8 February 2008 | $10

Science

Cities

AAAS

Volume 319, Issue 5864

COVER

The Ginza area of Tokyo in 2006. By 2030 the number of urban dwellers will have exploded to 4.8 billion people, roughly 60 percent of the projected world population, whereas only 13 percent lived in cities in 1900. The special section beginning on page 739 includes News stories, Reviews, and Perspectives that explore the ramifications of urban transformation.

Photo: Getty Images

DEPARTMENTS

EDITORIAL

SPECIAL SECTION

Cities

INTRODUCTION

NEWS

REVIEWS

PERSPECTIVES

>> Editorial p. 697; for related online material, see p. 691 or go to www.sciencemag.org/cities

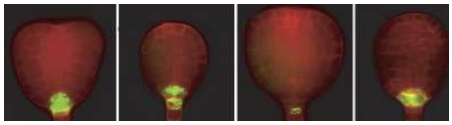
NEWS OF THE WEEK

NEWS FOCUS

CONTENTS continued >>

Published by AAAS

Science



SCIENCE EXPRESS

www.sciencexpress.org

CLIMATE CHANGE

Land Clearing and the Biofuel Carbon Debt
J. Fargione, J. Hill, D. Tilman, S. Polasky, P. Hawthorne
10.1126/science.1152747

Use of U.S. Croplands for Biofuels Increases Greenhouse Gases Through Emissions from Land Use Change
T. Searchinger et al.
Converting forests and grasslands to biofuel crop production results in a net carbon flux to the atmosphere for decades despite any displacement of fossil fuel use.
10.1126/science.1151861

PLANT SCIENCE

TOPLESS Mediates Auxin-Dependent Transcriptional Repression During *Arabidopsis* Embryogenesis
H. Szemenyei, M. Hannon, J. A. Long
A transcriptional co-repressor is part of the protein complex that inhibits developmental gene activation in *Arabidopsis* until the growth hormone auxin triggers its degradation.
10.1126/science.1151461

NEUROSCIENCE

Synaptic Protein Degradation Underlies Destabilization of Retrieved Fear Memory
S.-H. Lee et al.
Upon recollection, mouse memories of fearful situations become labile, as postsynaptic proteins are degraded by proteosomes and are then reconsolidated via protein synthesis.
10.1126/science.1150541

LETTERS

BOOKS *ET AL.*

POLICY FORUM

PERSPECTIVES

TECHNICAL COMMENT ABSTRACTS

MATHEMATICS

BREVIA

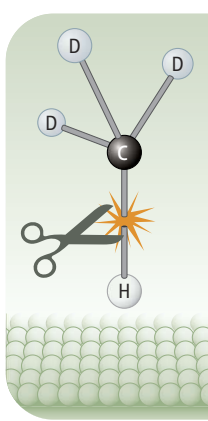
PHYSIOLOGY

RESEARCH ARTICLES

IMMUNOLOGY

REPORTS

PHYSICS



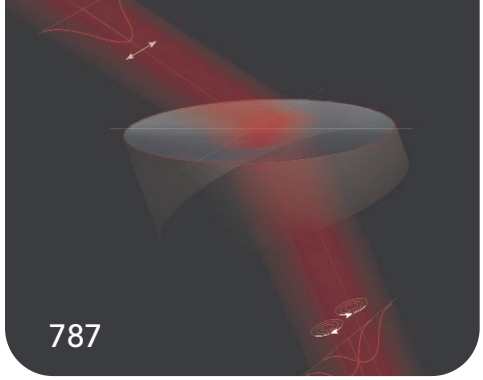
736

CONTENTS continued >>

Published by AAAS

Science

REPORTS *CONTINUED...*



AAAS
ADVANCING SCIENCE, SERVING SOCIETY

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): $144 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $770; 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. SCIENCE is printed on 30 percent post-consumer recycled paper. **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.

CONTENTS continued >>

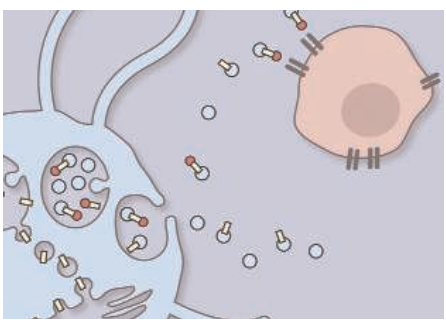
SCIENCE**NOW**

www.sciencenow.org DAILY NEWS COVERAGE

Team Uncovers New Evidence of Recent Human Evolution
Adaptation to disparate environments resulted in mutations related to obesity and diabetes.

Don't It Make Your Brown Eyes Blue?
Researchers locate genetic change that leads to baby blues, and it's not where they expected.

Move Over Beavers, Here Come Salmon
The big fish don't just swim upstream—they shape the stream.



Exosomes spread inflammatory signals.

SCIENCE **SIGNALING**

www.stke.org THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

PERSPECTIVE: Novel Roles for the NF-κB Signaling Pathway in Regulating Neuronal Function
M. C. Boersma and M. K. Meffert
Components of the NF-κB pathway may use multiple mechanisms to influence synaptic plasticity, learning, and memory.

PERSPECTIVE: Exosomes Secreted by Bacterially Infected Macrophages Are Proinflammatory
H. C. O'Neill and B. J. C. Quah
The release of bacterial components in vesicles secreted by infected macrophages helps promote inflammation.

Mentoring and your career.

SCIENCE **CAREERS**

www.sciencecareers.org CAREER RESOURCES FOR SCIENTISTS

Special Feature: Mentoring
E. Pain
What makes mentoring relationships successful?

A Gift That Keeps On Giving
S. Webb
An industry mentor helped physicist Joan Hoffmann navigate graduate school and launch her career.

Mentoring Opposites
C. Wald
A mentor and student turned their differences into strengths as they became scientific collaborators.

From the Archives: The Commandments of Cover Letter Creation
P. Fiske
A good cover letter highlights your qualifications and guides readers through the most important parts of your work history.

SCIENCE **ONLINE FEATURE**

VIDEO: Cities

An accompaniment to this week's special section exploring the benefits and challenges of urbanization.

www.sciencemag.org/cities

SCIENCE**PODCAST**

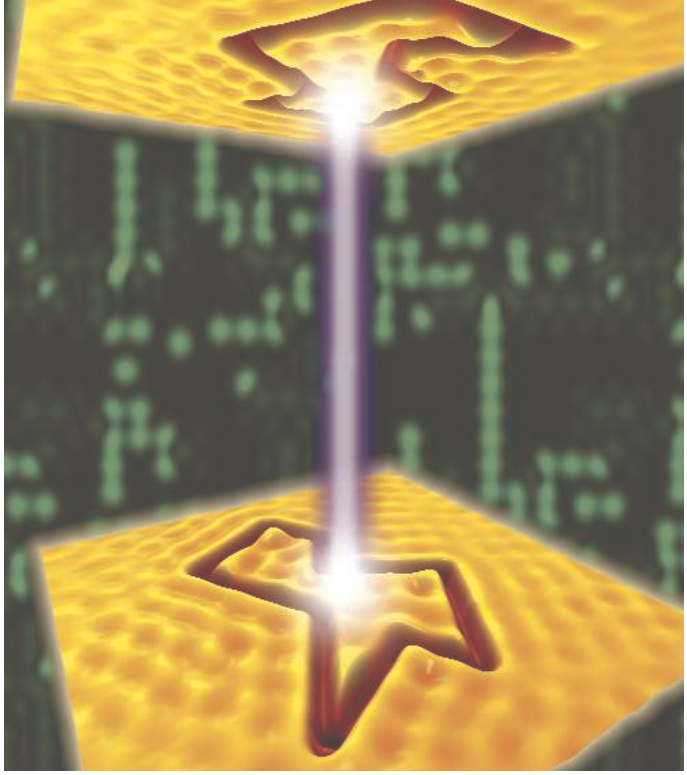
Download the 8 February *Science* Podcast to hear about greenhouse emissions from biofuel-dedicated land, the 2009 U.S. science budget, good mentoring relationships, reproducing in cities, and more.

www.sciencemag.org/about/podcast.dtl

CREDITS: (*SCIENCE CAREERS*) JOHN WIGHAM/CREATIVE COMMONS

Separate individual or institutional subscriptions to these products may be required for full-text access.

EDITED BY STELLA HURTLEY AND PHIL SZUROMI



<< Quantum Phase via Geometry

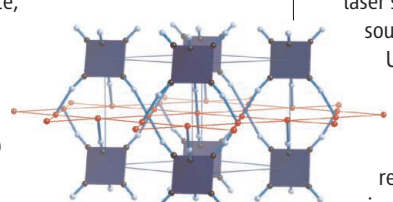
The phase of a wave function collapses when measurements are made, so additional information is needed to determine phase, and several methods have been developed based on interference with reference waves. **Moon *et al.*** (p. 782) describe a non-interferometric approach to phase-determination based on the isospectrality, which describes pairs of simple polygonal shapes that have the same frequency response—that is, if these shapes were drumheads, they could sound the same and be indistinguishable. The authors used scanning tunneling microscopy to position CO molecules on the Cu(111) surface at cryogenic temperatures to bound isospectral shapes. Despite the imperfect nature of this boundary, the spectral fingerprints of the two-dimensional electronic states in the terahertz range were the same within experimental error. The authors then used this property to extract the wave function phase. Phase extraction should be possible in two-dimensional quantum systems provided that the boundary shapes can be constructed.

Spin Hall Effect of Light

Hall effects manifest as the transverse movements of carriers of electronic current in the presence of an external field, and recent work has concentrated on the spin Hall effect, in which the effects depend on the spin of the electron and not just its charge. **Hosten and Kwiat** (p. 787, published online 10 January; see the Perspective by **Resch**) now report on the observation of an optical version of the spin Hall effect that developed a sensitive metrological technique capable of detecting displacements on the angstrom scale. When light refracts at an air-glass interface, there is an additional displacement of the light that depends on polarization. In this optical system, the polarization of the light interacts with a refractive index gradient in a manner analogous to how electronic spins are affected by electric fields.

Moving with the Heat

Most materials have a positive coefficient of thermal expansion—they expand when heated—but there are exceptions, such as cubic zirconium tungstate, which will contract over a wide temperature range. Silver(I) hexacyanocobaltate(III) is a framework material that has highly underconstrained Co–CN–Ag–NC–Co linkages. **Goodwin *et al.*** (p. 794) find that over a wide temperature range, this material can either contract or expand along orthogonal lattice directions with a coefficient an order of magnitude greater than that of most materials. They attribute these large movements to the framework flexing like a hinged lattice, of the sort commonly seen in garden fencing.



CREDITS (TOP TO BOTTOM): HARI MANOHARAN/STANFORD UNIVERSITY; GOODWIN *ET AL.*

Every Step You Take

As we walk, we expend energy not only in pushing off with our planted leg but also when our other leg decelerates as it makes contact with the ground. **Donelan *et al.*** (p. 807) have developed a mechanical device to harvest some of the expended energy in this deceleration step in much the same fashion as hybrid automobiles utilize regenerative braking. The device is light, fastens at the knee, and produces about 5 watts of power, potentially enough power to charge portable medical devices.

Bonds, Shaken and Sliced

Soon after the development of narrow-frequency laser sources, chemists attempted to use these sources to excite specific chemical bonds. Unfortunately, the deposited energy usually spread around the rest of the molecular framework too rapidly for the chosen bond to break. More recently, studies have shown that if collisions with other molecules or catalytic surfaces occur soon enough after vibrational excitation, reaction efficiencies can be selectively enhanced. **Killelea *et al.*** (p. 790; see the Perspective by **Mullins and Sitz**) now take this approach a step further to show that by exciting the C–H stretch in the CHD_3 isotopomer of methane just prior to collision with a nickel surface, they can achieve a 30:1 ratio of C-H to C-D bond cleavage, relative to a 1:3 ratio in a thermally equilibrated sample. Quantification of this selective scission required a technically demanding mass-resolved detection scheme of thermally desorbed products.

Sounding Out Earth's Core

Earth's solid inner core is predominantly a phase of iron at high pressures. One important clue for determining the properties of the core is that sound waves passing through Earth's solid inner core propagate fastest along the north-south direction, which suggests that there is a preferred alignment of iron crystals. **Belonoshko *et al.*** (p. 797) present numerical calculations which show that the body-centered cubic form of iron is strongly anisotropic to seismic waves and can match the observed 12% anisotropy, whereas the hexagonal close-packed form, previously thought to make up the inner core, is not.

Mixed-Up Microflora

The relationship between an animal host and the complex mixture of microbes it carries in its gut is a delicate one, and the exact role the host immune system plays in maintaining commensal homeostasis remains unclear. **Ryu *et al.*** (p. 777, see the Perspective by **Silverman and Paquette;** published online 24 January) examined the expression of antimicrobial proteins in

Continued on page 695

Published by AAAS

Continued from page 693

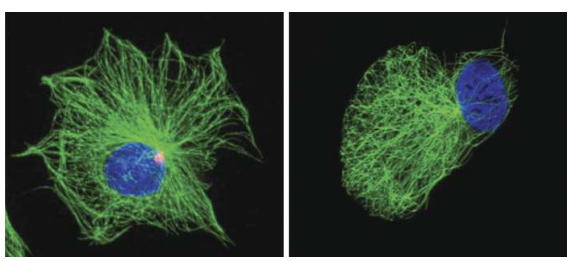
the gut of the fruit fly, *Drosophila melanogaster*. Although the key immune transcriptional regulator nuclear factor kappa B (NF-κB) was chronically activated in the flies by indigenous gut microflora, only a subset of NF-κB–regulated antimicrobial genes was actually expressed because of transcriptional repression exerted by the intestinal homeobox gene *Caudal*. Disruption of Caudal expression resulted in the expression of a different subset of antimicrobial peptides, as well as a dramatic change in the composition of the intestinal microflora that led to the apoptosis of intestinal epithelial cells and loss of host viability.

What a Tangled Food Web We Weave

Both direct and indirect effects can mediate bottom-up influences on the diversity and complexity of a food web. **Bukovinszky *et al.*** (p. 804) compared the effects of two related plants (domestic Brussels sprouts and feral *Brassica*) on their aphid herbivores, the aphids' parasitoid wasps, and the wasps' secondary parasitoids to examine how resource quality affects food web structure and complexity. In this multilevel trophic system, differences in the resource base cascaded through the system, via an array of direct and indirect effects, and led to substantial differences in the structure and complexity of the resulting food webs.

A Growing Role for Centrosomes

Genetic analyses of individuals with extreme forms of short stature can provide insights into the biological mechanisms regulating human growth. **Rauch *et al.*** (p. 816, published online 3 January; see the Perspective by **Delaval and Doxsey**) have identified a mutant gene responsible for microcephalic osteodysplastic primordial dwarfism type II (MOPD II). Adults with this rare inherited condition reach an average height of 100 centimeters, and although their brain is comparable in size to that of a three-month-old baby, they are of near-normal intelligence. The culprit gene is *PCNT*, which encodes pericentrin, a centrosomal protein implicated in mitotic spindle anchoring and chromosome separation during cell division. Although the precise mechanisms by which the cellular phenotype produces the size phenotype remains to be determined, it is intriguing that other inherited forms of microcephaly (disorders characterized by small brain size) have likewise been genetically linked to centrosomal and mitotic spindle genes.



CREDIT: RAUCH *ET AL.*

Tipping the Balance in T Cell Decisions

In the thymus, key developmental decisions are made that have far-reaching consequences for the immune system, perhaps most notable the commitment of thymocytes to becoming CD4 (helper) or CD8 (cytotoxic) T cells. The transcriptional factor Th-POK commits thymocytes that still express both CD4 and CD8 co-receptors to becoming CD4 T cells, but how is the alternate developmental option generated? **Setoguchi *et al.*** (p. 822) find that cells that would otherwise be destined to become $CD8^+$ T cells can be redirected to the CD4 lineage by loss of members of another transcription factor family, Runx. It seems that under normal circumstances, the Runx complex represses Th-POK expression and allows CD8 T cells to emerge.

The Heme Balancing Act

Heme, a component of several hemoproteins, is required in aerobic cells for oxygen transport and storage (hemoglobin and myoglobin), electron transfer and drug metabolism (cytochromes), and signal transduction (nitric oxide synthases). However, free heme is toxic, so its intracellular concentration must be carefully regulated. **Keel *et al.*** (p. 825) have generated mice lacking the heme export protein FLVCR (feline leukemia virus, subgroup C, receptor), and show that this factor is required for terminal red blood cell development. They suggest that heme toxicity may be a common pathophysiology in some erythroid disorders where free-heme-balance is perturbed. Additionally, FLVCR functions in the recycling of heme-iron from senescent red cells, and heme-iron trafficking via FLVCR is involved in systemic iron homeostasis.

Published by AAAS

David Baltimore is president of the AAAS and Robert A. Millikan Professor of Biology at the California Institute of Technology. E-mail: baltimo@caltech.edu.

Science for the Globe

SCIENCE AND TECHNOLOGY (S&T) CAN BE VIEWED FROM MANY ANGLES. AT THIS YEAR'S annual meeting of the American Association for the Advancement of Science (AAAS), which starts on 14 February in Boston, we look at them from a global perspective. Our ever-shrinking, flattening world invites a global focus on almost any issue. But in the United States, national competitiveness is often the key concern driving S&T policy, whereas the global perspective, which comes naturally to many scientists, is given short shrift. Indeed, a host of topics come to the fore when S&T are viewed globally, including international cooperation on big science projects, economic development, worldwide treatment and prevention of infectious disease, responses to climate change, and mitigation of global warming. In planning the annual meeting, it was evident that all elements had international dimensions, and the meeting reflects this reality.

Appropriately, this issue of *Science* focuses on the cities, which hold so much of the world's population at high density that they pose many of the most pressing problems of the world. Urbanization generates air pollution, first evident years ago in the United States but now a worldwide plague. Take Beijing, where running Olympic events this year will depend on temporary industrial and transportation shutdowns. This quick fix may work, but what's needed for long-term clean air in cities is an application of technology-based rules and processes, such as industrial emissions standards. Here, the United States leads. My adopted hometown of Pasadena, California, recovered its dramatic view of the San Gabriel Mountains because Los Angeles took the air-quality problem seriously. International technology-sharing can help cities in less developed nations solve their own air-quality problems.

Cleaning up emissions, whether particulate pollution or greenhouse gases, is one of the world's grand challenges for S&T. It has to be faced head on because solutions will be costly. This challenge to our inventiveness should be seen as an opportunity to create new industries. Countries that accept the challenge will reap huge rewards, as other nations recognize that they must find less-polluting ways to generate energy.

That recognition is coming fast, and the time to respond is now. The United States ought to take the lead here, because it is responsible for much of the world's burden of greenhouse gases and because it has such an effective engine of innovation in its universities and industries. For U.S. companies, there's a market incentive: the opportunity to be first movers in an international competition to solve a major environmental problem.

The AAAS annual meeting will explore many of these issues and, as a biologist, I am gratified that the program will also emphasize health. Bringing health benefits to the less-developed world has become a philanthropic priority for wealthy Americans such as the Gates family and for the governments of many developed countries. The concentration of resources for treating and preventing AIDS is heartening. However, as HIV researcher Daniel Halperin recently emphasized in the *New York Times*, donors could take a more balanced approach to improving health in the world's poor countries. Cleaning the air and water, treating sewage, dealing with diarrhea, and mounting immunization programs are all needed. So is improving the availability of health practitioners to treat non–HIV-related disease, which cannot be ignored even though AIDS is so widespread.

When I entered science in the early 1960s, my elder colleagues, particularly physicists who had worked on the atomic bomb, emphasized that S&T were not the province of one country but resources for the world. After World War II, U.S. politicians often took a more domestic focus, but the international activities of scientists such as the Pugwash Conferences reminded the nation that science and scientific concerns transcend national borders. The tensions between national security and science remain unresolved, and the terrorist challenges we now face have produced a greater emphasis on national concerns. I hope that this annual meeting will provide a counterforce, reminding us that in a shrinking world, the problems of any nation are the problems of every nation.

– David Baltimore

10.1126/science.1155011

GEOLOGY

The Ups and Downs of Stress

The Hawaiian Islands have formed sequentially as the Pacific Ocean crust has moved over a locus of melting in the mantle. As each island grows, the huge weight of cooled magma, a pile extending many kilometers above the ocean floor, bends the ocean crust downward. Two related large (M_w = 6.0 and 6.7) earthquakes struck the island of Hawaii on 15 October 2006 and have helped reveal important aspects of this process. Both earthquakes occurred in the mantle. One was particularly deep, 39 km below the surface, and implied local extension; the other was shallower, at a depth of about 19 km, and suggested lateral compression. Through finite-element modeling, McGovern shows that the different mechanisms reflect modification of the broad bending process by the different strengths of the lower crust and mantle, producing compression at depths shallower than about 32 km and extension below, with strain focused near the depth of the deeper quake. In addition, compression at a depth of 19 km would tend to restrict the ascent of magmas, consistent with the notion that the crust is being underplated by cooled magmas at this depth. — BH

Geophys. Res. Lett. **34**, L23305 (2007).

