm. (**F**) Representative growth cones in an animal overexpressing DLK-1 under the *unc-47* promoter. These growth cones (arrows) have a compact morphology similar to growth cones observed during development. Scale bar, 10 μm. (**G**) Distribution of all times of filopodia initiation in wild type and *dlk-1*. Each dot represents a filopodium. (**H**) Time of first filopodium initiation in wild type and *dlk-1*. Means ± SEM. (**I**) Rate of filopodia initiation in wild type and *dlk-1*. Means ± SEM. (**J**) Time to initiate regeneration after surgery in wild type and *dlk-1* overexpressing (OE) animals. Initiation is defined as the appearance of the filopodia that becomes a growth cone. Each dot represents a single axon. (**K**) Percentage of wild-type and *dlk-1* overexpressing (OE) regenerating axons that reached the dorsal cord after 18 to 24 hours. Error bars indicate 95% CI.

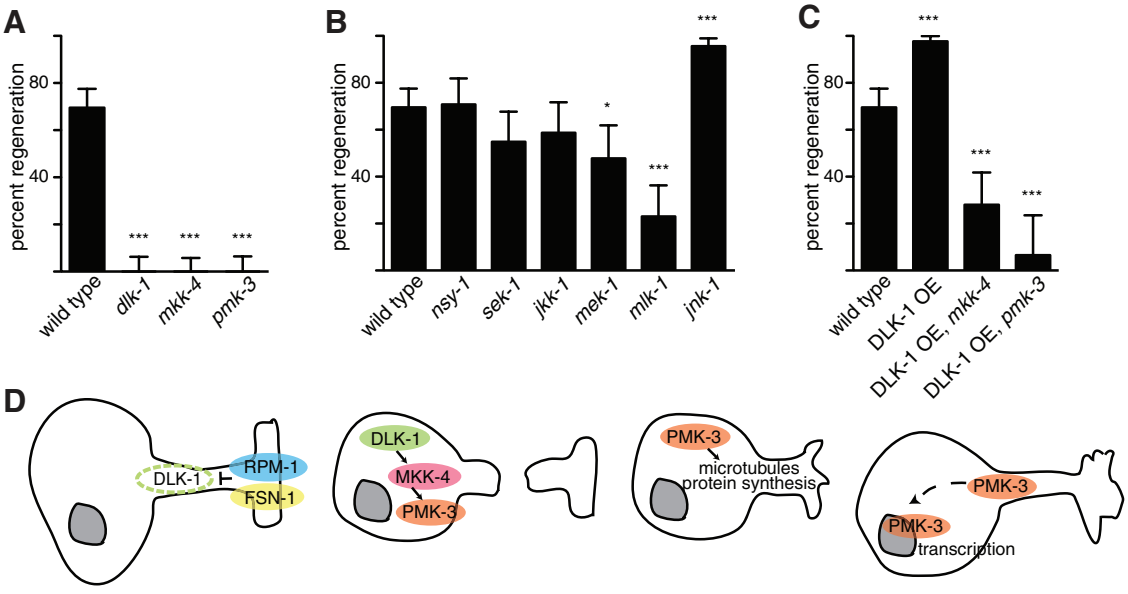


Fig. 4. MAP kinase signaling is required for axon regeneration. (**A**) Regeneration is eliminated by mutations in the DLK-1/MKK-4/PMK-3 MAP kinase module. (**B**) Other MAP kinase elements contribute to regeneration, but are not essential. (**C**) Activated DLK-1 has targets in addition to MKK-4 and PMK-3. (**D**) Model for function of MAP kinase signaling during axon regeneration. (A) to (C) Error bars indicate 95% CI.

defects in microtubule dynamics contribute to the axon outgrowth phenotype of *Phr1* mutant mice (*26*). Activated p38 may also control other targets that facilitate axon regeneration. p38 regulates local protein synthesis (*28*), which is required for regeneration (*19*). p38 is also likely to have functions in the nucleus, because it contributes to injury-induced changes in gene transcription (*29*). Activated p38 may reach the nucleus by retrograde transport. Retrograde transport in general is critical for regeneration (*30*), and transport of activated MAP kinases from axons to the cell body following axotomy has been observed in *Aplysia* sensory neurons (*31*) and in rodent sciatic nerves (*32*, *33*). Thus, regeneration may require activated PMK-3/p38 at the site of the break to regulate microtubule stability and protein expression and also may require PMK-3 to traffic to the nucleus to regulate gene transcription (Fig. 4D). The DLK-1 signaling pathway thus provides a critical link between axon injury and the process of regeneration.