PLANT SCIENCE

Pared to the Essentials

The RNA world is alive and well—deeply embedded in plants. Plant viruses, traveling messenger RNAs, gene-silencing RNAs, these and more are the various guises of RNA in plants. For infectious RNAs, structural motifs—sequences that form hairpins and loops and bulges—are necessary both for RNA replication and for RNA trafficking between cells. Using a genome-wide mutational analysis of potato spindle tuber viroid (PSTVd), which replicates in the host cell's nucleus, Zhong *et al*. have investigated what these two processes share in terms of structural motifs. Potatoes produced by plants infected by this viroid are smaller and lumpier than usual, and the viroids move through the leaf cells, into the phloem, and then on to distant parts. The PSTVd RNA adopts a rod-shaped structure, and all of the loops were essential for fully successful replication and trafficking. Loops toward one end of the rod and in the middle were critical for replication; damage to loop 11 produced viroids able to travel well but not so apt to replicate; finally, one loop in its native state actually seemed to repress replication. Comparisons with other types of viroid RNAs hint at a conservation of structure-function relationships for some loops. — PJH

Plant Cell **20**, 10.1105/tpc.107.056606 (2008).

APPLIED PHYSICS

Light amid Disorder

In semiconductors, the concept of an energy gap that separates conducting electrons from localized valence states is fundamental to understanding the materials' optical and electronic properties. Researchers discovered in the 1980s that photons can behave in a similar way: Optical materials fabricated with just the right periodic structures exhibit energy (or frequency) regions where light passes through and other energy zones where transmission of light is blocked. Just as semiconductor band gaps lead to a wide range of useful technological properties, photonic band gaps can do the same for optical materials. Researchers have assumed that in order to produce the photonic band gaps, high-quality crystalline materials are required. Edagawa *et al*. present computational results showing that amorphous diamond without lattice periodicity can also exhibit strong photonic band gaps. The results challenge the traditional view that photonic band gaps are strictly a consequence of Bragg reflection and interference in which electromagnetic waves are scattered from various planes formed by a periodic atomic lattice. Thus, a range of photonic band gap systems could potentially be synthesized from materials such as polymers, proteins, and colloids that lend themselves naturally to amorphous structures. — DV

Phys. Rev. Lett. **100**, 13901 (2008).

PHYSICS

Profiles in Charge

The availability of high-power lasers emitting intense pulses over femtosecond and picosecond time scales enables the study of high-field processes such as photodissociation and photoexcition of atoms and molecules in the laboratory. Such processes are relevant across a range of disciplines, from the study of photoinduced chemical reactions in the atmosphere to the more fundamental probing of the electronic excitations in atoms. When an intense laser pulse hits a cloud of atoms or molecules, the intensity profile of the laser pulse will produce a specific distribution of ions. After exciting a cloud of Xe atoms with intense laser pulses, Strohaber and Uiterwaal implement a time-of-flight technique that samples the pulse focal region with micrometer resolution, allowing the distribution of ions to be mapped out in three dimensions. On the flip side, the profile of the ion distributions can be used as an intensity sensor to aid the characterization and optimization of intense laser pulses. — ISO

Xe ion profile.

Phys. Rev. Lett. **100**, 23002 (2008).

Continued on page 701

CREDITS (TOP TO BOTTOM): NOAA COASTAL SERVICES CENTER; STROHABER AND UITERWAAL, *PHYS. REV. LETT.* **100**, 23002 (2008)

Continued from page 699

GENOMICS

A High-Salt Lifestyle

Bonneau *et al*. describe progress in an effort to link systems-level analysis to events at the molecular and organismal levels. Using experiments and computation, they have pooled transcriptome, protein-protein interaction, structural, and evolution-related data to generate a dynamic model of the halophilic organism *Halobacterium salinarum*. This model was trained on data sets that included more than 200 microarray experiments measuring responses to genetic perturbations and environmental factors (oxygen, sunlight, transition metals, ultraviolet radiation, and desiccation and rehydration). The model, known as EGRIN (environment and gene regulatory influence network) represents transcriptional regulation for 1929 of the 2400 genes in *H. salinarum*, and it was used to predict transcriptional changes after environmental or genetic perturbations (or combinations thereof) that had been held out of the training data sets. As an example, the gene *nhaC3* encodes a Na^+ extrusion pump that allows this organism to grow under high-salt conditions. Analyses of a map of protein-DNA interactions generated from ChIP-chip data could not dissect which of five possible transcriptional regulators governed expression of the gene, yet one of these was predicted by EGRIN to have the strongest effect, which was confirmed in laboratory experiments. — BJ

Cell **131**, 1354 (2007).

ENVIRONMENT

The Debt of Nations

Tracking the worldwide depletion of ecosystem resources is a complex international problem. Srinivasan *et al*. have used a simplified accounting framework to link populations who experience ecological damage to those who cause it. The largest and most blatant imbalance is the debt we (high-income countries) owe to low-income countries because of climate change. On a per capita basis, people in high-income countries are responsible for almost six times more greenhouse gas emissions than their low-income counterparts. Included in the tally is, for instance, the luxury debt accrued by high-income consumers of farmed shrimp; this demand encourages the destruction of coastal mangrove trees to clear the way for shrimp ponds. The resulting loss of storm protection is increasing the risk to adjacent cities as sea levels rise and coral reefs collapse (see also Grimm *et al*., Review, p. 756). Similarly, middle- and high-income countries consume most of the world's fish; nevertheless, several food-deficient African countries charge only modest access fees for the mining of their rich offshore fisheries. Despite the difficulties of measurement and the need to simplify, this analysis raises provocative questions about the division of responsibilities for environmental harm. — CA

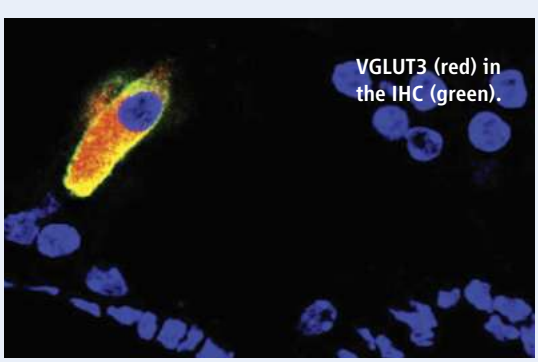
Proc. Natl. Acad. Sci. U.S.A. **105**, 10.1073/pnas.0709562104 (2008).

Science Signaling
AAAS

<< Hearing Essentials About Glutamate Transporters

Although three vesicular glutamate transporters (VGLUTs) have been identified, only two are found in identified glutamatergic neurons. In contrast, VGLUT3 is expressed in several populations of neurons that release other classical neurotransmitters, including inhibitory GABAergic interneurons in the hippocampus and cortex. Seal *et al*. found that mice lacking VGLUT3 were profoundly deaf: They failed to show a startle response to loud noises and did not exhibit auditory evoked potentials. Electrophysiological analysis revealed a defect in signaling from the inner hair cells (IHCs) of the cochlea to the auditory nerve, and morphological analysis showed abnormalities of IHC synapses. Immunofluorescence revealed that VGLUT3 was present in synaptic regions of the IHCs of wild-type mice. Whereas the conductances in the IHCs of the mice lacking VGLUT3 resembled those in wild-type mice, electrophysiological analysis indicated that these neurons failed to release glutamate. The authors conclude that VGLUT3 is essential for hearing and plays an important role in the regulation of cortical excitability. — EMA

Neuron **57**, 263 (2008).



CREDIT: SEAL *ET AL*., *NEURON* **57**, 263 (2008)

www.sciencemag.org

Science

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

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

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

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

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

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

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

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

AAAS was founded in 1848 and incorporated in 1874. Its mission is to advance science and innovation throughout the world for the benefit of all people. The goals of the association are to: foster communication among scientists, engineers and the public; enhance international cooperation in science and its applications; promote the responsible conduct and use of science and technology; foster education in science and technology for everyone; enhance the science and technology workforce and infrastructure; increase public understanding and appreciation of science and technology; and strengthen support for the science and technology enterprise.

Information for Authors

See pages 634 and 635 of the 1 February 2008 issue or access www.sciencemag.org/about/authors

EDITOR-IN-CHIEF **Donald Kennedy**
EXECUTIVE EDITOR **Monica M. Bradford**
DEPUTY EDITORS **R. Brooks Hanson, Barbara R. Jasny, Katrina L. Kelner**
NEWS EDITOR **Colin Norman**

Editorial supervisory senior editor Phillip D. Szuromi; **senior editor/perspectives** Lisa D. Chong; **senior editors** Gilbert J. Chin, Pamela J. Hines, Paula A. Kiberstis (Boston), Marc S. Lavine (Toronto), Beverly A. Purnell, L. Bryan Ray, Guy Riddihough, H. Jesse Smith, Valda Vinson, David Voss; **associate editors** Jake S. Yeston, Laura M. Zahn; **online editor** Stewart Wills; **associate online editors** Robert Frederick, Tara S. Marathe; **book review editor** Sherman J. Suter; **associate letters editor** Jennifer Sills; **editorial manager** Cara Tate; **senior copy editors** Jeffrey E. Cook, Cynthia Howe, Harry Jach, Barbara P. Ordway, Trista Wagoner; **copy editors** Lauren Kmec, Peter Mooreside; **editorial coordinators** Carolyn Kyle, Beverly Shields; **publications assistants** Ramatoulaye Diop, Chris Filiatreau, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Scott Miller, Jerry Richardson, Brian White, Anita Wynn; **editorial assistants** Carlos L. Durham, Emily Guise, Patricia M. Moore, Jennifer A. Seibert; **executive assistant** Sylvia S. Kihara; **administrative support** Maryrose Madrid

News senior correspondent Jean Marx; **deputy news editors** Robert Coontz, Eliot Marshall, Jeffrey Mervis, Leslie Roberts; **contributing editors** Elizabeth Culotta, Polly Shulman; **news writers** Yudhijit Bhattacharjee, Adrian Cho, Jennifer Couzin, David Grimm, Constance Holden, Jocelyn Kaiser, Richard A. Kerr, Eli Kintisch, Andrew Lawler (New England), Greg Miller, Elizabeth Pennisi, Robert F. Service (Pacific NW), Erik Stokstad; **intern** Elsa Youngsteadt; **contributing correspondents** Jon Cohen (San Diego, CA), Daniel Ferber, Ann Gibbons, Robert Irion, Mitch Leslie, Charles C. Mann, Virginia Morell, Evelyn Strauss, Gary Taubes; **copy editors** Rachel Curran, Linda B. Felaco, Melvin Gatling; **administrative support** Scherraine Mack, Fannie Groom; **bureaus** New England: 207-549-7755, San Diego, CA: 760-942-3252, FAX 760-942-4979, Pacific Northwest: 503-963-1940

Production director James Landry; **senior manager** Wendy K. Shank; **assistant manager** Rebecca Doshi; **senior specialists** Jay Covert, Chris Redwood; **specialist** Steve Forrester; **Preflight director** David M. Tompkins; **manager** Marcus Spiegler; **specialist** Jessie Mudjitaba

Art director Kelly Buckheit Krause; **associate art director** Aaron Morales; **illustrators** Chris Bickel, Katharine Sutliff; **senior art associates** Holly Bishop, Laura Creveling, Preston Huey, Nayomi Kevitiyagala; **associate** Jessica Newfield; **photo editor** Leslie Blizard

***Science* International**

Europe (science@science-int.co.uk) **editorial: international managing editor** Andrew M. Sugden; **senior editor/perspectives** Julia Fahrenkamp-Uppenbrink; **senior editors** Caroline Ash, Stella M. Hurtley, Ian S. Osborne, Stephen J. Simpson, Peter Stern; **editorial support** Deborah Dennison, Rachel Roberts, Alice Whaley; **administrative support** John Cannell, Janet Clements, Jill White; **news: europe news editor** John Travis; **deputy news editor** Daniel Clery; **contributing correspondents** Michael Balter (Paris), John Bohannon (Vienna), Martin Enserink (Amsterdam and Paris), Gretchen Vogel (Berlin); **intern** Elizabeth Quill

Asia Japan Office: Asca Corporation, Eiko Ishioka, Fusako Tamura, 1-8-13, Hirano-cho, Chuo-ku, Osaka-shi, Osaka, 541-0046 Japan; +81 (0) 6 6202 6272, FAX +81 (0) 6 6202 6271; asca@os.gulf.or.jp; **asia news editor** Richard Stone (Beijing: rstone@aaas.org); **contributing correspondents** Dennis Normile (Japan: +81 (0) 3 3391 0630, FAX 81 (0) 3 5936 3531; dnormile@gol.com); Hao Xin (China: + 86 (0) 10 6307 4439 or 6307 3676, FAX +86 (0) 10 6307 4358; cindyhao@gmail.com); Pallava Bagla (South Asia: +91 (0) 11 2271 2896; pbagla@vsnl.com)

Africa Robert Koenig (contributing correspondent, rob.koenig@gmail.com)

EXECUTIVE PUBLISHER **Alan I. Leshner**
PUBLISHER **Beth Rosner**

Fulfillment Systems and Operations (membership@aaas.org); **director** Waylon Butler; **customer service supervisor** Pat Butler; **specialists** Laurie Baker, Latoya Casteel, LaVonda Crawford, Vicki Linton; **data entry supervisor** Cynthia Johnson; **specialists** Tarrika Hill, Erin Layne, Sheila Thomas

Business Operations and Administration director Deborah Rivera-Wienhold; **assistant director, business operations** Randy Yi; **senior financial analysts** Michael LoBue, Jessica Tierney; **financial analyst** Nicole Nicholson; **rights and permissions: administrator** Emilie David; **associate** Elizabeth Sandler; **marketing director** John Meyers; **marketing managers** Allison Pritchard, Darryl Walter; **marketing associates** Aimee Aponte, Alison Chandler, Mary Ellen Crowley, Marcia Leach, Julianne Wielga, Wendy Wise; **international marketing manager** Wendy Sturley; **marketing executive** Jennifer Reeves; **marketing/member services executive** Linda Rusk; **site license sales director** Tom Ryan; **sales manager** Russ Edra; **sales and customer service** Mehan Dossani, Iquo Edim, Kiki Forsythe, Catherine Holland, Phillip Smith, Philip Tsolakidis; **electronic media: manager** Lizabeth Harman; **project manager** Trista Snyder; **assistant manager** Lisa Stanford; **senior production specialist** Walter Jones; **production specialists** Nichele Johnston, Kimberly Oster

Advertising director worldwide ad sales Bill Moran

Product (science_advertising@aaas.org); **consumer & sponsorship sales manager** Tina Morra: 202-326-6542; **midwest** Rick Bongiovanni: 330-405-7080, FAX 330-405-7081; **west coast/w. canada** Teola Young: 650-964-2266; **east coast/e. canada** Christopher Breslin: 443-512-0330, FAX 443-512-0331; **uk/europe/asia** Michelle Field: +44 (0) 1223-326-524, FAX +44 (0) 1223-325-532; **japan** Mashy Yoshikawa: +81 (0) 33235 5961, FAX +81 (0) 33235 5852; **senior traffic associate** Deiandra Simms

Commercial Editor Sean Sanders: 202-326-6430

Classified (advertise@sciencecareers.org); **us: recruitment sales manager** Ian King: 202-326-6528, FAX 202-289-6742; **inside sales manager: midwest/canada** Daryl Anderson: 202-326-6543; **northeast** Alexis Fleming: 202-326-6578; **southeast** Tina Burks: 202-326-6577; **west** Nicholas Hintibidze: 202-326-6533; **sales coordinators** Erika Foard, Rohan Edmonson, Shirley Young; **international: sales manager** Tracy Holmes: +44 (0) 1223 326525, FAX +44 (0) 1223 326532; **sales** Marium Hudda, Alex Palmer, Alessandra Sorgente; **sales assistant** Louise Moore; **japan** Mashy Yoshikawa +81 (0)3 3235 5961, FAX +81 (0)3 3235 5852; **advertising production operations manager** Deborah Tompkins; **senior production specialists** Robert Buck, Amy Hardcastle; **senior traffic associate** Christine Hall; **publications assistant** Mary Lagnaoui

AAAS Board of Directors retiring president, chair John P. Holdren; **president** David Baltimore; **president-elect** James J. McCarthy; **treasurer** David E. Shaw; **chief executive officer** Alan I. Leshner; **board** John E. Dowling, Lynn W. Enquist, Susan M. Fitzpatrick, Alice Gast, Linda P. B. Katehi, Cherry A. Murray, Thomas D. Pollard, Kathryn D. Sullivan

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

Published by AAAS

Human Universals

In 1967, psychologist Paul Ekman visited New Guinea to test the idea proposed by Charles Darwin a century earlier that human facial expressions are universal. Last month, the Exploratorium science museum in San Francisco, California, celebrated the 40th anniversary of his trip. The museum's new Mind exhibit displays some of Ekman's photos for the first time, including this montage of indigenous South Fore men. Ekman asked each to show how he would look if he (from left) learned that his child had died, met friends for the first time that day, saw a dead pig in the road, or was about to fight with someone. Anthropologists now agree, says Ekman, that such expressions are biologically determined, as Darwin had thought.

Sarkozy's Bitter Pill

French scientists are reacting with growing fury to plans by President Nicolas Sarkozy to overhaul basic research.

In a 21 January speech paying homage to France's 2007 physics Nobelist Albert Fert, Sarkozy lashed out at the research system as plagued by "balkanization" and threatened by "paralysis." He said major agencies such as CNRS should be turned into "funding bodies rather than performers of research" to "implement science policy specified by the government." The government may also go ahead with a controversial plan to replace open-ended job commitments for scientists with 4-year contracts.

Some scientists are livid. Sarkozy wants to "implement rapidly a new stage in demolishing our [basic research] system," said chemist Henri-Edouard Audier, a board member in the leading researchers' union. The president's attack "thoroughly blackens the situation to show that everything is so rotten that it must be destroyed," Audier added.

Physicist Bertrand Monthubert, president of the Let's Save Research movement, said protest plans are being laid. The last time researchers took to the streets was in 2004, to protest budget cuts.

A TEST FOR STRING THEORY, IF ONLY

Skeptics of string theory complain that the purported "theory of everything" is so complicated and flexible that it's impossible to test experimentally. But theoretical cosmologists Rishi Khatri and Benjamin Wandelt, both of the University of Illinois, Urbana-Champaign, claim they've devised such a test.

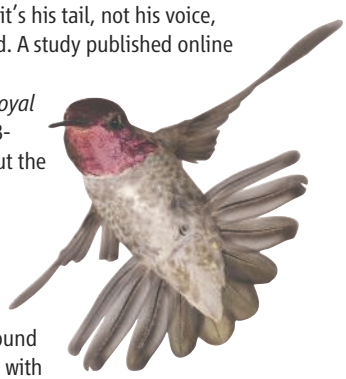
According to string theory, every fundamental particle is really a wriggling filament stretching less than a billionth of a billionth of the width of a proton. In some models, space is riven with long cosmic strings, defects in spacetime that affect the distribution of matter. Detectable evidence for them should show up in the mottled distribution of hydrogen gas in the early universe, Khatri and Wandelt calculate in a paper to be published in *Physical Review Letters*. All you need to see it is an array of radio telescopes covering 10,000 square kilometers.

A practical proposal? "Nobody would be able to build a 10,000-square-kilometer array," chuckles Yervant Terzian, a radio astronomer at Cornell University. Terzian is a member of a team proposing to build a 1-square-kilometer array, and that alone will cost more than $2 billion, he says.

Plan for the Square Kilometer Array—just multiply by 10,000.

The Tail Tells

The male Anna's hummingbird emits an emphatic squeak as he swoops above his intended mates, but it's his tail, not his voice, that makes the sound. A study published online 29 January in the *Proceedings of the Royal Society B* settles a 68-year-old debate about the source of the chirp.



Ornithologist Christopher Clark of the University of California, Berkeley, suspected that the sound had something to do with the birds' unusually shaped outer tail feathers. So his team took high-speed videos of the birds doing dive displays in a California park. The videos revealed that a quick tail flick coincided with the squeak, and the displays of four consistent chirpers were silenced when researchers trimmed their tails.

In the lab, the researchers found that isolated feathers produced a continuous whine when subjected to an air stream. They fluttered like little flags, generating the same frequency regardless of wind speed. "We were blown away," Clark says, when they realized that the feathers worked like musical reeds, a mechanism previously unknown in birds.

The study "fully solves" the question of the Anna's chirp and is likely to explain other nonvocal bird sounds as well, says biomechanist Douglas Altshuler of the University of California, Riverside. What's more, the tail sound is nearly identical to part of the birds' vocal song. It's "very wild," Altshuler says, that the birds evolved to make the same sound in two completely different ways.

CREDITS (TOP TO BOTTOM): PAUL EKMAN; *PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON*/CHRIS CLARK; SKA PROJECT OFFICE AND XILOSTUDIOS

Published by AAAS

Three Q's >>

Josephine Briggs, 63, a nephrologist and former National Institutes of Health (NIH) administrator, this week became the second director of its 9-year-old National Center for Complementary and Alternative Medicine (NCCAM). She replaces Stephen Straus, who died in 2007.

Q: NCCAM is probably the most controversial institute at NIH. That didn't deter you?

I think it's also an area that's very high on the public profile. The aim [of the institute] is absolutely no relaxation in the notion of rigorous science. We're going to try very hard to continue to support only very top-ranked, careful, rigorous science.

Q: Have you ever tried alternative medicine yourself?

Well, I'm a regular exerciser. I do some yoga. It's sort of part of the whole mind-body interface.

I pay attention to this literature. A lot of people, including a lot of scientists, are using these agents. I think the real rationale for NCCAM's investment is the enormous public interest in this. I think we all want answers as to which of these approaches really will help with the symptoms of aging.

Q: Is there anything that has come out of alternative medicine that you think clearly works?

There are small, promising areas. The tai chi for shingles was a very nice study (nccam.nih.gov/research/results/past/index.htm).

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

CREDITS (TOP TO BOTTOM): PAUL FETTERS, COURTESY OF HHMI; CAROLYN NAPOLI/UNIVERSITY OF ARIZONA; DEBORAH FEINGOLD/CORBIS

MOVERS

LINKING UP. Plant geneticist Richard Jorgensen was trained as an engineer, a background that should help the University of Arizona, Tucson, researcher lead a $50 million initiative to build computational tools and interdisciplinary teams for plant biology.

Jorgensen, 56, is known as the "Petunia Man" for his work on gene silencing and flower color in that species. Over the years, his research has increasingly required the analysis of large data sets about proteins and genes. This effort, along with his 5-year stint as editor-in-chief of *The Plant Cell*, convinced him that the plant community needed better ways to compile and analyze information across disciplines as disparate as genomics and ecology. Last week, the multiuniversity consortium he proposed to accomplish that goal, called the iPlant Collaborative, got a 5-year grant from the U.S. National Science Foundation.

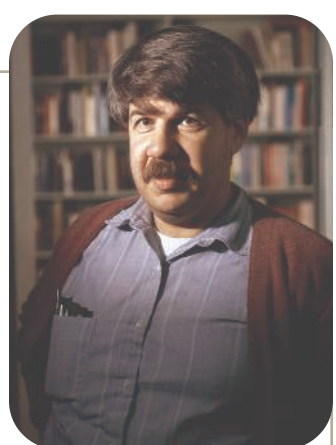
Plant biologist Robert Last of Michigan State University in East Lansing says Jorgensen's experience makes him an ideal person to lead the effort, which will involve "bringing together biologists, information scientists, computer infrastructure [engineers], and informatics experts to work as teams."

IN BRIEF

The Israel-based Wolf Foundation has announced prizes to honor a host of accomplishments in basic and applied science, from geometry to pest management. **William Moerner** of Stanford University in Palo Alto, California, and **Allen Bard** of the University of Texas, Austin, will be awarded the chemistry prize for their contributions to single molecule spectroscopy. **John Pickett** of Rothamsted Research in Hertfordshire, U.K., **James Tumlinson** of Pennsylvania State University in State College, and **W. Joe Lewis** of the U.S. Department of Agriculture in Tifton, Georgia, will share the agriculture prize for helping develop better pest-control methods. The mathematics prize will honor **Pierre Deligne** and **Phillip Griffiths**, both of the Institute for Advanced Study in Princeton, New Jersey, and **David Mumford** of Brown University for their contributions to arithmetic, complex differential geometry, algebraic theory, and several other topics. And the medicine prize will recognize **Howard Cedar** and **Aharon Razin**, both of the Hebrew University of Jerusalem in Israel, for helping to advance the understanding of how gene expression is controlled. Each prize is worth $100,000.

Bill Read has been named director of the National Hurricane Center (NHC) in Miami, Florida. A meteorologist who once served in the U.S. Navy, Read has served as the center's acting head since the controversial exit of Bill Proenza last summer (*Science*, 13 July 2007, p. 181).

Celebrities >>



EVOLUTION OF AN IDEA. Sometime this year, visitors to the Stanford University library will be able to look into the mind of Stephen Jay Gould, the Harvard paleontologist and popularizer of science.

Gould, who died in 2002, bequeathed his notes, letters, and office library to Stanford in hopes it would put the materials online. Hyperlinks among the digitized documents would allow users to trace the evolution of Gould's ideas from handwritten marginal notes to outlines, manuscripts, letters, and final published works. But the library does not yet have the funds to do the required scanning and exhaustive indexing, says Henry Lowood, curator of Stanford's History of Science and Technology Collections.

In the meantime, Lowood's staff is still cataloging the monumental stack of materials it began receiving in 2004, with the goal of putting it on public display. "Steve was a pack rat," says artist Rhonda Shearer, Gould's widow. Organizing his collections—which included not only academic texts but also Beanie Babies and baseball memorabilia—"could sometimes feel glacial," she says.

Published by AAAS

Bluetongue blues
710

Prizes or patents?
713

Ethnic clashes. Youth gangs run amok (*left*) after burning houses in a Nairobi slum last month. The political violence struck hardest in the capital and in western Kenya.

AFRICA

Kenyan Scientists Endure Violent Unrest, University Closings

Deadly ethnic clashes in Kenya over the past month, sparked by a disputed presidential election and tribal tensions, have closed public universities, disrupted numerous field research projects, and caused the postponement or cancellation of several scientific conferences in the normally peaceful East African nation. But progress in mediation efforts offered hope this week that the tensions can be defused.

More than 900 people have died and a quarter-million have been driven from their homes—mainly in western Kenya's Rift Valley and in Nairobi's slums (see map, below)—as a result of the violence that has struck the country since President Mwai Kibaki was declared the winner of the election, which opposition candidate Raila Odinga has alleged was rigged. Former United Nations Secretary-General Kofi Annan helped mediate a framework agreement on 1 February, but clashes continued in some areas, causing some institutes to scale back or even cancel projects that required risky travel.

"There have been phenomenal problems with field research and logistics," says entomologist Christian Borgemeister, director general of the International Centre of Insect Physiology and Ecology (ICIPE) in Nairobi, Africa's leading center for the study of insect vectors. The unrest, he says, has had "a severe impact" on research projects at the center's Mbita Point Field Station at Lake Victoria—a hot spot of ethnic strife—and some of ICIPE's 300 staffers have lost their homes or are sheltering refugees.

Although the worst violence has struck southwest Kenya, riots in Nairobi have also disrupted field projects. Demographer Alex Ezeh, executive director of the African Population and Health Research Center in Nairobi, says the unrest led him to "postpone indefinitely" parts of a longitudinal demographic and health surveillance system in two Nairobi slums.

In part because of Kenya's traditional stability, Nairobi is a major center for pan-African research organizations, which have been riding out the storm. Geologist Judi Wakhungu, executive director of the African Centre for Technology Studies in Nairobi, says, "Our major concern is safety." Although the center's researchers have not been hurt, she says, "our agendas and work plans have been disrupted by postponing activities until calm is restored." The International Livestock Research Institute in Nairobi, whose field researchers investigate livestock maladies and help develop vaccines, has temporarily "reduced its field projects in Kenya as a result of security concerns," says spokesperson Susan MacMillan.

Agricultural research centers have been especially hard hit. Last week, a gang of armed youths drove an estimated 500 employees, including some scientists, from the Kenya Agricultural Research Institute and the Kenya Forestry Research Institute, northwest of Nairobi.

Although rumors have flown that international organizations might leave Nairobi if the civil strife continues, a spokesperson for the United Nations Environment Programme, which employs about 350 staffers at its world headquarters there, says no such plan is being considered. "Nairobi is an important research hub; it would be a disaster for any of the major institutes or organizations to move because of this," says Mohamed Hassan, a Sudanese mathematician who is president of the Nairobi-based African Academy of Sciences, which is also committed to keeping its headquarters in Kenya.

Kenyan university administrators were trying to gauge when they might resume classes. Private colleges were open, but students had not yet returned this week at most public universities. Physicist Frederick N. Onyango, vice chancellor of Maseno University in the hard-hit Nyanza Province, told *Science* that classes there won't be started until April. "The main reason is the violence in the city, which has destroyed the shops of university suppliers," he says. Although many lecturers are on leave, some research laboratories are still functioning. At Masinde Muliro University of Science and Technology in Kakamega in western Kenya, rioters armed with machetes used ▶

CREDIT: (TOP) MAXPPP/LANDOV

gasoline to set afire three hostels rented by university students. Meanwhile, officials at Moi University in the Rift Valley town of Eldoret reported that some agricultural research had been disrupted.

At the University of Nairobi, "systems are at the moment functioning minimally," says Benson Estambale, who directs the school's Institute of Tropical and Infectious Diseases. Even so, the university's College of Health Sciences resumed classes in late January. "We hope things will cool down soon," says pediatrics professor Nimrod O. Bwibo.

Outside scientists who conduct research in Kenya are taking a careful look at the situation. Ecologist David Harper of the University of Leicester, U.K., who heads the Earthwatch Institute's Lakes of the Rift Valley project, says he hopes for a quick settlement, but if violence continues in Naivasha and Nakuru—troubled cities near lakes being studied—his research groups have the option of shifting focus temporarily to other Rift Valley lakes. So far, foreign medical assistance and research programs in Kenya have continued despite community disruptions. The U.S. Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, evacuated eight of its American researchers and their families—as well as about 40 of the 900 Kenyan staffers—from the Kisumu field research station to Nairobi. But the facility's director, epidemiologist Kayla Laserson, said Monday that she planned to return this week to supervise vaccine trials and research projects. Kenya is home to the CDC's largest overseas presence.

The Annan-led mediation entered a crucial phase this week, with both sides calling for an end to the violence and opposition leader Odinga asking for foreign peacekeepers. Researchers hoped for an early resolution that would restore Kenya's reputation for safety. Says Estambale: "We are traumatized and not believing what has gone wrong" in recent weeks. **–ROBERT KOENIG**

BIOCHEMISTRY

Lifting the Veil on Traditional Chinese Medicine

DALIAN, CHINA—Genome, proteome, metabolome … *herbalome*? In the latest industrial assault on nature's biochemical secrets, a Chinese team in this seaside city is about to embark on a 15-year effort to identify the constituents of herbal preparations used as medications for centuries in China.

The Herbalome Project is the latest—and most ambitious—attempt to modernize traditional Chinese medicine (TCM). The venerable concoctions—as many as 400,000 preparations using 10,000 herbs and animal tinctures—are the treatment of choice and often the only recourse for many in China. In the 1970s, TCM tipped off researchers to qinghaosu, a compound in sweet wormwood whose derivatives are potent antimalaria drugs. But TCM's reputation has been blackened by uneven efficacy and harsh side effects, prompting critics to assail it as outmoded folklore. "TCM is not based on science but based on mysticism, magic, and anecdote," asserts biochemist Fang Shi-min, who as China's self-appointed science cop goes by the name Fang Zhouzi. He calls the Herbalome Project "a waste of research funds."

Hoping to rebut TCM critics, Herbalome will use high-throughput screening, toxicity testing, and clinical trials to identify active compounds and toxic contaminants in popular recipes. "We need to ensure that TCM is safe and also show that it is not just qinghaosu," says Guo De-an, who leads TCM modernization efforts at the Shanghai Institute of Materia Medica and is not involved in Herbalome. Initial targets are cancer, liver and kidney diseases, and illnesses that are difficult for Western medicine to treat, such as diabetes and depression.

The Dalian Institute of Chemical Physics (DICP), one of the biggest and best-funded institutes of the Chinese Academy of Sciences, won a $5 million start-up grant to develop purification methods; the Ministry of Science and Technology is reviewing the project with a view to including it as a $70 million initiative in the next 5-year plan to start in 2010. A planning meeting will be held at a Xiangshan Science Conference—China's equivalent of a Gordon Research Conference—in Beijing this spring.

Several TCM power players have thrown their weight behind the initiative. "It's the right time to start this project," says chemist Chen Kai-xian, president of the Shanghai University of Traditional Chinese Medicine. Herbalome should appeal to pharmaceutical firms, as it could identify scores of drug candidates, says Hui Yongzheng, chair of the Shanghai Innovative Research Center of Traditional Chinese Medicine.

In many parts of the world, traditional medicine recipes are handed down orally from one generation to the next. But in China, practitioners more than 2000 years ago began to compile formulations in compendia. Although in major cities Western medicine has largely supplanted TCM, many Chinese still believe in TCM's power as preventive medicine and as a cure for chronic ailments, and rural Chinese depend on it. "For most of us, when we feel unwell, we want to take TCM," says chemist Liang Xinmiao of DICP.

Medicine man. Liang Xinmiao's Herbalome Project aims to identify active ingredients and toxins in thousands of traditional Chinese preparations.

Since the Mao Zedong era, the government has strongly supported TCM, in part because it was too expensive to offer Western medicine to the masses. It remains taboo for Chinese media to label TCM as pseudoscience. "Criticizing TCM is unthinkable to many Chinese and almost like committing a traitorous act," says Fang. ▶

CREDIT: R. STONE/*SCIENCE*

Published by AAAS

Proponents insist that TCM has much to offer. But for every claimed TCM success, there are reports of adverse effects from natural toxins and contaminants such as pesticides. Dosages are hard to pin down, as preparations vary in potency according to where and when herbs are harvested. Quality can vary from manufacturer to manufacturer and from batch to batch. "That's why many people don't trust TCM," says Guo. In the modernization drive, quality control is a paramount concern.

Herbalome intends to take modernization to a whole new level. The initiative is the brainchild of Liang, who believes many TCM recipes are effective. "The problem is, we don't know why it works," he says. The main hurdle is the complexity of the preparations. As an example, Liang shows a chromatograph of Hong Hua, or "red flower," a preparation applied externally for muscle pain. In many samples chemists deal with, one peak usually represents one compound, Liang says. But for Hong Hua, each peak is many compounds, and fractionating these yields more multicompound peaks like nested matryoshka dolls. Hong Hua is composed of at least 10,000 compounds, says Liang: "We know only 100."

Faced with such complexity, "we must invent new methodologies," says Liang. "This is the battleground of the Herbalome project." For starters, his 45-person team at DICP is developing new separation media. Herbs will be parsed into "multicomponents": groups of similar constituents. To determine which substances are beneficial or toxic, his group plans to devise Herbalome chips in which arrays of compounds are screened for their binding to key peptides. The expanded Herbalome project would involve researchers at many institutes in China and abroad.

Herbalome has potential pitfalls. One is a concern that Western companies will develop blockbuster drugs—and walk away with the spoils—by modifying compounds identified by the project. To counter this possibility, says Guo, "we're encouraging scientists not to rush to publish and do structure modifications [to identify drug candidates] first." Teams would then apply for patents on groups of similar structures.

Not all practitioners embrace TCM's demystification. "Some are afraid that the traditions will be lost," says Chen. But Hui says that modernization is necessary "to reconcile the knowledge-oriented, deductive process of Western medicine with the experience-oriented, inductive process of TCM." Fang has a different take: "Can you marry astrology and astronomy, alchemy and chemistry? It never works."

Hui insists that TCM can coexist with Western medicine. Liang hopes his Herbalome project will prove Hui right.

–RICHARD STONE

With reporting by Li Jiao in Beijing.

ANIMAL DISEASE

Exotic Disease of Farm Animals Tests Europe's Responses

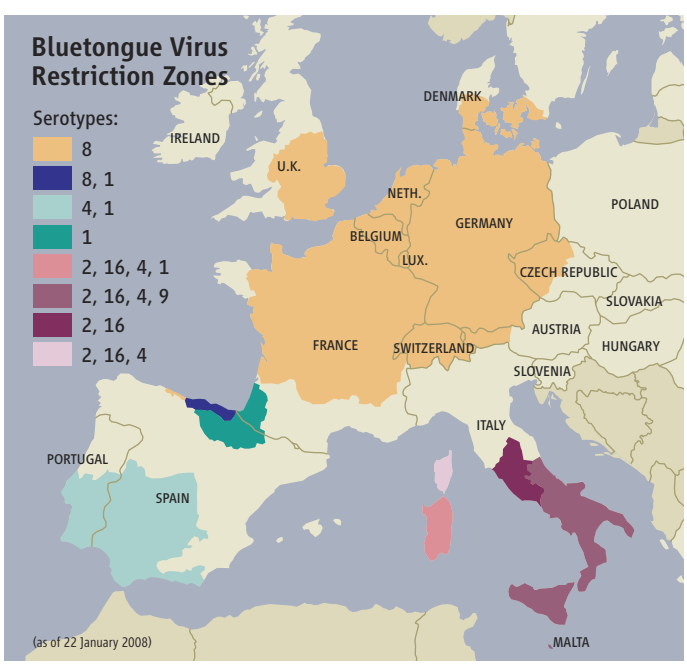
BRUSSELS—A race against the clock is on at farms in northern and central Europe. The question: Can bluetongue, an exotic insect-borne viral disease that unexpectedly popped up here in 2006 and expanded aggressively in 2007, be stopped in 2008? For the past year and a half, manufacturers have been scrambling to produce a vaccine against the particular viral strain, called serotype 8; a handful are ready. But as a recent meeting* here showed, a number of logistical, scientific, and political problems still threaten to hobble the fight.

It's unclear, for instance, whether the tens of millions of vaccine doses needed, preferably as early as May, can be delivered in time. Countries are also still pondering their vaccination strategies and which vaccine to use. The debate is complicated by a fundamental and, for the moment, unanswerable question: Can bluetongue still be wiped off the map entirely, or has climate change created such favorable conditions that the disease is here to stay?

Bluetongue, of which 24 different serotypes exist, is transmitted by *Culicoides*, or biting midges. The virus causes high fevers and swelling of the face, lips, and tongue. Sheep are most susceptible; mortality rates vary widely, by strain and location, but may reach 10% or more. Other ruminants—including cattle and goats—can be infected as well. The disease, for which no treatment exists, is not transmissible between animals.

Until 10 years ago, bluetongue was barely known in Europe. The virus was found primarily in tropical and subtropical zones in Africa, Asia, and the Americas. But starting in 1998, serotypes 1, 2, 4, 9, and 16 moved from Africa and the Middle East into southern Europe. The biggest surprise came in 2006, when serotype 8, presumably originating in sub-Saharan Africa, caused outbreaks in the Netherlands, Belgium, and Germany. Scientists had no idea the disease could gain a foothold there, because the best-known vector, *C. imicola*, doesn't occur at this latitude; recent studies suggest that local species such as *C. obsoletus* make just as good vectors, however.

The serotype 8 outbreak has dwarfed the previous incursions in southern Europe, both in geographic spread and in number of infected animals. Last summer, the virus spread to eight new countries. Unless drastic measures are taken, 2008 promises to be "disastrous," says Peter Mertens of the Institute for Animal Health in Pirbright, U.K. Belgium lost 15% of its sheep last year, he says; if the same ▶



Virgin territory. Northern and central Europe had never seen bluetongue before a major outbreak of serotype 8 took off in 2006. Southern Europe is home to five other serotypes.

SOURCE: EUROPEAN COMMISSION

* "Conference on Vaccination Strategy Against Bluetongue," 16 January.

Published by AAAS

Open your mouth. Sheep are the bluetongue virus's main victims.

CREDIT: A. J. CANN

happened in the United Kingdom, home to 34 million sheep, "that's a lot of dead animals." The disaster could rival the foot-and-mouth outbreak of 2001, he adds.

Although bluetongue vaccines are available, every strain requires its own vaccine, and some of them carry risks. The first vaccines against bluetongue were live, attenuated viruses, many of them produced by Onderstepoort Biological Products in South Africa, where several serotypes of bluetongue are endemic. Such vaccines are made by growing a virus in cell culture or eggs for many generations until it is weakened enough not to cause disease. They are easy and cheap to produce in large quantities. Spain, Portugal, France, and Italy have all used them in the past 8 years, in most cases successfully. But a vaccine against serotype 16, which France used to battle an outbreak on the island of Corsica in 2004, turned out to be pathogenic and transmissible by midges.

Live vaccines have also been linked to higher abortion rates and decreased milk production, which is why the European Food Safety Authority has recommended that countries use a new generation of inactivated (killed) vaccines. Five companies have now developed killed vaccines against serotype 8.

Whether they can avert disaster this year remains to be seen. Most countries have hesitated to order massive amounts of the vaccine. As of last week, only the United Kingdom and the Netherlands had ordered 22.5 million and 6 million vaccine doses, respectively, from Intervet, a Dutch company. Yet the vaccine takes some 5 months to produce, says a spokesperson for Intervet; countries that order now won't have the vaccine by May, when the virus could start rearing its head again. There could be critical shortages—and painful questions about how to distribute a short supply—if countries don't order soon, warns Declan O'Brien, managing director of IFAH, a Brussels-based industry group.