References and Notes

1. L. C. Case, M. Tessier-Lavigne, *Curr. Biol.* **15**, R749 (2005).
2. F. Rossi, S. Gianola, L. Corvetti, *Prog. Neurobiol.* **81**, 1 (2007).
3. B. P. Liu, W. B. Cafferty, S. O. Budel, S. M. Strittmatter, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 1593 (2006).
4. S. Neumann, C. J. Woolf, *Neuron* **23**, 83 (1999).
5. M. Hammarlund, W. S. Davis, E. M. Jorgensen, *J. Cell Biol.* **149**, 931 (2000).
6. M. Hammarlund, E. M. Jorgensen, M. J. Bastiani, *J. Cell Biol.* **176**, 269 (2007).
7. K. Nakata *et al.*, *Cell* **120**, 407 (2005).
8. Materials and methods are available as supporting material on *Science* Online.
9. M. F. Yanik *et al.*, *Nature* **432**, 822 (2004).
10. Z. Wu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15132 (2007).
11. B. Grill *et al.*, *Neuron* **55**, 587 (2007).
12. D. Gitler, M. E. Spira, *J. Neurobiol.* **52**, 267 (2002).
13. S. Chierzi, G. M. Ratto, P. Verma, J. W. Fawcett, *Eur. J. Neurosci.* **21**, 2051 (2005).
14. S. Neumann, F. Bradke, M. Tessier-Lavigne, A. I. Basbaum, *Neuron* **34**, 885 (2002).
15. J. Qiu *et al.*, *Neuron* **34**, 895 (2002).
16. J. Silver, J. H. Miller, *Nat. Rev. Neurosci.* **5**, 146 (2004).
17. K. M. Knobel, E. M. Jorgensen, M. J. Bastiani, *Development* **126**, 4489 (1999).
18. C. V. Gabel, F. Antonie, C. F. Chuang, A. D. Samuel, C. Chang, *Development* **135**, 1129 (2008).
19. P. Verma *et al.*, *J. Neurosci.* **25**, 331 (2005).
20. J. J. Ewbank, *WormBook* (23 January 2006); doi/10.1895/wormbook.1.83.1, www.wormbook.org.
21. D. Nihalani, D. Meyer, S. Pajni, L. B. Holzman, *EMBO J.* **20**, 3447 (2001).
22. M. Mata, S. E. Merritt, G. Fan, G. G. Yu, L. B. Holzman, *J. Biol. Chem.* **271**, 16888 (1996).
23. A. Daviau *et al.*, *J. Biol. Chem.* **281**, 31467 (2006).
24. A. M. Schaefer, G. D. Hadwiger, M. L. Nonet, *Neuron* **26**, 345 (2000).
25. C. A. Collins, Y. P. Wairkar, S. L. Johnson, A. DiAntonio, *Neuron* **51**, 57 (2006).
26. J. W. Lewcock, N. Genoud, K. Lettieri, S. L. Pfaff, *Neuron* **56**, 604 (2007).
27. H. Erez *et al.*, *J. Cell Biol.* **176**, 497 (2007).
28. D. S. Campbell, C. E. Holt, *Neuron* **37**, 939 (2003).
29. H. Zrouri, C. Le Goascogne, W. W. Li, M. Pierre, F. Courtin, *Eur. J. Neurosci.* **20**, 1811 (2004).
30. S. Hanz, M. Fainzilber, *J. Neurochem.* **99**, 13 (2006).
31. Y. J. Sung, M. Povelones, R. T. Ambron, *J. Neurobiol.* **47**, 67 (2001).
32. E. Perlson *et al.*, *Neuron* **45**, 715 (2005).
33. A. J. Reynolds, I. A. Hendry, S. E. Bartlett, *Neuroscience* **105**, 761 (2001).