Countries' slow response is a result of bureaucratic rules—most procure vaccines through time-consuming competitive contracts—and questions about how to use the vaccine. In theory, it's possible to wipe serotype 8 off the northern European map, says Eugène van Rooij of the Central Veterinary Institute in Lelystad, the Netherlands. But that would require an extremely rigorous, multiyear vaccination campaign in each country; it's no use for Germany to go for elimination if, say, Switzerland and Belgium do not. Nor is it clear that the costs of such an operation would outweigh the benefits in reduced disease and mortality.

Complicating matters, an elimination plan would probably work only if vaccination became compulsory—but farmers are divided on that question, says Klaas Johan Osinga, vice chair of the animal health and welfare working group within COPA-COGECA, an international farmer's organization in Brussels. Although sheep farmers—especially those in affected areas—are eager to vaccinate, cattle farmers, who have not been hit as hard, tend to be more wary of a vaccination campaign.

As a result, "everybody is sort of looking at each other," Van Rooij says. The most likely result is that a vaccination campaign will aim to reduce disease rather than eliminate it altogether from northern and central Europe. Still, the European Commission hopes to achieve at least 80% coverage, a threshold that past experience suggests will all but halt spread of the disease. The commission has offered to cofinance national vaccination campaigns, provided they try to reach that 80% target.

Also under debate is whether countries should consider using live vaccines if companies can't produce enough of the killed variety. "I wouldn't be keen on using them," says Mertens. But Vincenzo Caporale of the World Organisation for Animal Health in Paris, who helped develop a live vaccine while at the Istituto Zooprofilattico Sperimentale in Teramo, Italy, says that by ruling out such vaccines prematurely, northern Europe is exposing southern Europe to unnecessary risks of serotype 8 invasion.

Even if the spread of serotype 8 is halted this year, that may not be the end of the story. In a 2006 paper in *Nature*, Mertens and colleagues proposed that global warming has created more favorable conditions for European *Culicoides* populations and the virus. That might mean the continent is in for a lot more trouble. "There are 24 serotypes. If one of them can survive in northern Europe, then who knows what will arrive next," says Mertens. **–MARTIN ENSERINK**

Go Code Orange, Labs Urged

This week, the Society for Neuroscience released guidelines (sfn.org/animals) to help universities and institutions protect researchers from attacks by animal-rights extremists. Concerned about recent incidents involving vandalism of researchers' homes and threats and harassment of family members (*Science*, 21 December 2007, p. 1856), the society urges research institutions to take active steps, such as working with local police to ensure rapid responses to attacks off campus, and encourages university leaders to forcefully condemn attacks when they occur. "It has all the right elements," says Roberto Peccei of the University of California, Los Angeles, where several recent attacks have occurred.

–GREG MILLER

Coal Plant Burnt

As part of the 2009 budget request to Congress (see p. 714), the U.S. Department of Energy (DOE) wants to cancel a $1.8 billion coal power plant after cost estimates nearly doubled. The project, an industry partnership called FutureGen, was to demonstrate by 2012 the first-ever coal-fired plant designed to capture carbon dioxide from coal while producing hydrogen for power. Now DOE, which blames the cost overruns on skyrocketing material and labor prices, wants to work with industry to make several less sophisticated plants by 2015 that will capture CO_2 for underground storage but will not produce hydrogen. Lawmakers—including senator and presidential candidate Barack Obama (D–IL), who represents Mattoon, Illinois, where the plant was to be built—will try this year to force DOE to stick to its original plan. **–ELI KINTISCH**

AIDS Research Taking Time Out

In the wake of yet more disappointing results from human studies of AIDS vaccines last fall (*Science*, 16 November 2007, p. 1048), the U.S. National Institute of Allergy and Infectious Diseases (NIAID) plans to hold a daylong summit on 25 March to reassess how it invests the nearly $600 million it spends annually on the field. The summit, which will be webcast and open to the public, came about after 14 leading AIDS researchers sent NIAID Director Anthony Fauci a letter contending that NIAID was investing too heavily in developing products and should spend more of its budget on basic research. "The real issue is the balance that we want between discovery research and development," says Fauci. "We need to take a time out and talk to people in the field."

–JON COHEN

RESEARCH FUNDING

Prizes Eyed to Spur Medical Innovation

MAASTRICHT, THE NETHERLANDS—If the World Health Organization offered a $10 billion award for a malaria vaccine, would that persuade major pharmaceutical companies to go after the prize? Could a $100 million prize encourage development of a reliable, cheap, and fast diagnostic assay for tuberculosis? And would those monetary awards prove to be the cheapest, or fastest, way to achieve such medical innovations?

Provocative questions such as those were at the core of a 2-day workshop* here last week addressing whether prize incentives can stimulate the creation of new drugs and therapies. For some speakers, prizes offer a chance to spur medical research on neglected diseases, including those that strike people in developing nations who can afford little health care. Others took a more radical view: A national or global medical prize scheme could eliminate drug patents, stimulate drug development, and lower escalating health care costs. "A prize is a [research] incentive, the same way a monopoly is an incentive," says James Love, director of the think tank Knowledge Ecology International (KEI) in Washington, D.C.

NOTABLE AWARDS	AMOUNT	GOALS	WINNER
Napoleon's Food Preservation Prize (1795)	12,000 francs	Ease food-supply problems for invading armies	Nicolas François Appert's canning process (1809)
Ansari X Prize (1995)	$10 million	Private, crewed reusable spacecraft	SpaceShipOne (2004)
Methuselah Mouse Prize (2003)	$4.5 million fund	Extend longevity or slow aging of mice	Multiple winners
Prize4Life (2006)	$1 million	Biomarker for ALS	No winners so far

Prize-worthy. Noting the long history of scientific and technological prize contests, James Love argues that a national scheme of awards for medical innovations should replace drug patents.

Cosponsored by KEI and UNU-MERIT, a research and training center run jointly by United Nations University and Maastricht University, the workshop drew several dozen economists, intellectual-property specialists, public-health officials, and drug-development experts to discuss a concept that's attracting more attention. For example, U.S. Senator Bernie Sanders (I–VT) has introduced a bill, the Medical Innovation Prize Act, written with Love's help, that would replace medical patents with an estimated $80 billion annual award fund. Although the bill is unlikely to go anywhere now, Sanders hopes to get a Senate hearing this year to publicize the concept. "There is growing interest and political feasibility for trying prizes in a variety of contexts," says Stephen Merrill of the U.S. National Academies, who recently examined how the U.S. National Science Foundation could set up a prize system to stimulate innovation (*Science*, 26 January 2007, p. 446).

Prize contests have long been used to steer efforts toward particular discoveries or technological accomplishments, and they're becoming popular again (*Science*, 30 September 2005, p. 2153). One well-known early success was the British government's 18th century prize to find a way for seafarers to gauge longitude. More recently, the $10 million Ansari X Prize for a private, reusable, crewed spacecraft prompted an estimated $100 million to $400 million in space-flight research before Burt Rutan's SpaceShipOne won it in 2004.

Although perhaps not as prevalent as technology competitions, medical prizes are attracting sponsors. Pierre Chirac of Médecins sans Frontières said at the meeting that his group was considering an award for the desperately needed TB diagnostic test. And in 2006, Prize4Life, a nonprofit group founded by a patient with amyotrophic lateral sclerosis (ALS), announced a $1 million prize for a biomarker that can track the fatal disease's progression—a key for any drug development. Prize4Life hopes to launch two more contests, including a $2.5 million prize for a treatment that proves effective in a common mouse model of ALS.

Such modest awards pale in comparison to the mammoth prize system Love advocates through the Sanders bill. Financed annually with 0.6% of the United States's gross domestic product—about $80 billion at the moment—the Sanders plan would give annual awards to medical innovations based on the health impact for the nation—assessed using a measurement known as quality-adjusted life years that gauges improvements in life expectancy. Instead of the government granting patents to a company, a board that would include business and patient representatives, as well as government health officials, would each year judge any new products and award their developers a share of the fund.

At the Maastricht meeting, intellectual-property specialist William Fisher III of Harvard Law School argued that prize schemes have some advantages. Patents, said Fisher, guide medical research away from vaccines, which may require at most a few doses per person but arguably have the most health impact, and toward treatments for the rich and the development of "me-too" drugs, copies of an already successful drug with just enough differences to be patentable. "Prizes can offset all three" of those biases, he says.

PhRMA, a trade group in Washington, D.C., that represents pharmaceutical and biotech firms, has strongly criticized the Sanders bill as a step toward socialized medicine. And yet it is intrigued by new incentives, if the patent system stays intact. "It's an interesting idea to add prizes for neglected diseases to the existing system," says Shelagh Kerr of PhRMA, who attended the workshop.

Prize incentives are, however, unlikely to sweep the medical research world. Philanthropic and patient groups may offer new awards, but governments may be more cautious. "We're no longer in the Longitude Prize era. We pay scientists many millions to do research," says David King, former science adviser to the U.K. government. "How do you decide how much money to award?" adds economist Aidan Hollis of the University of Calgary in Canada, noting that governments typically don't know in advance what social value a medical treatment will have.

The workshop itself offered an ironic morsel of evidence that prizes are not perfect incentives. Organizers offered a €1500 award for the best paper on using monetary prizes to stimulate private investment in medical research, but no entries have been submitted thus far. The contest has now been extended to mid-April.

–JOHN TRAVIS

* "Medical Innovation Prizes as a Mechanism to Promote Innovation and Access," 28–29 January.

CREDIT: (IMAGE) HERMAN PIJPERS/UNU-MERIT

Published by AAAS

A Science Budget of Choices and Chances

In his final year, President George W. Bush has submitted a request for 2009 funding with few new wrinkles—and with probably little chance of being adopted

Budgets are about choices, U.S. presidential science adviser John Marburger told reporters this week as he explained what his boss is asking Congress to support in 2009. And what President George W. Bush has chosen for science funding is exactly what he has requested for the past few years: Give a big boost to agencies that support the physical sciences, flat-line basic biomedical research, and put NASA between a rock and a hard place.

The betting in Washington is that the Democratic Congress won't grant a lame-duck Republican president his wish, and that it is likely to delay approving any part of the Administration's overall $3.1 trillion budget request for the fiscal year that begins on 1 October until after the November elections. But in the meantime, the president's support for some disciplines at the expense of others has left science lobbyists uncertain about how to react. As Robert Berdahl, president of the 62-member Association of American Universities, puts it: "Question: Is the president's [2009] budget good or bad for the vital research and education that is performed by America's research universities? Answer: Yes."

The big winners are the three agencies that are part of what the Bush Administration has labeled the American Competitiveness Initiative (ACI). The $6 billion National Science Foundation (NSF) would receive a 13.6% jump, the Department of Energy's (DOE's) $4 billion Office of Science would get a 17.5% hike, and the $500 million core research programs at the National Institute of Standards and Technology (NIST) would get a 22% bump. The large boosts compensate for double-digit increases that were in the cards for all three agencies in 2008 until a last-minute budget deal erased most of their gains (*Science*, 4 January, p. 18), prompting the early termination of some experiments and scheduled layoffs at two DOE national laboratories.

There is bipartisan agreement about the value of a healthy science budget. "The president is right that basic research included in his American Competitiveness Initiative (ACI) is important to our economy and our future," says Representative Bart Gordon (D–TN), chair of the House Science Committee. But Gordon is very unhappy with some of the choices made along the way, including what he sees as miserly increases in several education programs at NSF, the lack of proposed funding for a new high-risk research agency at DOE that he championed, and the Administration's repeated attempts to eliminate technology and manufacturing programs at NIST.

Biomedical organizations lament not just the lost research opportunities but also the impact on the next generation of scientists. "This is a real deterrent for any young investigators who were holding out hope that biomedical research was a viable career path," says Robert Palazzo, president of the Federation of American Societies for Experimental Biology in Bethesda, Maryland. "They have their answer today."

Marburger disagrees that a flat budget means a gloomy future for biomedical researchers. "Frankly, I think that an argument can be made that better management [of NIH] can bring about much better productivity even with flat resources," he says. "The private sector does it all the time." And he says that those who advocate 6% annual growth for NIH to capitalize on its 5-year doubling that ended in 2003 will have to wait their turn. "It will be necessary to increase biomedical research in the future, but it's important that we first fix this problem in the physical sciences."

AGENCY	FY 2007	FY 2008 (all figures in $ millions)	FY 2009 (request)	% CHANGE
National Institutes of Health†	29,137	29,465	29,465	+0.0%
National Science Foundation	5,916	6,032	6,854	+13.6%
Research	4,758	4,821	5,593	+16.0%
Education	696	725	790	+8.9%
NASA—Science	4,610	4,706	4,442	+1%‡
Department of Energy—Science	3,797	4,018	4,722	+17.5%
Defense Department				
Basic research*	1,525	1,450	1,699	+17%
DARPA*	2,908	2,959	3,286	+11%
Department of Commerce				
NIST core labs	493	520	634	+22%
NOAA research	564	585	582	–0.5%
EPA—Science	561	542	550	+1.5%
FDA (includes user fees)	2,008	2,270	2,400	+5.7%
CDC—Overall	6,060	6,124	5,691	–7.1%
USDA—Competitive research	190	191	257	+34.6%
U.S. Geological Survey—Research	564	583	545	–6.5%

† Includes $300 million for Global AIDS Fund.
‡ Due to an accounting change from the 2008 budget.
* Excludes earmarks.

SOURCE: CONGRESS, FEDERAL AGENCIES

Published by AAAS

Not all the physical sciences are treated equally in the president's 2009 budget, however. The head of research and engineering at the Defense Department, John Young, successfully lobbied the White House for a 17% boost in basic science, to $1.7 billion, after activists complained that the military shouldn't have been left out of the 3-year-old ACI. But NASA's science chief, Alan Stern, didn't fare nearly as well: His $4.6 billion portfolio received only a 1% increase.

Despite the negligible growth, Stern sees room for a long-stalled mission to the outer planets as well as an ambitious multibillion-dollar flight to retrieve a sample from the surface of Mars. Agency officials hope to decide by the end of this year whether to send the outer-planets spacecraft to the Jupiter or Saturn system by the end of the decade. Past plans to send a large robot to Jupiter's moon foundered on high cost estimates that proved beyond NASA's means, and Stern says the Mars community may have to forgo other missions if it wants to focus on a sample return.

The moon also shines brightly in Stern's effort to encourage lower cost missions. One payload set to orbit in 2011 would study the moon's atmosphere for a mere $100 million. But these and other missions will never get off the ground, he warns, unless project managers keep a tight rein on costs.

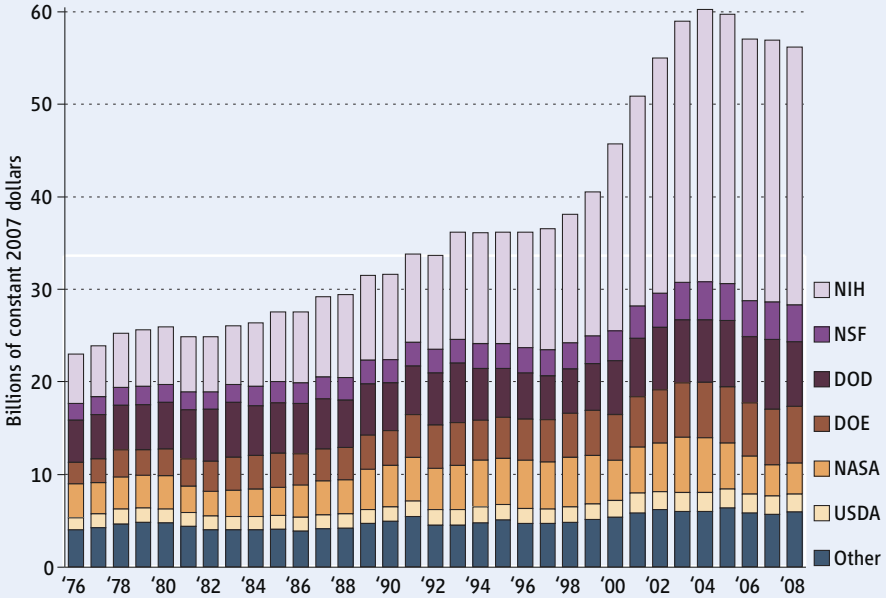
A similar policy of no cost overruns at NSF has left a $331 million Ocean Observatories Initiative high and dry after it was pulled from the queue of major new facilities pending a final review later this year. "It was a big surprise to us," says Steven Bohlen of the Consortium for Ocean Leadership. "NSF had given us every indication that we were ready to go" after the agency completed a preliminary review of the project in December. NSF Director Arden Bement says it's impossible to know a project's true costs until all aspects have been vetted, although he admits that delays will inevitably drive up the price. "It's a balancing act," he says. "We also need to follow our rules."

For NIH Director Elias Zerhouni, the first rule is "for NIH to keep its pipeline of new investigators up." He says his proposed 2009 budget should allow him to achieve that goal even if it means a slight drop in success rates for grant applications. And speaking as the head of an agency treading water, he reminds his constituents that the situation could be worse: The overall discretionary budget for NIH's parent agency, the Department of Health and Human Services, declines by 3.1% in the president's request. Small consolation, indeed.

–JEFFREY MERVIS

With reporting by Yudhijit Bhattacharjee, Jennifer Couzin, and Andrew Lawler.

Up, Up ... and Down



SOURCE: AAAS

Feast and famine. Overall federal spending on research hasn't kept up with inflation since 2004 despite the continuing growth of the budget in current dollars.

A Broken Record?

When you're hungry, it can be hard to remember an earlier era of plenty. And meager rations are more difficult to swallow if they come after promises of a feast. U.S. researchers may well be feeling such nutritional ambivalence as they ponder the Bush Administration's legacy toward science.

When presidential science adviser John Marburger proclaimed this week that the president's 2009 budget request to Congress (see main text) was "wonderful" for science, it marked the seventh consecutive year that he has delivered the same verdict. But the facts he used to buttress his argument in 2008 are quite different from those he marshaled in 2001 after George W. Bush took office.

One constant is that each year the overall request for federal research and development (more than half of which goes to developing weapons) sets a record—at the same time the rest of the domestic budget is being reined in. The 2009 request of $147 billion is up from $95 billion in the 2002 request. But the ingredients have changed, especially within the $29 billion for basic research.

The centerpiece of the president's first research budget was a 14% hike at the U.S. National Institutes of Health (NIH); in contrast, the U.S. National Science Foundation (NSF) was given only a 1.3% increase, and the U.S. Department of Energy's (DOE's) Office of Science was slated for a reduction. Fast-forward to the 2009 request, and it's DOE and NSF laying claim to double-digit increases, with NIH standing pat. In fact, the physical sciences were out of favor until the president's 2007 budget request packaged large increases for NSF, DOE science, and the National Institute of Standards and Technology under the rubric American Competitiveness Initiative.

If enacted by Congress, the request for NIH would mark its sixth year of declining budgets in real terms after a 5-year doubling that ended in 2003. That turnabout has led to prolonged howls from a biomedical community wondering if it might have been better off with steady increases rather than a cycle of boom and bust. And the big proposed boosts for NSF and DOE, although welcomed, don't erase the sting from a string of broken promises by the Administration and Congress—made first to NSF in 2002 and then to both agencies last summer—to put their budgets on the same doubling path that NIH enjoyed.

In fact, NSF's budget this decade may be a microcosm of how science as a whole has fared during the Bush years. After a windfall at the end of the Clinton Administration, NSF recorded solid gains for three straight years before suffering a drop in fiscal year 2005. Apart from a last-minute boost last year after the new Democratic majority took office, NSF's budget in the second Bush term has lost ground to inflation.

Republicans and Democrats alike say they want to do better. But until those words are replaced with new dollars, science agency officials will hope for the best while they prepare for the worst.

–JEFFREY MERVIS

Published by AAAS

Near-Term Energy Research Prospers

Business is booming at the U.S. Department of Energy's National Renewable Energy Laboratory (NREL) in Golden, Colorado, DOE's flagship facility for greening the nation's energy supply. Its budget has skyrocketed by 80% in the past 2 years, to $378 million. In addition to hiring more than 100 scientists, the lab has launched programs to integrate windmills into the nation's electrical grid, broadened work to facilitate solar panel manufacturing, and beefed up its biofuels research. "Everybody's busy; we're expanding," says Robert Thresher, who manages the lab's wind energy science program.

Although the president's 2009 request would keep research dollars at NREL steady, its recent rapid growth reflects the strong support in Congress for research aimed at tackling global warming by making near-term adjustments to the country's existing energy sources. NREL gets most of its money from DOE's $1.5 billion office of Energy Efficiency and Renewable Energy (EERE), which has received boosts of 27% and 18% in the past 2 years. At the same time, legislators have rejected increases of similar magnitude requested for the past 2 years by the president for DOE's $4 billion Office of Science, which typically funds research that is less likely to provide immediate answers to the nation's energy problems. Undeterred, President George W. Bush has asked for 17.5% more for the Office of Science in his 2009 budget.

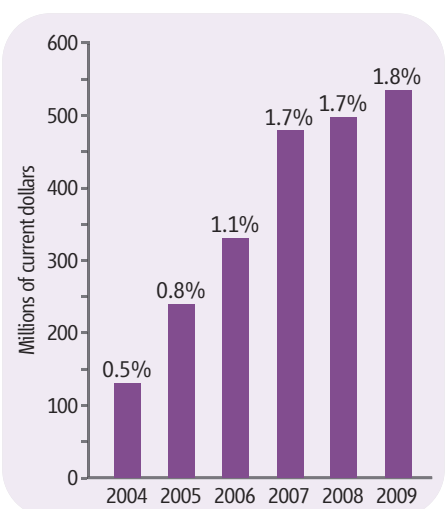
"I'm happy that EERE got a big boost [in 2008], but there are mid- and longer term research priorities that need to be attended to," says Nobelist Steven Chu, director of Lawrence Berkeley National Laboratory in California. Other energy researchers also lament the zeroing out in 2008 of a $150 million contribution to the $6 billion International Thermonuclear Experimental Reactor being built in Cadarache, France. The project, to design and build a prototype fusion reactor, represents a field whose goal of a cheap, sustainable source of energy has remained stubbornly out of reach for decades.

Filling the hopper. The National Renewable Energy Laboratory is expanding work on biofuels.

CREDIT: DOE/NREL/JIM YOST; SOURCE: NIH

NIH Hopes for More Mileage From Roadmap

While the budget of the U.S. National Institutes of Health (NIH) has languished for the past 5 years, one piece of the $29.3 billion agency—the NIH Roadmap—has taken off. Since its creation in 2003, the transinstitute initiative has grown to a half-billion-dollar-a-year program. And the Roadmap, which gets an 8% bump in NIH's 2009 request from President George W. Bush (see main text), remains Director Elias Zerhouni's signature effort to give patients a better chance of enjoying the fruits of basic research.



On the rise. The Roadmap accounts for a small but growing share of NIH's budget.

Some biomedical researchers love the Roadmap's emphasis on tools, translational research, and the opportunity to collaborate and stretch their wings. Others slam it, claiming it takes funding away from single-scientist, hypothesis-driven research. Congress has embraced the idea as a way to give the NIH director more control over the 27 institutes and centers by making it a permanent part of the agency.

The Roadmap has three main components—basic science, clinical, and research teams—comprising some 40 programs. They range from interdisciplinary training grants to the Molecular Libraries Initiative, a major project to develop chemical probes. The latter, says Elizabeth Wilder, acting associate director of the NIH Office of Portfolio Analysis and Strategic Initiatives, is an example of something whose "scope and size" exceeded the budget of the National Institute of General Medical Sciences, which otherwise might be a good home for it. The Roadmap is meant to be an incubator for projects that would eventually be adopted by an institute or terminated.

Supporters say some of the Roadmap's technology-intensive projects are already paying off. The 10 teams involved with the molecular libraries have found dozens of useful probes for studying basic biology and disorders such as Gaucher disease and schistosomiasis, says pharmacologist Bryan Roth of the University of North Carolina, Chapel Hill, who was part of a review panel last year. The effort has also set an example for other academic screening programs.

Some researchers have questioned the Roadmap's emphasis on big biology at the same time success rates for individual investigator-initiated R01 grants have plummeted. Zerhouni's answer is that the Roadmap constitutes only 1.8% of NIH's budget, that it funds many R01s, and that it has not shifted funds away from unsolicited research grants.

Other Roadmap pieces include the Pioneer Awards, given out on the basis of the track record of an investigator rather than for a specific project. Although the program has won praise for channeling a sliver of NIH's budget into high-risk research, the selection process was tweaked after women were shut out of the first round of nine winners (*Science*, 22 October 2004, p. 595). Other Roadmap ideas have required honing, too. NIH solicited proposals for translational research centers, for example, before deciding to use the money instead to revamp its clinical research support program.

In 2006, when Congress reauthorized NIH's programs, it enshrined the Roadmap by including language creating a permanent "common fund." But the law caps the fund's size at 5% of the overall NIH budget and limits its annual rise to no more than the rate of biomedical inflation.

–JOCELYN KAISER

Published by AAAS

Most scientists believe that the United States should invest in both near-term research on existing energy sources and long-range research to develop new ones, and DOE Under Secretary for Science Raymond Orbach says that Congress is making a mistake by not recognizing the importance of both approaches. "Without transformational [basic] discoveries, we don't have a very hopeful future" in energy, he says.

Aides on Capitol Hill say that Congress wound up embracing near-term energy projects because it was the right thing to do—and because legislators allocated limited funds to the most pressing needs. "There was no thinking 'Oh, we're doing this in solar, but we're not doing that,' " says an aide to Representative Pete Visclosky (D–IN), chair of the House panel that funds DOE. Emphasizing near-term over long-range research was difficult, says the aide, but it was a "conscious priority decision when you're in an energy crisis." The aide noted that U.S. government spending on applied energy research is only 40% of 1978 levels.

Legislators last year did agree to protect DOE's down payment on three $125 million bioenergy centers that combine academic-style genomic studies with industrial chemical engineering to find new biofuels. (DOE wants $75 million more for them in 2009.) And Orbach has reintroduced a program intended to bridge the divide—proposing $100 million in collaborative, multiscientist basic research funds in areas including energy storage, solid state lighting, and the computational analysis of underground CO_2—after canceling it when his 2008 budget fell far short of expectations (*Science*, 1 February, p. 554).

One likely flash point in the 2009 budget is the Advanced Research Projects Agency–Energy, which an influential 2005 National Academies' panel recommended as a way to support "out-of-the-box transformational research" in energy. Congress created the office last year, but DOE didn't request any money for it in 2009, as Orbach says the work would be "duplicative." Although he's not an appropriator, House Science and Technology committee chair Bart Gordon (D–TN) has pledged to push for 2009 funding for what he sees as a "small but aggressive" agency.

Although DOE spends $4 on long-term, fundamental physical science for every $3 on near-term, more applied energy research, the Bush Administration made a failed attempt to grow the latter in this week's budget request. *Science* has learned that the White House quietly asked DOE staff last summer to propose a roughly $500 million research initiative that could be packaged as DOE's contribution to climate change. What DOE submitted called for, among other things, innovations for existing and futuristic coal and nuclear plants, CO_2 sequestration studies, and expanded efforts to develop renewables.

In the end, however, White House budget officials slashed it by $343 million after questioning DOE's argument that research in coal and nuclear was likely to result in the biggest payoff in reducing carbon emissions. The Office of Management and Budget's funding levels "severely undermine … the President's stated policy on climate change," shot back Energy Secretary Samuel Bodman in a punchy 29 November 2007 letter. In the end, the White House put $200 million of DOE's increase into coal research and kept funding flat for renewables. Those levels "will help achieve a more secure and reliable energy future," says Bodman.

–ELI KINTISCH

Better sense. A pilot satellite to be launched in 2010 will carry key climate sensors.

CREDIT: BALL AEROSPACE

Earth Gets a Closer Look

In his last year in office, President George W. Bush wants Congress to beef up what scientists have called an anemic federal effort to scan Earth's atmosphere, land, and oceans from space. That's cheered researchers who rely on satellite-based remote sensing for a variety of studies, including monitoring global change, although they say more needs to be done.

The 2009 budget requests for NASA and the National Oceanic and Atmospheric Administration propose two new satellites to quantify soil moisture and ice sheets and the proposed restoration of instruments removed from the troubled National Polar-orbiting Operational Environmental Satellite System (NPOESS) program (*Science*, 31 August 2007, p. 1167). They closely follow several recommendations of the decadal study on earth observing issued last year by the National Research Council (NRC) of the U.S. National Academies, which called on the government to commit roughly $7 billion through 2020 to "renew its investment in Earth-observing systems."

Overall, NASA wants to spend $910 million over 5 years on five missions culled from a list of 15 in the decadal survey, with $103 million proposed for 2009. They've named two: a Soil Moisture Active-Passive mission, proposed for a 2012 launch, to help quantify soil's role in the global carbon cycle and help predict landslides, and ICESat-II, for 2015, which would measure the depth of ice sheets—crucial to calculate and forecast sea-level rise and forest canopy heights. NASA hasn't determined how much it will spend on each. In all, it's "a solid down payment on the recommended program," says biogeochemical modeler Berrien Moore of the University of New Hampshire, Durham, co-chair of the survey. "Looks like they listened to us somewhat," says climate modeler Warren Washington of the National Center for Atmospheric Research in Boulder, Colorado. Even so, fellow NRC panel co-chair Richard Anthes of the University Corporation for Atmospheric Research, also in Boulder, notes that NASA's proposed funding levels for Earth sensing are less than half the decadal recommendations.

Scientists are also applauding the Administration for beginning to address the impact of changes to NPOESS made when the Pentagon restructured the program in 2006. Agency officials said last week that they want to restore, at least in part, three of six sensors that had been stripped from the $12.5 billion program. The most recent step involves a sensor that would measure the reflected radiation from Earth, a crucial factor in quantifying global warming (*Science*NOW, 1 February). Agency officials propose replacing that sensor with a slightly less capable one for a pilot mission, called NPP, that is set for launch in 2010. "NPP's become a gap filler," says David Ryan of Northrop Grumman, which is building NPOESS.

–ELI KINTISCH

Gold standard. Silicon solar cells dominate the market, but new competitors are rising fast.

SOLAR ENERGY

Can the Upstarts Top Silicon?

Several nascent technologies are improving prospects for turning the sun's rays into electricity. The success of any one of them could mean a big boost for solar power

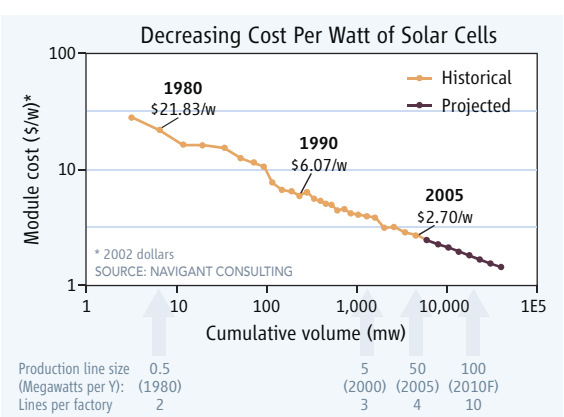
These are bright days for backers of solar power. The exuberance that previously pumped up dot-coms and biotech companies migrated in 2007 to solar energy, one of the hottest sectors in the emerging market for clean energy. Last year, solar energy companies around the globe hauled in nearly $12 billion from new stock offerings, loans, and venture capital funds. And although the markets have taken a bath in recent weeks due to investor fears about a coming recession in the United States, enthusiasm for solar's future remains strong. The industry is growing at a whopping 40% a year. And the cost of solar power is dropping and expected to rival the cost of grid-powered electricity by the middle of the next decade (see figure, right).

Still, there are clouds overhead. Solar power accounts for only a trivial fraction of the world's electricity. Silicon solar panels—which dominate the market with a 90% share—are already near their potential peak for converting solar energy to electricity and thus are unlikely to improve much more. A typical home's rooftop loaded with such cells can't produce enough power to meet the home's energy needs. That limitation increases the need for large-scale solar farms in sunny areas such as the American Southwest, which are far from large population centers. The bottom line is that the future of solar energy would be far brighter if researchers could make solar cells more efficient at converting sunlight to electricity, slash their cost, or both.

That's just what a new generation of solar-cell technologies aims to do. A raft of those technologies was on display here* late last year, as researchers reported how a broad array of recent advances in chemistry, materials science, and solid state physics are breathing new life into the field of solar-energy research. Those advances hold out the promise of solar cells with nearly double the efficiency of traditional silicon-based solar cells and of plastic versions that cost just a fraction of today's photovoltaics (PVs). "It's a really exciting time [in solar energy research]," says chemist David Ginger of the University of Washington, Seattle.

In the past few years, Ginger and others point out, solar researchers have hit upon several potential breakthrough technologies but have been stymied at turning that potential into solar cells able to beat out silicon. "The next couple of years will be important to see if we can overcome those hurdles," Ginger says. Although most of these novel cells are not yet close to commercialization, even one or two successes could dramatically change the landscape of worldwide energy production.

* Materials Research Society meeting, Boston, Massachusetts, 26–30 November 2007.



Heading for parity. Solar electricity still costs about five times as much as electricity from coal. But many experts expect economies of scale could close the gap by 2015.

Minding the gap

Beating silicon is a tall order. Although the top lab-based silicon cells now convert about 24% of the energy in sunlight into electricity, commercial cells still reach only 15% to 20%. In such traditional solar cells, photons hitting the silicon dump their energy into the semiconductor. That excites electrons, kicking them from their staid residence in the so-called valence band, where they are tightly bound to atoms, into the higher-energy conduction band, where they lead a more freewheeling existence, zipping through the material with ease. But if photons don't have enough energy to push electrons over this "band gap," the energy they carry is lost as heat. So is any energy photons carry in excess of the band gap. Given the sun's spectrum of rays and the fact that only certain red photons have the amount of energy that closely matches silicon's band gap, single silicon cells can convert at most 31% of the energy in sunlight into electricity—a boundary known as the Shockley-Queisser limit.

Engineers can boost the efficiency with a number of conventional strategies. One is to layer several light-absorbing materials that capture different portions of the solar spectrum—for example, by having one cell that absorbs mostly blue photons, while others absorb yellow and

CREDITS: (TOP) COURTESY OF DUPONT; (BOTTOM) SOURCE: APPLIED MATERIALS

Published by AAAS

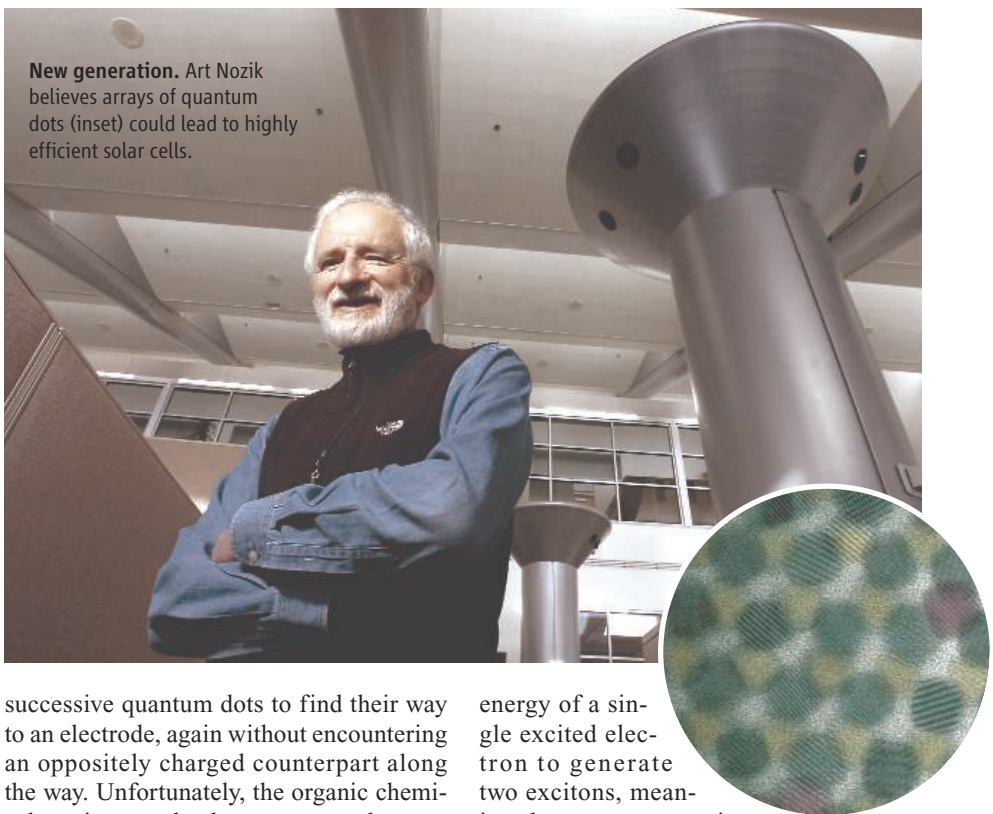
red photons. But such "tandem" cells are expensive to produce and thus are currently used primarily for high-end applications such as space flight.

But there may be other ways to capture more energy from the sun. One strategy that has drawn a lot of attention in recent years is to find materials that generate multiple electronic charges each time they absorb a photon. Traditional silicon solar cells generate just one. In them, a layer of silicon is spiked with impurity atoms so that one side attracts negatively charged electrons, while the other attracts positively charged electron vacancies, known as holes. Most light is absorbed near the junction between the two layers, creating electrons and holes that are immediately pulled in opposite directions.

In 1997, however, chemist Arthur Nozik of the National Renewable Energy Laboratory (NREL) in Golden, Colorado, and colleagues predicted that by using tiny nanosized semiconductor particles called quantum dots to keep those opposite charges initially very close together, researchers could excite two or more electrons at a time. A paired electron and hole in close proximity, they reasoned, increases a quantum-mechanical property known as the Coulomb interaction. The greater this interaction, the more likely it is that an incoming energetic photon with at least twice the band-gap energy will create two electron-hole pairs with exactly the energy of the band gap—instead of one electron-hole pair with excess energy above the band gap, the other possible outcome that quantum mechanics allows. At least that's how Nozik and his colleagues see it. Several other models of multiple-electron excitation exist, and theorists are still debating just what is behind the effect.

Four years ago, researchers led by Victor Klimov of Los Alamos National Laboratory in New Mexico reported the first spectroscopic evidence showing that multiple electron-hole pairs, known as excitons, were indeed generated in certain quantum dots. Nozik's team and others have since found the same effect in silicon and other types of quantum dots. And according to calculations by Nozik and NREL colleague Mark Hanna, the multiple exciton generation (MEG)–based solar cells hit with unconcentrated sunlight have a maximum theoretical efficiency of 44%. Using special lenses and mirrors to concentrate the sunlight 500-fold, they predicted, could boost the theoretical efficiency to about 80%—twice that of conventional cells hit with concentrated sunlight.

But reaching those higher efficiencies isn't easy. "One big hang-up is that no one has yet shown that you can extract those extra electrons," Nozik says. To harvest electricity, researchers must first break apart the pairs of electrons and holes, using an electric field across the cell to attract the opposite charges. That must happen fast, as electrons in excitons will collapse back into their holes within about 100 trillionths of a second if left side by side. If those charges can be separated, they must hop between successive quantum dots to find their way to an electrode, again without encountering an oppositely charged counterpart along the way. Unfortunately, the organic chemical coatings used to keep quantum dots stable and intact push the particles apart from one another, slowing down the charges.



New generation. Art Nozik believes arrays of quantum dots (inset) could lead to highly efficient solar cells.

Still, Nozik's group seems to be making progress. At the Materials Research Society meeting, Nozik reported preliminary results on solar cells made with arrays of lead selenide quantum dots. In such cells, a layer of quantum dots, and their organic coats, is spread between two electrodes. According to Nozik, spectroscopic studies indicate that two or three excitons are generated for every photon the dots absorb. And the researchers managed to separate the charges and get many of the electrons out, boosting the efficiency of the solar cells to about 2.5%, up from 1.62% from previous MEG-based cells.

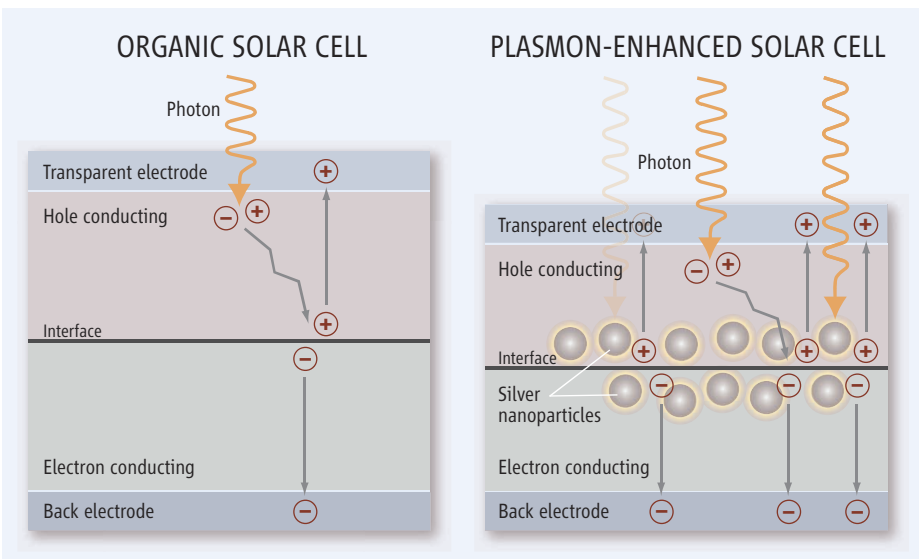
To boost that efficiency further, Nozik says, one key will be to pack quantum dots closer together and in more regular arrays, making it easier for electronic charges to hop from one dot to the next to the electrodes where they are collected. Nozik's group is already experimenting with strategies for doing that, such as shrinking the organic groups that coat each dot and keep them separated from one another. Another needed improvement will be to find quantum dot materials better at generating multiple excitons. Nozik's lead selenide dots, for example, must be hit with about 2.5 times the energy of a single excited electron to generate two excitons, meaning that extra energy is wasted. In the November 2007 issue of *Nano Letters*, however, Klimov and his colleagues reported that dots made from indium arsenide generate two excitons almost as soon as the energy of the incoming photons exceeds twice the band gap.