34. This work is dedicated to the memory of Craig H. Neilsen. We thank C.-B. Chen, D. Gard, W. Davis, J. White, and K. Schuske for helpful discussions. Supported by the Craig H. Neilsen Foundation, the McKnight Endowment Fund for Neuroscience, and NIH 1R21NS060275 to M.B. and NIH NS034307 to E.M.J.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1165527/DC1
Materials and Methods
Table S1
References
Movie S1

4 September 2008; accepted 10 December 2008
Published online 22 January 2009;
10.1126/science.1165527
Include this information when citing this paper.

Solenoid Valve Manifold

An injection manifold mount system enables the precise injection of a fluid into a flow stream, so any ported Lee VHS (very high speed) microdispense solenoid valve can be mounted so that the outlet port is in close proximity to the flow stream. This provides the dual benefit of minimal captive capillary volume and increased injected volume repeatability. Lee VHS two-way solenoid valves are suitable for applications requiring precise liquid control down to nanoliter dispense volumes. The new mount system is suitable for laboratory, biomedical, and chemistry applications, and especially for flow injection analyzers.

Lee Products
For information +44-(0)-1753-886664
www.leeproducts.co.uk

Immunogenecity Package

The Biacore T100 immunogenecity software package supports screening, confirmation, and characterization of antidrug antibodies. The software for the Biacore T100 label-free protein interaction analysis system supports screening, confirmation, and characterization of antidrug antibodies during preclinical and clinical development. The software addresses the problem of drug interference by enabling measurement of antidrug antibodies in the presence of excess amounts of drug. It also includes recommendations for development of immunogenecity assays to reduce development time.

Biacore/GE Healthcare
For information +44-(0)20-7457-2020
www.biacore.com

Downflow Workstations

The DWS range of Downflow Workstations provide operator safety for carrying out routine work. The units operate at low noise levels. Because they recirculate air, they do not exhaust expensive conditioned or heated air into the atmosphere. They are designed to provide a small bench-mounted unit with unrestricted access for operations that are difficult to perform in a conventional fume hood. The downflow action takes the contaminated air away from the operator. An alarm sounds when the airflow falls to an unacceptable level. The main filter can be chosen from 14 different types of carbon, which include specialty media for vapors of organics, solvents, acids, mercury, and formaldehyde. High efficiency particulate air filters are also available.

Air Science USA
For information 800-306-0656
www.air-science.com

Multiplex Immunoassays

A new line of MultiBead multiplex immunoassays can simultaneously quantify the levels of up to 22 analytes in a single sample and in various matrices, including serum, plasma, urine, culture supernatants, and cell lysates. MultiBead assays do not require specialized instruments, but work on commercially available dual-laser flow cytometers, making access to instrumentation convenient and economical. Proprietary data analysis software is available with the products.

Assay Designs
For 800-833-8651
www.assaydesigns.com

Animal-Free Culture Supplements

The first product under the CellPrime brand, Transferrin recombinant supplement is designed to reduce risk and ease regulatory concerns for biopharmaceutical manufacturers by improving the consistency and productivity of industrial cell culture processes. Recombumin is a yeast-derived recombinant human albumin that has a structure identical to human serum albumin. It is the first animal-free recombinant human albumin approved for use in the manufacture of human therapeutics. LongR3 IGF-1 is an analog of human insulin-like growth factor 1 designed for use in industrial cell culture, providing an animal-free alternative to recombinant insulin.

Novozymes Biopharma
For information +44-115-9551209
www.biopharma.novozymes.com

Gene Expression Profiling

The SOLiD System high throughput genomic analysis platform can generate genomewide expression profiling data at the single-cell level. Gene expression signatures at the single-cell level have the potential to reveal details about cell fate and can potentially accelerate the discovery of biomarkers for disease. The generation of comprehensive profiles from single stem cells is expected to reveal clues about molecular variation between genetically identical stem cells and the underlying molecular mechanisms that trigger pluripotent stem cells to differentiate into specific cell types, including cancer cells. Such detailed characterizations will give researchers a better understanding of how these cells can be used in regenerative therapies for damaged cells and organs.

Applied Biosystems
For information 800-955-6288
www.appliedbiosystems.com

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.