Other groups are hoping to use quantum dots as steppingstones to cross the band gap in conventional semiconductor materials. The idea is to seed a semiconductor with an array of quantum dots, which will absorb photons that have too little energy to raise electrons above the band gap. The photons would excite electrons in the quantum dots to an intermediate level between the valence and conduction bands, then a hit from a second low-energy photon would boost them the rest of the way into the conduction band.

CREDITS: MARTY CAIVANO/DAILY CAMERA; (INSET) A. NOZIK

Theoretical work by Antonio Luque of the Universidad Politécnica de Madrid in Spain suggests that such cells could achieve a maximum efficiency of 63% under concentrated sunlight. But here, too, the potential has been hard to realize. In practice, adding quantum dots to materials such as an alloy of gallium arsenide seems to cause more losses than gains; the quantum dots also seem to attract electrons and holes and promote their recombination, thus losing the excess energy as heat.

Last year, Stephen Forrest and Guodan Wei, both of the University of Michigan, Ann Arbor, suggested a way around that problem: designing energetic barriers into their solar cells that discourage free charges from migrating to the quantum dots. At the meeting, Andrew Gordon Norman of NREL reported that his team has managed to grow such structures. The cells didn't outperform conventional GaAs cells, because too few quantum dots were packed into the structure to absorb enough low-energy photons to offset recombination losses. But Norman says he's working on solving that problem.



Better reception. In an organic solar cell, sunlight frees an electron (–) and an electron vacancy, or hole (+), which migrate to the border between different materials and then to oppositely charged electrodes (*left*). Adding metal nanoparticles (*right*) increases the light absorption and the number of charges generated.

A silver lining

Many of the approaches to boosting the efficiency of solar cells require expensive materials or manufacturing techniques, so they are likely to increase capital costs. Some groups are exploring low-cost alternatives: light-absorbing plastics or other organic materials that can be processed without the expensive vacuum deposition machines most inorganics require. Unfortunately, organics waste much of the incoming light because they typically absorb only a relatively narrow range of frequencies in the solar spectrum. The key to boosting their efficiency, some groups believe, could be precious metals.

A layer of tiny silver or other metal nanoparticles added to a solar cell encourages an effect known as surface plasmon resonance, in which light triggers a collective excitation of electrons on the metal's surface. This causes the nanoparticles to act like antennas, capturing additional energy and funneling it to the active layer of the material to excite extra electrons (see figure). At the meeting, electrical engineer Peter Peumans of Stanford University in Palo Alto, California, reported that when he and his colleagues added a layer of silver nanoparticles atop a conventional organic solar cell, they increased the efficiency of the device by 40%. Even though the overall efficiency of Peumans's devices is still dismally low—less than 1%—Ginger says the big jump in efficiency is "very promising."

Peumans notes that the silver nanoparticles work best when placed at the interface between two semiconducting layers in organic solar cells, one of which preferentially conducts electrons, the other, holes. In organic solar cells, excitons must migrate to just such an interface so that they can split into separate charges, which are then steered to opposite electrodes.

Other researchers have found in recent years that they can increase the efficiency of their organic solar cells by expanding the surface area of this interface. Instead of having flat layers lying atop one another like pages in a book, they create roughened layers that interpenetrate one another, a configuration known as a bulk heterojunction. Last year, researchers led by physicist Alan Heeger of the University of California, Santa Barbara, reported that they could use low-cost polymers to create tandem bulk heterojunction solar cells with an overall energy conversion efficiency of 6.5% (*Science*, 13 July 2007, p. 222). At the time, Heeger said that he expected that further improvements to the cells would propel them to market within 3 years.

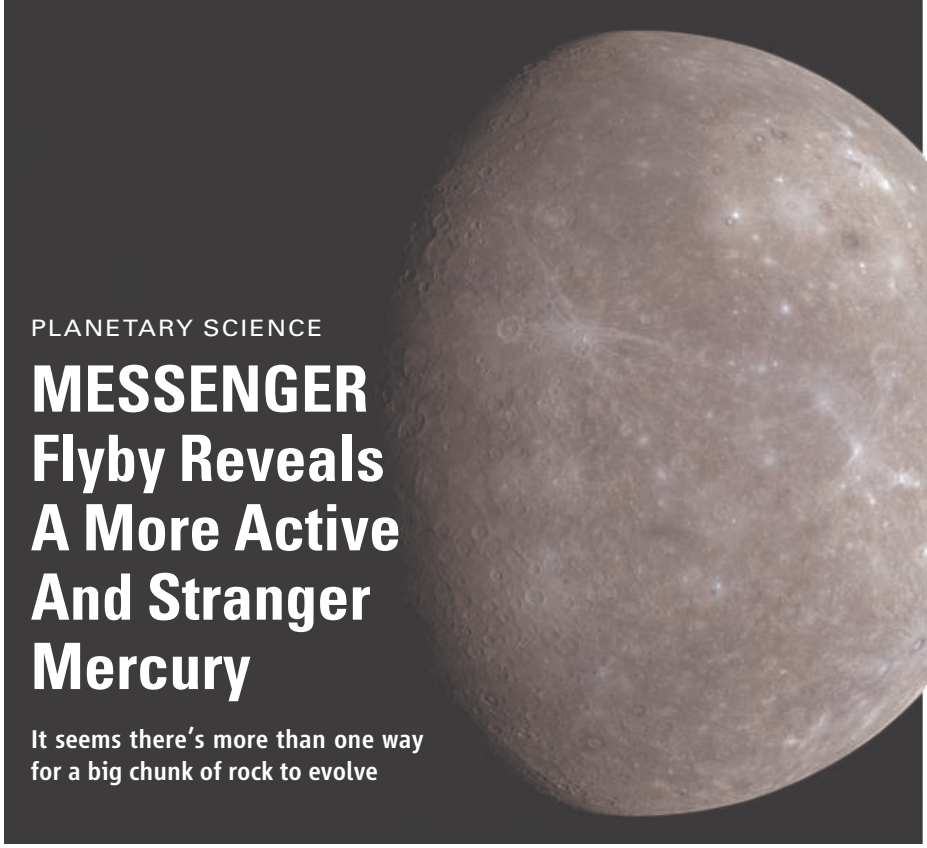
Researchers are pursuing several strategies to improve these cells. One approach that may pay off down the road, Peumans says, is to incorporate metal nanoparticles into the random surface in these solar cells. That's not likely to be easy, he adds. But it may be possible to outfit the nanoparticles with chemical tethers that encourage them to bind to tags designed into the interface of the material. In theory, Peumans says, that would offer researchers the best of both worlds.

In addition to improving solar cell efficiencies, researchers and companies are also working on a host of technologies to make them cheaper. Nanosolar in San Jose, California, for example, has spent millions of dollars perfecting a new roll-to-roll manufacturing technology for making solar cells from thin films of copper indium gallium selenide atop a metal foil. Although they haven't reported the efficiency of their latest cells, they began marketing them in December 2007. Konarka, another roll-to-roll solar cell company in Lowell, Massachusetts, is working on a similar technology with plastic-based PVs. Other groups, meanwhile, are pushing the boundaries on everything from replacing quantum dots with nanowires that can steer excited charges more directly to the electrodes where they are harvested to using modified ink-jet printers to spray films of quantum dots and other solar-cell materials.

For now, there appears to be no shortage of ideas about creating new high-efficiency, low-cost cells. But whether any of these ideas will have what it takes to beat silicon and revolutionize the solar business remains the field's biggest unknown. "There are a lot of ways to beat the Shockley limit on paper, but it's difficult to realize in the real world," Nozik says. So far, it's not for want of trying.

–ROBERT F. SERVICE

CREDIT: ADAPTED FROM M. BRONGERSMA, P. PEUMANS, S. FAN; ILLUSTRATION: L. CREVELING/*SCIENCE*



PLANETARY SCIENCE

MESSENGER Flyby Reveals A More Active And Stranger Mercury

It seems there's more than one way for a big chunk of rock to evolve

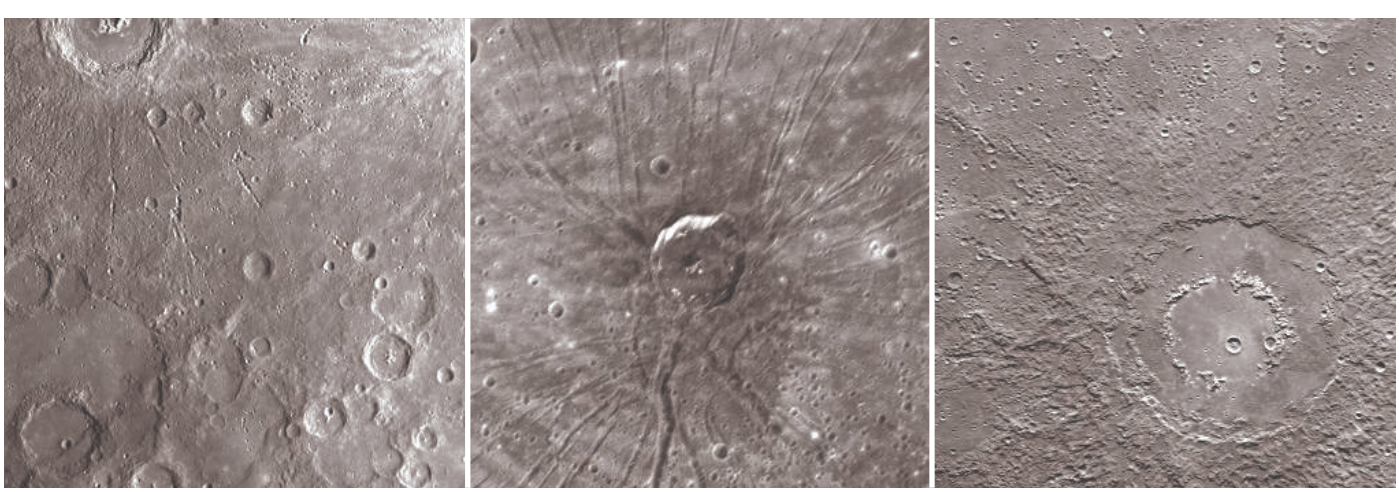
WASHINGTON, D.C.—The untrained eye might have trouble distinguishing the latest images from Mercury—released last week at a NASA press conference here—from countless images of Earth's moon, but don't tell that to Sean Solomon. Mercury "was not the place we expected," says Solomon, principal investigator of the MESSENGER mission to Mercury. "It was not the moon." The first close look at the innermost planet in 33 years and the first look ever at one-third of it revealed a new side to the innermost planet: much more volcanism than seen before, deeply excavating impact craters, and—unique in the solar system—"The Spider."

The Mariner 10 spacecraft last flew by Mercury in 1975, returning images suggesting that lava once flowed across the surface, at least in places. But volcanism "wasn't accepted by everyone," says imaging team member Louise Prockter of Johns Hopkins University's Applied Physics Laboratory in Laurel, Maryland. Now, "there's very little doubt there has been widespread volcanic resurfacing of Mercury." She pointed to impact craters hundreds of kilometers across with floors so smooth that they must have been partially filled by lava (below, left, double pair in lower left; below, right, concentric pair lower right). Team member Robert Strom of the University of Arizona, Tucson, also found that the side of Mercury seen by Mariner 10 turns out to be more heavily cratered by impacts than the side seen for the first time by MESSENGER. That means that lava has flooded the MESSENGER side even more extensively than the other side. "There's been a lot of volcanic activity on Mercury," says Strom.

The moon has its volcanic flooding, too—witness the dark "seas," or maria, that shape the man (or woman) in the moon—but MESSENGER found a mercurial variation on such light-dark patterning. Caloris is a huge—1550-kilometer-wide—impact basin glimpsed by Mariner 10 but now seen in its entirety by MESSENGER (left, in false color, brighter spot in upper right of disk). On the moon, such giant impact basins were often filled with dark lava to form maria, but Caloris has the opposite pattern. Its interior is lighter and is surrounded by a darker ring. Perhaps the Caloris impact excavated deep, lighter-colored rock and left it at the surface without flooding it with lava, says Solomon, director of the Carnegie Institution of Washington's Department of Terrestrial Magnetism in Washington, D.C. Including smaller craters with distinctive dark rims, "we've got a variety of natural drill holes into Mercury's interior," says Solomon.

Then there's The Spider (below, center). More than 50 troughs radiate from near the center of Caloris where a 40-kilometer-wide crater has formed. Whether the crater has anything to do with the radiating troughs, Prockter can't say; no one has ever seen anything like The Spider. One possibility is that the formation of Caloris somehow created a plume of molten rock that rose beneath the basin's center, bulging the basin floor upward and cracking the crust to form the troughs. The crater would then have been an accidental impact. MESSENGER returns in October for another look at Mercury on its way to entering orbit in 2011.

–RICHARD A. KERR



CREDITS: NASA/JOHNS HOPKINS UNIVERSITY APPLIED PHYSICS LABORATORY/CARNEGIE INSTITUTION OF WASHINGTON

New views. MESSENGER revealed craters smoothed by volcanic flooding (*left* and *right*) and a cryptic feature dubbed The Spider (*center*).

Splashing around. Hyenas enjoy playing in a tub of water at the Berkeley colony.

WILDLIFE BIOLOGY

Berkeley Hyenas Face An Uncertain Future

Long studied for their unusual anatomy and behavior, the world's only research colony of spotted hyenas faces a funding crisis

BERKELEY, CALIFORNIA—The foggy, eucalyptus-studded hills above the San Francisco Bay are a world away from the African savanna, but the spotted hyenas that live here seem content. On a recent afternoon, they excitedly jostled one another to get a better look—and sniff—at some visitors passing by their enclosure at the Field Station for the Study of Behavior, Ecology, and Reproduction at the University of California (UC), Berkeley. Pink tongues darted through the chainlink fence to lick a keeper's outstretched hand. "These animals really get to you," she says. "It doesn't take long for your heart to be stolen."

Long maligned in myths and movies as dangerous freeloaders with a high-pitched giggle, spotted hyenas are in fact intelligent animals with fascinating biology and behavior, says Stephen Glickman, a psychologist and integrative biologist at UC Berkeley and director of the field station, home to the only captive hyena research colony in the world. Since establishing the colony in 1985, Glickman has worked hard to repair the reputation of his charges and to attract the interest of other scientists.

Many have been hooked by the unique reproductive anatomy of the female spotted hyena: She sports an elongated clitoris roughly the size of the male's penis, through which she urinates, mates, and gives birth. Why this structure evolved, how it develops in the embryo, and what it might have to do with the female's dominant status in hyena society are among the questions that intrigue biologists. The animals also have digestive and immune systems that enable them to swallow chunks of bone that would give a lion indigestion and to feast on rotted, anthrax-ridden carcasses with no ill effects.

Other researchers say that by virtue of his enthusiasm and easygoing manner, Glickman has created a remarkably diverse network of scientists who use the hyena colony to pursue such questions. "You've got a dozen or more senior investigators across the entire range of biological sciences collaborating without any kind of formal arrangement," says Elihu Gerson, an independent San Francisco–based sociologist who is studying the dynamics of the collaboration. "It's extraordinarily unusual."

But that collaboration now faces an uncertain future. Last fall, the U.S. National Institute of Mental Health (NIMH), which funded the colony for 22 years through an R01 grant to Glickman, did not renew his grant. Despite positive comments about the project's recent work, the priority score was too low. "The basic problem is that there's no precedent I know of for a research grant with the number of collaborators we have and the variety of projects," Glickman says. An emergency $200,000 grant from the U.S. National Science Foundation (NSF) will keep the colony going for another 15 to 18 months while he and colleagues look for a longer term solution. Meanwhile, he's had to downsize the colony by about a third, arranging for 10 hyenas to be sent to zoos and animal parks and euthanizing two older animals.

Puzzling endowment. The female hyena's elongated clitoris raises many questions for biologists.

"I can't stand the thought of them shutting this down," says Kay Holekamp, a former student and longtime collaborator of Glickman's, now a behavioral ecologist at Michigan State University in East Lansing. "This is such an invaluable resource; it would be a tragedy if it were lost," says molecular endocrinologist Geoffrey Hammond of the University of British Columbia in Vancouver, Canada.

Crazy anatomy

In the break room at the field station, Glickman turns on a small television and plays a video montage of hyena behavior at the colony. Silver-haired and soft-spoken, with a voice that betrays a hint of his Brooklyn upbringing, Glickman narrates as a female hyena gives birth to a 2-kilogram cub. This cringe-inducing process—imagine pushing a golf ball through a soda straw—makes the downside of the female hyena's strange anatomy abundantly clear. Hyena moms typically have two cubs in a litter, and about 60% of cubs born to a first-time mom are stillborn, Glickman says. The cub in the video survives, but before it can free itself from the amniotic sac, a sister, born just minutes earlier, attacks with a bite to the back of the neck and a vigorous shake. "This little brain is organized for aggression at birth," Glickman says.

For decades, researchers have suspected that the female hyena's heightened aggression and masculinized anatomy might share common biological roots. "It looks as if the female hyenas have traded off certain reductions in fertility for the ability to be more aggressive and dominate males and feed themselves and their offspring in tight times," Glickman says. That hypothesis has driven work at the colony from the beginning.

In the early 1980s, a UC Berkeley graduate student named Laurence Frank was studying

CREDITS (TOP TO BOTTOM): CHRISTINE DREA; LAURA SMALE

Published by AAAS

wild hyenas in Kenya. He'd noticed that although both sexes hunt, the females drove off the males once a kill was made. (Despite their reputation as scavengers, hyenas kill most of what they eat.) As the kill was consumed and the remaining share shrunk, high-ranking females drove off subordinates. Cubs of high-ranking females were allowed to feed, but cubs of low birth were shoved off. The female dominance hierarchy was clearly central to hyena society, says Frank, now a conservation biologist with the Wildlife Conservation Society.

Frank, as others before him did, wondered whether high levels of male sex hormones might prime the female hyena brain for aggression and account for her "crazy anatomy." Frank wanted to test the idea by injecting antiandrogen drugs but soon concluded this wasn't feasible in wild hyenas. He began looking into alternatives. By this time, Glickman, who was on Frank's Ph.D. thesis committee, had become interested, and he submitted a grant proposal to NIMH to fund a research colony at the field station, a few kilometers from the main campus. "It was really a pretty wild-ass idea to suggest that a large, dangerous carnivore would make a convenient lab animal," Frank says. "To our utter amazement, we got the grant."

In 1985, Frank, Glickman, and co-workers arranged export permits and flew to Kenya, where they paid Maasai tribesmen to lead them to hyena dens with young cubs. In two trips, they collected 20 cubs, which they bottle-fed and kept in their tents at night until they made the trip back to California.

In a series of experiments published in the early 1990s, Glickman and colleagues established that hyena fetuses are exposed to unusually high androgen levels in the womb. In other mammals, this causes females to develop a penislike structure. Androgens also spur aggression, and an explanation seemed near at hand.

But a seminal experiment published in 1998 revealed that androgens aren't the whole story. Glickman and colleagues administered androgen-blocking drugs to pregnant female hyenas—a treatment that prevents penis formation in males in other mammals. The researchers predicted that the drugs would have the same effect on hyena males and would prevent a pseudopenis from forming in females. Neither prediction came true.

"That was an extraordinary finding," says Gerald Cunha, an emeritus developmental biologist and anatomist at UC San Francisco (UCSF) who collaborates with Glickman. The hyenas have forced biologists to think beyond the dogma on genital development, which essentially says that if a fetus gets androgens, it develops a penis, and if not, it develops a vaginal opening, Cunha says: "You're going to have to come up with something creative to explain this unusual set of circumstances."

New directions

A full account of how the female pseudopenis develops has eluded the hyena researchers so far, but other dogma-challenging results have emerged. One study suggested that the female sex hormone estradiol plays a role in the development of the male hyena's penis, a finding that may have implications for understanding a common birth defect in boys, says Laurence Baskin, the chief of pediatric urology at UCSF. In a condition called hypospadias, the urethra exits the underside of the penis instead of at the tip. The condition happens in approximately one in 125 boys, Baskin says, and something similar occurs in male hyenas born to moms given drugs that block estrogen synthesis. The cause of hypospadias isn't known, Baskin says, but the hyena findings suggest that there may be a previously unrecognized estrogen-sensitive phase of male sexual development. If so, it's possible that exposure to certain chemicals in the environment during that period could have a role in causing hypospadias, Baskin says.

Baskin says work with hyenas has already had clinical applications for another condition he treats: congenital adrenal hyperplasia. In about one in 10,000 girls, a malfunctioning enzyme results in abnormally high androgen levels and development of a penislike structure that is similar to the female hyena's pseudopenis. Baskin recently collaborated with the Berkeley team to map the nerves in the hyena pseudopenis and says the findings have helped refine the surgeries he does to restore more typical female anatomy in girls with the condition. "It's already paid off in terms of … preserving their sexual function," he says.

Although prenatal androgens apparently aren't necessary for development of the female's pseudopenis, recent evidence suggests that they do influence hyena behavior. In 2006, Holekamp and colleagues reported in *Nature* that high-ranking females in the wild have higher androgen levels late in pregnancy than do low-ranking females and that cubs born to these androgen-fueled moms exhibit more aggression. Preliminary findings from the Berkeley colony back this up: Females treated with antiandrogen drugs in utero seem to display less aggression toward males as adults.

Such studies may also provide insights into how hormones set up sex differences in the brain, says behavioral neuroendocrinologist Nancy Forger of the University of Massachusetts, Amherst. Forger and colleagues have been examining male and female hyena brains from the Berkeley colony for differences in the distribution of vasopressin, a hormone linked to aggressive behavior. Vasopressin expression differs between males and females in a consistent pattern across a wide range of vertebrates, Forger says. "So far, we can say that hyenas don't show that difference, and the pattern may even be reversed."

She says she's concerned about the outlook for the colony. "There's never been anything like this before, and I don't think there ever will be again."

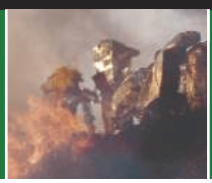
Glickman has been busy investigating funding options at NSF and at the National Institutes of Health, including the resource grants used to fund chimpanzee colonies and other shared facilities. But the uniqueness of the project—not to mention the research subjects—means it's not a perfect fit for many of the existing funding mechanisms, Glickman says. "It's tough because we don't fit the template."

–GREG MILLER

Aggressive by nature. Complementary studies in wild and captive hyenas are helping researchers unravel the links between androgens and aggression.

CREDIT: LAURA SMALE

Risky business 728



Balancing CO_2 and economics in China 730

LETTERS | BOOKS | POLICY FORUM | EDUCATION FORUM | PERSPECTIVES

LETTERS

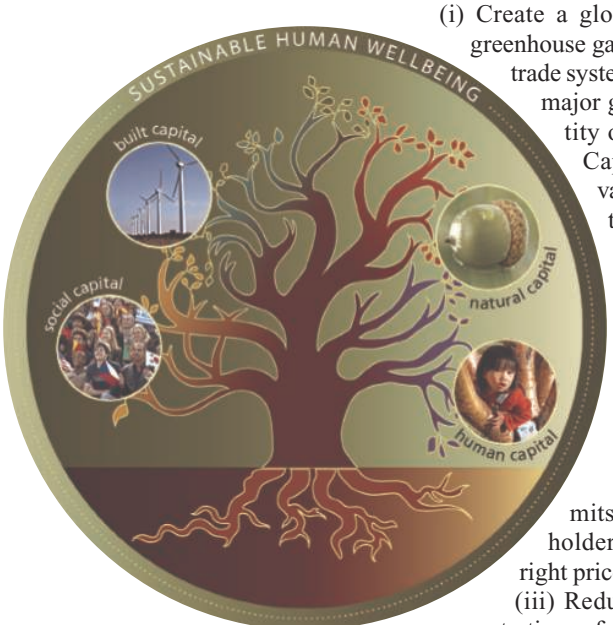
edited by Jennifer Sills

Creating an Earth Atmospheric Trust

STABILIZING CONCENTRATIONS OF GREENHOUSE GASES IN THE EARTH'S ATMOSPHERE AT A level that will control climate change will require drastic departures from business as usual. Here, we introduce one response to this challenge that may seem visionary or idealistic today, but that could become realistic once we reach a tipping point that opens a window of opportunity for embracing major changes.

The core of this system is the idea of a common asset trust. Trusts are widely used and well-developed legal mechanisms designed to protect and manage assets on behalf of specific beneficiaries. Extending this idea to the management and protection of a global commons, such as the atmosphere, is a new but straightforward extension of this idea. Because the atmosphere is global, the Earth Atmospheric Trust would be global in scope; however, initial implementation at a regional or national scale may be necessary. We provide an outline of the steps that must be taken to create and manage such a system.

(i) Create a global cap-and-trade system for all greenhouse gas emissions. We believe a cap-and-trade system is preferable to a tax, because the major goal is to cap and reduce the quantity of emissions in a predictable way. Caps set quantity and allow price to vary; taxes set price and allow quantity to vary.



Balancing act. Sustainable human well-being depends on a balance of built, human, social, and natural capital assets. An Earth Atmospheric Trust might better manage the atmospheric commons to control climate change.

(ii) Auction off all emission permits, and allow trading among permit holders. This essential act will send the right price signals to emitters.

(iii) Reduce the cap over time to stabilize concentrations of greenhouse gases in the atmosphere at a level equivalent to 450 parts per million of carbon dioxide (or lower).

(iv) Deposit all the revenues into an Earth Atmospheric Trust, administered by trustees serving long terms and provided with a clear mandate to protect Earth's climate system and atmosphere for the benefit of current and future generations.

(v) Return a fraction of the revenues derived from auctioning permits to all people on Earth in the form of an annual per capita payment. This dividend will be insignificant to the rich but will be enough to be of real benefit to many of the world's poor people. At the current annual rate of global emissions of 45 gigatons CO_2 equivalent and an auction price of $20 to $80 per ton, the Trust's total annual revenues would be $0.9 to $3.6 trillion. If half the revenues were returned equally to all 6.3 billion people, payment would amount to $71 to $285 per capita per year.

(vi) Use the remainder of the revenues to enhance and restore the atmospheric asset, to encourage both social and technological innovations, and to administer the Trust. These funds could be used to fund renewable energy projects, research and development on new energy sources, or payments for ecosystem services such as carbon sequestration.

No system is perfect. A system designed on these general principles would be fair; it would be efficient and relatively immune to political manipulation, and it would help to alleviate global poverty.

We encourage those interested in adding their name to a growing list of supporters of this idea to visit www.earthinc.org/earth_atmospheric_trust.php.

PETER BARNES,[1] ROBERT COSTANZA,[2] PAUL HAWKEN,[3] DAVID ORR,[4] ELINOR OSTROM,[5,6] ALVARO UMAÑA,[7] ORAN YOUNG[8]

[1]Tomales Bay Institute, Point Reyes Station, CA 94956, USA. [2]Gund Institute for Ecological Economics, University of Vermont, Burlington, VT 05405, USA. [3]Natural Capital Institute, Sausalito, CA 94965, USA. [4]Lewis Center for Environmental Studies, Oberlin College, Oberlin, OH 44074, USA. [5]Workshop in Political Theory and Policy Analysis, Indiana University, Bloomington, IN 47408, USA. [6]Center for the Study of Institutional Diversity, Arizona State University, Tempe, AZ 85287, USA. [7]InterAmerican Development Bank, Washington, DC 20577, USA. [8]Donald Bren School of Environmental Science and Management, University of California, Santa Barbara, CA 93106, USA.

The Latest Buzz About Colony Collapse Disorder

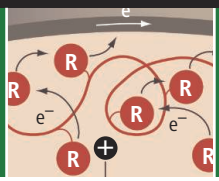
THE REPORT "A METAGENOMIC SURVEY OF microbes in honey bee colony collapse disorder" (D. L. Cox-Foster *et al*., 12 October 2007, p. 283) identified Israeli acute paralysis virus (IAPV) as a putative marker for colony collapse disorder (CCD). It also purports to show a relationship between U.S. colony declines as early as 2004 and importations of Australian honeybees. We believe these links are tenuous for several reasons: (i) Importations of Australian honeybees to the United States did not commence until 2005. (ii) No evidence is presented for a causal link between IAPV and CCD. Koch's postulates, as modified for

CREDIT: SARAH MONTGOMERY/GUND INSTITUTE FOR ECOLOGICAL ECONOMICS

Published by AAAS

Human dwarfism

732



Bendable batteries

737

viruses by Rivers (*1*), were not demonstrated. Several CCD colonies were free of IAPV and the "shivering phenotype," and the death of bees close to the hive associated with IAPV in Israel (*2*) was not observed in CCD colonies. (iii) The case definition for CCD is ambiguous, and the symptoms are indistinguishable from those of the normal winter colony collapse reported in the United States since the late 1980s and attributed to Nosema infection and/or the secondary effects of varroa (*3*). Many scientists are unconvinced that CCD is a new disorder (*4*). (iv) Members of the Kashmir bee virus complex (including IAPV) persist as nonacute (harmless) infections in honeybee colonies (*5*). They are opportunists and only cause acute infection in association with a primary pathogen (such as *Nosema apis*) (*6*). (v) Neither CCD nor large-scale, unexplained mortality events have occurred in the Australian bee industry. The implication that the absence of varroa in Australia may explain the absence of CCD is incorrect. Modeling has shown that fast-replicating viruses (such as IAPV) cannot cause colony collapse when associated with varroa (*7*). (vi) Other countries reporting CCD (such as Greece, Poland, and Spain) have not imported bees from Australia.

A followup paper by coauthors on the *Science* Report has now been published in the *American Bee Journal* (*8*) describing isolation of IAPV from specimens of *Apis mellifera* collected within the United States in 2002. This is more than 2 years prior to the commencement of importation of Australian packaged bees. It would now be appropriate for the authors of the *Science* Report to issue a retraction of the claims linking CCD to importation of Australian bees.

Future collaboration between United States and Australian scientists can only lead to a better understanding of colony collapse and IAPV and result in more secure trade for package honeybees to meet the growing demands of the United States pollination industry.

DENIS ANDERSON[1] AND IAIN J. EAST[2]

[1]CSIRO Entomology, Canberra, ACT 2601, Australia. [2]Office of the Chief Veterinary Officer, Department of Agriculture, Fisheries and Forestry, Barton, ACT 2600, Australia.

References

1. T. M. Rivers, *J. Bacteriol*. **33**, 1 (1937).
2. E. Maori, E. Tanne, I. Sela, *Virology* **362**, 342 (2007).
3. M. Sanford, *Bee Cult.* **135**, 38 (2007).
4. E. Stokstad, *Science* **316**, 970 (2007).
5. D. L. Anderson, A. J. Gibbs, *J. Gen. Virol*. **69**, 1617 (1988).
6. D. L. Anderson, *Am. Bee J*. **131**, 767 (1991).
7. S. J. Martin, *J. Appl. Ecol*. **38**, 1082 (2001).
8. Y. Chen, J. D. Evans, *Am. Bee J.* **147**, 1027 (2007).

Response

IN THEIR LETTER, ANDERSON AND EAST SUGgest that CCD is an ambiguous disorder consistent with normal winter losses. We do not agree. CCD is characterized by a rapid loss of adult bees; excess brood, in all stages, abandoned in the hive; low levels of varroa; and a lack of dead bees in or near the hive. In CCD, levels of varroa do not reach those associated with normal winter losses, distinguishing CCD from colony declines attributed to parasitic mites. Although Anderson and East imply that we claim to have determined the cause of CCD, the final paragraph of our paper states, "We have not proven a causal relationship between any infectious agent and CCD...."

The notion that all viruses within a phylogenetic group can only present as a single syndrome is invalid. Differences in virulence are common even among closely related viruses (*1*) and may reflect differences in the host, the microbe, or both. Indeed, genetically distinct lineages of IAPV sequences found in Israel differ in pathogenicity (*2*). With regards to varroa, most evidence points to a link between bee viruses and varroa and indicates that varroa acts as both a vector and an activator of latent viruses (*3*). Finally, given work from Anderson describing "Disappearing Disorder," it is not clear that Australia is free of unexplained losses of honey bees (*4*).

We appreciate that research on products important to international trade may lead into politically and economically sensitive territory. However, trade issues should not color research. Anderson and East note that subsequent work from our group indicates the presence of IAPV in bees in the United States as early as 2002 (*5*), predating recognition of CCD or the formal importation of bees from Australia. Infectious agents, including IAPV, do not respect national boundaries. IAPV is not confined to the United States or Australia. It has also been found in bees in Israel and royal jelly from Manchuria. We anticipate that with the new focus on IAPV and the distribution of diagnostic reagents, we will learn that it is even more widely distributed. Nonetheless, IAPV lineages have now been found in U.S. bees; one of them correlates genetically with IAPV found in bees in Australian shipments. The presence of IAPV strains in older U.S. samples does not eliminate a role for this virus in CCD.

DIANA COX-FOSTER,[1] SEAN CONLAN,[2] EDWARD C. HOLMES,[3,4] GUSTAVO PALACIOS,[2] ABBY KALKSTEIN,[1] JAY D. EVANS,[5] NANCY A. MORAN,[6,8] PHENIX-LAN QUAN,[2] DAVID GEISER,[7] THOMAS BRIESE,[2] MADY HORNIG,[2] JEFFREY HUI,[2] DENNIS VANENGELSDORP,[1,9] JEFFERY S. PETTIS,[5] W. IAN LIPKIN[2]

[1]Department of Entomology, The Pennsylvania State University, University Park, PA 16802, USA. [2]Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, NY 10032, USA. [3]Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA. [4]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA. [5]Bee Research Laboratory, USDA-ARS, Beltsville, MD 20705, USA. [6]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA. [7]Department of Plant Pathology, The Pennsylvania State University, University Park, PA 16802, USA. [8]The Center for Insect Science, University of Arizona, Tucson, AZ 85721, USA. [9]The Pennsylvania Department of Agriculture, Bureau of Plant Industry–Apiculture, Harrisburg, PA 17110, USA.

References

1. A. C. Brault *et al*., *Nat. Genet.* **39**, 1162 (2007).
2. E. Maori *et al*., *J. Gen. Virol.* **88**, 3428 (2007).
3. M. Shen, X. Yang, D. Cox-Foster, L. Cui, *Virology* **342**, 141 (2005).
4. D. Anderson, Rural Industries Research and Development Council Publication #04/152 (2004).
5. Y. Chen, J. D. Evans, *Am. Bee J.* **147**, 1027 (2007).

More Toxin Tests Needed

IN HIS EDITORIAL "TOXIC DILEMMAS" (23 November 2007, p. 1217), D. Kennedy mentions tris(2,3-dibromopropyl) phosphate. In 1978, results published by the National Cancer Institute clearly showed that this flame retardant was carcinogenic in both sexes of rats and mice, causing cancers of the kidney, lung, liver, and forestomach (*1*).

Kennedy was an integral partner in the formation in 1978 of the National Toxicology Program (NTP), and as FDA Commissioner, he was an early chairman of the NTP Executive Committee (*2*). Since its inception, NTP has conducted nearly 600 chemical carcinogenesis bioassay studies,

nearly half of which have shown evidence of carcinogenic activity (*3*). However, this represents only about 0.6% of available chemicals on the market. Likewise, the International Agency for Research on Cancer has evaluated only about 950 chemicals for carcinogenic activity; of these, about 100 were found to be human carcinogens, another 69 were classified as probably carcinogenic to humans, and 246 were classified as possibly carcinogenic to humans (*4*). The number of chemicals that have not yet been tested is staggering, and it becomes even more formidable when one considers mixtures of chemicals, together with the thousands of new chemicals that enter the marketplace every year.

We live in a chemical soup, and alternative methods of testing chemicals, such as in vitro short-term testing, have failed at identifying carcinogens. The NTP, the major testing program in the world, starts at most only five new bioassays per year. We must test more chemicals for carcinogenicity than are currently being evaluated.

JAMES HUFF

National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA.

References and Notes

1. National Toxicology Program, *TR-76 Bioassay of Tris (2,3-Dibromopropyl) Phosphate for Possible Carcinogenicity (CAS No. 126-72-7)*; available online at http://ntp.niehs.nih.gov/go/6907.
2. The NTP Executive Committee was made up of governmental research and regulatory agencies, including the Centers for Disease Control and Prevention, Consumer Product Safety Commission, Environmental Protection Agency, Food and Drug Administration, National Cancer Institute, National Center for Toxicological Research, National Institute of Environmental Health Sciences, National Institutes of Health, National Institute for Occupational Safety and Health, and Occupational Safety and Health Administration.
3. National Toxicology Program, Department of Health and Human Services, Long-Term Study Reports and Abstracts; available online at http://ntp-server.niehs.nih.gov/ntpweb/index.cfm?objectid=D16D6C59-F1F6-975E-7D23D1519B8CD7A5.
4. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans; available online at http://monographs.iarc.fr/ENG/Classification/crthall.php.

The Inimitable Field of Cosmology

IN THE NEWS FOCUS ARTICLE "A SINGULAR conundrum: How odd is our universe?" (28 September 2007, p. 1848), A. Cho perpetuated misunderstanding of science with the statement, in part from James Gunn, that "'Cosmology may look like a science, but it isn't a science' because it's impossible to do repeatable experiments." In the truly natural sciences (such as geology, oceanography, atmospheric science, and ecology), rigorous observation and interpretation are commonly used, rather than "repeatable experiments" à la Karl Popper—except in those few cases where a small-scale experiment is meaningful (*1*). It would be better to say that cosmology is science—it just isn't Popperian physics.

L. BRUCE RAILSBACK

Department of Geology, University of Georgia, Athens, GA 30602–2501, USA.

Reference

1. C. M. Condit, L. B. Railsback, The Transilience Project (www.gly.uga.edu/railsback/Transilience/Transilience.html).

Response

WITH ALL DUE RESPECT, I THINK THAT L. B. Railsback misses the point that I was attempting to make. The pursuit of understanding the cosmos is certainly a scientific pursuit and makes use of many of the most powerful tools of science. Unfortunately, there is only one observable universe, and while it is quite possible in principle, and probably in practice, to formulate theories that describe its observed behavior perfectly on the largest scales, those theories could well be unverifiable by any doable experiment.

In geophysics and astrophysics, the experimenter is nature, not the scientist, but repeated experiments can be done and the results can be observed. This is not so in cosmology for phenomena on the largest scales. Further confusion stems from our belief that the structure we are observing is stochastic on scales up to and beyond the current particle horizon. As discussed in the News Focus article, we may be unlucky enough to live in a volume in which some large-scale quantity assumes a very unlikely value within the framework of some otherwise seemingly successful theory. It then becomes a very subjective matter of whether this observation does or does not rule out the theory in question.

Whatever one's view on the Popperian definition, verification by whatever technique is a cornerstone of science; I am merely saying that this can be impossible for crucial and interesting aspects of cosmological inquiry.

JAMES GUNN

Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA.

CORRECTIONS AND CLARIFICATIONS

Editors' Choice: "Cooler in the forest" (7 December 2007, p. 1525). The final sentence should have been "Thus, contrary to some assertions, conversion of open fields to wooded fields will not necessarily lead to local increases in temperature."

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "Clustering by Passing Messages Between Data Points"

Michael J. Brusco and Hans-Friedrich Köhn

Frey and Dueck (Reports, 16 February 2007, p. 972) described an algorithm termed "affinity propagation" (AP) as a promising alternative to traditional data clustering procedures. We demonstrate that a well-established heuristic for the *p*-median problem often obtains clustering solutions with lower error than AP and produces these solutions in comparable computation time.

Full text at www.sciencemag.org/cgi/content/full/319/5864/726c

RESPONSE TO COMMENT ON "Clustering by Passing Messages Between Data Points"

Brendan J. Frey and Delbert Dueck

Affinity propagation (AP) can be viewed as a generalization of the vertex substitution heuristic (VSH), whereby probabilistic exemplar substitutions are performed concurrently. Although results on small data sets (≤900 points) demonstrate that VSH is competitive with AP, we found VSH to be prohibitively slow for moderate-to-large problems, whereas AP was much faster and could achieve lower error.

Full text at www.sciencemag.org/cgi/content/full/319/5864/726d

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.

Comment on "Clustering by Passing Messages Between Data Points"

Michael J. Brusco[1]* and Hans-Friedrich Köhn[2]

Frey and Dueck (Reports, 16 February 2007, p. 972) described an algorithm termed "affinity propagation" (AP) as a promising alternative to traditional data clustering procedures. We demonstrate that a well-established heuristic for the *p*-median problem often obtains clustering solutions with lower error than AP and produces these solutions in comparable computation time.

Frey and Dueck (*1*) described an algorithm for analyzing complex data sets termed "affinity propagation" (AP). The algorithm extracts a subset of representative objects or "exemplars" from the complete object set by exchanging real-valued messages between data points. Clusters are formed by assigning each data point to its most similar exemplar. The authors reported that "[a]ffinity propagation found clusters with much lower error than other methods, and it did so in less than one-hundredth the amount of time" (*1*). We demonstrate that an efficient implementation of a 40-year-old heuristic for the well-known *p*-median model (PMM) often provides lower-error solutions than AP in comparable central processing unit (CPU) time.

For consistency with AP in (*1*), we present the PMM as a sum of similarities maximization problem, while recognizing that this is equivalent to the more common form of minimizing the sum of dissimilarities (e.g., distances or costs). The PMM is a general mathematical problem that can be concisely stated as follows: Given an $m \times n$ similarity matrix, S, select p columns from S such that the sum of the maximum values within each row of the selected columns is maximized (*2*). Thus, each row is effectively assigned to its most similar selected column (exemplar) with the goal of maximizing overall similarity. One classic example of the PMM occurs in facility location planning: Locate p plants such that the total distance (or cost) required to serve m demand points is minimized. In data analysis applications where S is an $n \times n$ matrix of negative squared Euclidean distances between objects, clustering the n objects using the PMM corresponds to the selection of p exemplars to minimize error, which is defined as the sum of the squared Euclidean distances of each object to its nearest exemplar.

Lagrangian relaxation methods enable the exact solution of PMM instances with $n \leq 500$ objects (*3*, *4*). For larger problems, a vertex substitution heuristic (VSH) developed in (*5*) has been the standard for comparison for nearly four decades. The VSH begins with the random selection of a subset of p exemplars, which is iteratively refined by evaluating the effects of substituting an unselected point for one of the selected exemplars. Frey and Dueck assert that this type of strategy "works well only when the number of clusters is small and chances are good that at least one random initialization is close to a good solution" (*1*). To the contrary, the VSH is remarkably effective and is often the engine for metaheuristics such as tabu search (*6*) and variable neighborhood search (*7*).

We compared AP to an efficient implementation of VSH (*7*) across eight data sets from the clustering literature: I: Hartigan's (*8*) birth/death rates data; II: Fisher's (*9*) iris data; III: Lin and Kernighan's (*10*) circuit-board data; IV: Reinelt's (*11*) circuit-board data; V, VI, and VII: Grötschel and Holland's (*12*) European city coordinates; and VIII: Olivetti face images as used in (*1*). The measure of similarity for all data sets was negative squared Euclidean distance. The AP preference vectors for I through VII were established based on median similarity, as recommended in (*1*). For data set VIII, we used the preference vector from (*1*).

Affinity propagation was applied to each test problem (*13*), and the selected number of exemplars (p), the observed error, and computation time were stored. Next, 20 restarts of the VSH were applied assuming the value of p selected by AP. Each restart used a different randomly selected set of p exemplars as the initial solution. The minimum error observed across the 20 restarts, as well as the computation time, were stored for each test problem. Finally, we obtained optimal solutions for problems I through VII, as well as a reasonably tight lower bound for problem VIII, using Lagrangian relaxation methods.

The results in Table 1 are discordant with the claims in (*1*) that AP is an appreciably more efficient algorithm that produces solutions with less error than standard iterative refinement heuristics. Affinity propagation and the VSH yielded the same error for problem I; however, the VSH obtained a better solution than AP for the other seven problems. Moreover, the AP error value exceeded the optimum by 3% or more for four of the eight problems, whereas the VSH error never exceeded the optimum by more than 0.27%. In terms of efficiency, the maximum ratio of VSH to AP computation time was 2.79:1, which is drastically less than the 100:1 ratio reported in (*1*).

We also applied VSH to the document summary and airline travel routing data sets from (*1*). These are asymmetric similarity matrices with violations of the triangle inequality and were included to refute the implication in (*1*) that the PMM is inapplicable for such conditions. The VSH accommodates the asymmetry and captures differential preferences as a "fixed charge" in the objective function. For the document summary data, AP produces a four-cluster solution with a net similarity index of –10,234.33. The four-cluster VSH solution yielded a better index of –10,216.63. For the travel routing data, the similarity index of –92,154 for the seven-cluster VSH solution was better than the corresponding figure of –92,460 for AP. The VSH was slightly faster than AP for both test problems.

In summary, using a classic heuristic for the PMM, we frequently obtained better solutions than AP in comparable computation time. Moreover, like AP, the PMM has sufficient flexibility to accommodate nonmetric forms of S. With respect to another purported advantage of AP, the selection of the number of exemplars, AP merely replaces the problem of choosing p with the problem of setting the preference vector. Thus, the problem of choosing p is not resolved by AP. In light of these issues, we contend that AP,

Table 1. A computational comparison of AP and VSH.

Test problem (label)	*n* (points)	*p* (exemplars)	*f** = Optimal error (squared distance)	% > *f** AP (min %)	% > *f** VSH (min %)	Total CPU time VSH:AP (ratio)
I	70	6	863.41	0.00	0.00	0.46:1
II	150	6	43.98	4.50	0.00	0.59:1
III	318	11	44,882,137	0.22	0.00	1.20:1
IV	724	22	21,396,140.7195	1.88	0.25	2.70:1
V	202	17	2,065.5737	4.10	0.00	1.31:1
VI	431	16	45,316.3623	5.27	0.10	1.22:1
VII	666	17	127,447.545	3.53	0.21	1.20:1
VIII	900	62	9,665.84*	0.71	0.27	2.79:1

*The optimal solution for problem VIII could not be verified. The reported value is a lower bound.

[1]College of Business, 307 Rovetta Business Building, Florida State University, Tallahassee, FL 32306–1110, USA. [2]Department of Psychological Sciences, 19 McAlester Hall, University of Missouri-Columbia, Columbia, MO 65211, USA.

*To whom correspondence should be addressed. E-mail: mbrusco@cob.fsu.edu

although a welcome addition to the clustering literature, does not possess the relative advantages over existing procedures that were touted in (*1*). We invite analysts to compare AP to VSH on other data sets and render their own opinions (*14*).

References and Notes

1. B. J. Frey, D. Dueck, *Science* **315**, 972 (2007).
2. N. Mladenović, J. Brimberg, P. Hansen, J. A. Moreno-Pérez, *Eur. J. Operational Res.* **179**, 927 (2007).
3. G. Cornuejols, M. L. Fisher, G. L. Nemhauser, *Manage. Sci.* **23**, 789 (1977).
4. P. Hanjoul, D. Peeters, *Eur. J. Operational Res.* **20**, 387 (1985).
5. M. B. Teitz, P. Bart, *Oper. Res.* **16**, 955 (1968).
6. E. Rolland, D. A. Schilling, J. R. Current, *Eur. J. Operational Res.* **96**, 329 (1997).
7. P. Hansen, N. Mladenović, *Location Sci.* **5**, 207 (1997).
8. J. A. Hartigan, *Clustering Algorithms* (Wiley, New York, 1975).
9. R. A. Fisher, *Ann. Eugen.* **7**, 179 (1936).
10. S. Lin, B. W. Kernighan, *Oper. Res.* **21**, 498 (1973).
11. G. Reinelt, www.iwr.uni-heidelberg.de/groups/comopt/software.TSPLIB95/index.html (2001).
12. M. Grötschel, O. Holland, *Math. Prog.* **51**, 141 (1991).
13. Analysis was performed using Matlab function apcluster.m, downloaded 12 December 2007 from www.psi.toronto.edu/affinitypropagation.
14. All algorithms and data sets used to obtain results for this Technical Comment are available at http://garnet.acns.fsu.edu/~mbrusco.

25 September 2007; accepted 14 January 2008
10.1126/science.1150938

SOCIOLOGY

Facing the Flames

Eugene A. Rosa

If cutting and digging a fireline in the midst of a crown fire blowup that bellows the sound of wood exploding under its intense heat is not a real risk, few activities would qualify. Lines are constructed with a Pulaski (a combination ax and adze or grub hoe), the principal tool of wildland firefighters on the front line in forest fire suppression. A key risk is to become trapped in a burnover.

Despite low rates of fatality and injury, wildland fire fighting falls in the wide range of occupations where performing one's duties is clearly risky business. Furthermore, firefighting risks are increasing due to the accelerated frequency, intensity, and erratic behavior of wildland fires—changes traceable to global warming, past management practices that allowed fuel accumulation, and residential encroachment of the forests (the "wildland-urban interface").

What do we know about such risky occupations and the people who choose them? Very little. Overwhelmingly, the social science of risk is framed by the reductionistic structure of psychology and economics, relies principally on surveys and other pencil-and-paper data with predesigned expectations, and rarely connects the thoughts of respondents with actual behavior.

Matthew Desmond's *On the Fireline* holds considerable promise for addressing this pivotal gap in the social science literature on risk. The book, which began as a master's thesis at the University of Wisconsin, draws from his five summers of experience on a 14-member engine crew in the Arizona wilds. Desmond's in-depth account provides a zoom lens into how U.S. Forest Service firefighters perceive and act on risks under working conditions. Why do they risk, and what "practical logic" do they use to "understand risk and death"?

Firefighting, like the favored Pulaski, is a hybrid that welds "desired risk" (*1*) with the instrumental needs of forest management. To understand it, the author takes us into the encampments of the seasonal firefighters—motivated by adventure (but not a lust for danger) and financial incentives—who can answer these questions. There we learn that firefighting combines manual skills with love of the outdoors and self-reliance, cultivates distaste for urban sophistication, rewards the ease of firefighters to be filthy, enables the gritty idiom of that macho culture (the f-word punctuates most sentences), and provides not-infrequent occasions to party.

On the Fireline
Living and Dying with Wildland Firefighters

by Matthew Desmond

University of Chicago Press, Chicago, 2007. 379 pp. $24, £15. ISBN 9780226144085.

When fighting a fire (an "incident" in the parlance of the fire agencies), firefighters are immersed in the deep preoccupation of doing their job—risk is clearly in sight but out of mind. But, what are their thoughts about risk during the frequent down time? The "grunts," Desmond's principal subjects, almost never think of their work in terms of risk or give much thought to their safety or to dying—although supervisors do. The Forest Service ensures this in its training and retraining by inculcating the ideas that firefighters are responsible for their own safety and competency and that by following the Ten Standard Fire Orders and Eighteen Watch Out Situations (known as the ten and eighteen) they have little to worry about.

Seeking safety. A crew deploys their fire shelters after being trapped by the Santiago fire (October 2007, Orange County, California).

A firefighter's death is the clearest safety failure, and each prompts a formal investigation. Although such investigations typically identify multiple causes, the finger of attribution nearly always points at the dead firefighter. This practice is consistent with the literature on organizations, where up to 80% of system failures are assigned to individuals. But Desmond deepens our understanding of this pattern by showing that the ten and eighteen, unrealizable in practice, are not just sets of rules to be followed but also an affirmation of the Forest Service's code of individual responsibility that prefigures attribution of blame: "*Trust only one person: yourself. You are responsible for your own safety and actions on the fireline.*" Tracing death to the missteps of the individual firefighter deflects attribution away from the firefighting system and its managers and preempts the active learning at the core of adaptive management.

Because of its captivating prose (despite an excess of tropes and other obeisances to postmodern discourse) and its masterful crafting of a French tradition in sociology (that emphasizes social order via symbols and rituals) to understand risk, the book may become an instant hit in mainstream sociology. It may even be fast-tracked to "classic" status alongside historian Stephen J. Pyne's personal account of firefighting (*2*).

However, owing to Desmond's slavish commitment to a narrow sociological framing, O*n the Fireline* fails to engage the sizable social science literature on risk as well as the work of psychologists, geographers, sociologists, political scientists, and economists who study wildland fire. Justification for this neglect stems from Desmond's characterization of that literature with an ever-thinning man of straw, the rational actor. His treatment overlooks four decades of research that shows the axioms of rationality to be woefully inadequate to describe risk choices despite serving as a useful normative decision aid (*3*). This strategic inattention handicaps the book's success in the science and risk communities, positioning it as an interesting, but tangential, contribution to the field.

References

1. G. E. Machlis, E. A. Rosa, *Risk Anal.* **10,** 161 (1990).
2. S. J. Pyne, *Fire on the Rim: A Firefighter's Season at the Grand Canyon* (Weidenfeld and Nicolson, New York, 1989).
3. C. C. Jaeger, O. Renn, E. A. Rosa, T. Webler, *Risk, Uncertainty, and Rational Action* (Earthscan, London, 2001).

The reviewer is at the Department of Sociology and Thomas S. Foley Institute for Public Policy and Public Service, Washington State University, Pullman, WA 99164–4020, USA. E-mail: rosa@wsu.edu

10.1126/science.1153554

CREDIT: KAREN TAPIA-ANDERSON/LOS ANGELES TIMES

TECHNOLOGY

Change of Time over Changing Time

Thomas S. Mullaney

BROWSING

Transit Maps of the World. Mark Ovenden. Penguin, New York, 2007. 144 pp. Paper, $25, C$27.50. ISBN 9780143112655.
Diagrams of rail transit systems have become icons of urban life. Ovenden's comprehensive survey includes over 200 cities, with series of historical maps charting the evolution of the more complex systems. The text covers the growth of each metro, but it is focused on the graphic designs used to help travelers get from one station to another.

Everyone knows that people, places, and things change over time. In his most recent and final work, the late scientist-turned-historian Ian R. Bartky reminds us that time itself, often taken as the common denominator of change, is also in flux. Concatenating these two observations, we are left with a recursive variation on an old theme: the change of time over changing time.

One Time Fits All: The Campaigns for Global Uniformity deals with five campaigns for global temporal standardization: the selection of an International Date Line, the establishment of the "universal day," the creation of the common initial meridian, the subdivision of the globe into time zones, and the institution of Daylight Saving Time. Each of these campaigns is tied to separate yet overlapping histories, a multiplicity of actors, and a series of complex debates, all of which Bartky's highly accessible account disentangles.

One Time Fits All
The Campaigns for Global Uniformity
by Ian R. Bartky
Stanford University Press, Stanford, CA, 2007. 320 pp. $49.95. ISBN 9780804756426.

After a career as a physical chemist at the U.S. Bureau of Standards, Department of Commerce, and Army laboratories, Bartky (who died in December) became an expert on the history of timekeeping. Like its predecessor (*1*), this book is based on extensive archival and library work, to which he brought a meticulous attention to detail. And he again digests an abundance of very technical material and presents it in a thoroughly comprehensible format.

In the book's first section, which begins with Magellan's 16th-century circumnavigation of the globe and closes in 1921, Bartky discusses the emergence and resolution of a question that had scarcely been relevant in the era before regular transoceanic voyages: the problem of temporal reckoning that surfaces when (in modern parlance) one crosses the International Date Line. The second section deals with a narrower span of time (1870 to 1925) but a much broader—sometimes even dizzying—array of scientists, politicians, commercialists, and industrialists. Its eight chapters address the selection of a common initial meridian, the standardization of the uniform day, and the division of the globe into 24 time zones. Each chapter unfolds like a short play, with half of the dramatis personae supporting temporal standardization and the other half standing in opposition. The third section and epilogue bring the question of temporal standardization from 1883 to the present, looking at the issue of daylight saving from the moment it was proposed through the changes made in the United States in 2007.

It is also important to note what readers will not find in the book. They will not learn much about the social history of standardized time, either in terms of its popularization or the influence that clock time had on the way people experienced their daily lives. This is a book primarily concerned with policymakers. Readers should not expect to dwell long on any one aspect of the history of temporal standardization. At moments, the book reads like a précis of a much longer work. In addition, the book does not include critical analyses of the concepts of standardization or modern rationality per se. Bartky clearly identifies with the standardizers and, like many historians of standardization, writes with a presentist orientation. Lastly, readers will not learn about the world beyond Europe and North America.

As a historian of China, I understand that this absence of the non-Western world is partly justified by a parallel absence in the sources. China, for example, was not represented at any of the major events associated with the International Prime Meridian, Greenwich Mean Time, or the network of metrological organizations that competed, and collaborated, to develop and enforce worldwide standards. On the other hand, the continued preoccupation of scholars with European and North American standardization prevents them from asking important questions. For example, why doesn't the United States, Canada, or Russia observe one time zone? If this sounds like an absurd question, remember that the Chinese government in 1949 opted to do just that. Spanning some 62 degrees of longitude (from eastern Heilongjiang province to western Xinjiang), a country that "should" have five time zones recognized only one, the single standard known colloquially as "Beijing Time" (or technically as UTC + 8). To Bartky I would have liked to have asked: Is Beijing Time part of global time, or is it a symbolic secession from a standard that China had no part in creating?

The reviewer is at the Department of History, Stanford University, Stanford, CA 94305–2024, USA. E-mail: tsmullaney@stanford.edu

Reference

1. I. R. Bartky, *Selling the True Time: Nineteenth-Century Timekeeping in America* (Stanford Univ. Press, Stanford, CA, 2000).

10.1126/science.1152851

CREDIT (TOP): COURTESY PENGUIN BOOKS

Published by AAAS

SUSTAINABLE DEVELOPMENT

Climate Change—the Chinese Challenge

Controlling CO_2 emissions without hindering economic development is a major challenge for China and the world.

Ning Zeng,[1]* Yihui Ding,[2] Jiahua Pan,[3] Huijun Wang,[4] Jay Gregg[5]

In 2006, China's carbon dioxide emission rate reached 1.6 GtC (gigatons of carbon or 10^{15} g carbon) per year (see chart, below) (*1*–*3*). Economic growth is projected to continue at higher than 7% per year; at this rate, Gross Domestic Product (GDP) would quadruple in 20 years. The associated high CO_2 emission rate would substantially affect the goal of avoiding dangerous climate change as set by the United Nations Framework Convention on Climate Change (UNFCCC). The conflict between economic development and keeping atmospheric greenhouse gases at a manageable level poses one of the greatest challenges of this century.

The Impact of Climate Change on China

China will be one of the worst-impacted regions in the world if climate changes as predicted (*4*). Three main industrial centers of China are on lowland areas: the Gulf of Bohai region with the Beijing-Tianjin axis, the Yangtze River delta radiating inland from Shanghai, and the Pearl River delta encompassing Hong Kong and Guangzhou. A sea-level rise of a meter would inundate 92,000 km^2 of land in these three regions (*5*).

Mountain glaciers have melted by 21% over the past 50 years in northwestern China (*5*). Under the projected climate change scenarios, temperature in the Tibet region would rise 3° to 6°C by 2100 (*4*). The melting of the permafrost might threaten the newly completed Qinghai-Tibet Railway (*6*). Aside from the Yangtze and the Yellow rivers, several major Asian rivers, such as Ganges and Mekong, originate from the Tibetan plateau, and the changing water resources may lead to tension with the neighboring countries.

The geography and climatology of China already give rise to frequent extreme events. The summer storms moving eastward along the river systems, which are generally oriented west to east, dump large amounts of rainfall that cause severe flooding. Indeed, the Three Gorges Dam was motivated largely on the premise of flood control. In the summer of 2006, Chongqing and Sichuan in the upper Yangtze Basin experienced a once-every-100-years drought, followed by a similarly rare flood in 2007, a harbinger of the projected intensification of extreme events in southern and eastern China (*5*).

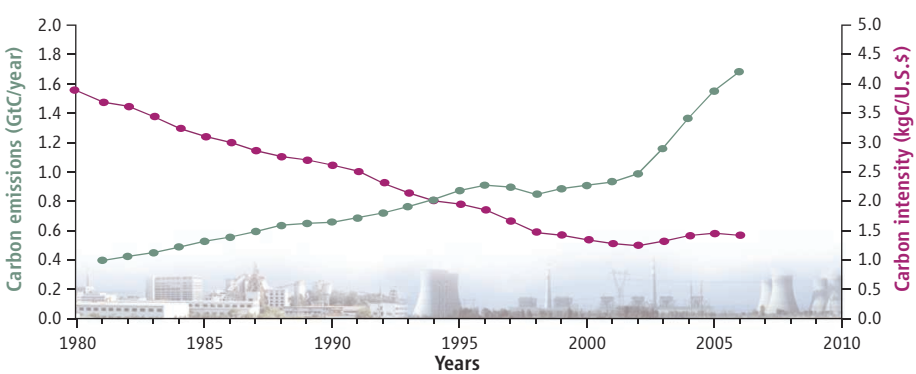
Half of the country's land area is arid or semiarid. Water shortage in northern China over the last three decades led to the ongoing construction of the South-North Water Diversion Project, a gigantic project that will divert water from three points of the Yangtze River basin to the north. Global warming is likely to enhance such drying. China's agricultural output could be reduced by 5 to 10% by 2030 (*5*), thus adding stress to a country that has 20% of the world's population and only 7% of the arable land. Similarly, major ecosystem impacts are expected with the loss of tundra and mountain forests and the intensification of fires (*5*).

Drivers of CO_2 Emissions and Coal Use

China's GDP has grown by 9.5% per year over the last 27 years, while its CO_2 emissions have increased by only 5.4% per year (*1*–*3*), corresponding to a large decrease in carbon intensity (carbon emission per unit of GDP) (see chart, below)—a remarkable achievement, as energy consumption generally grows faster than GDP during the early stages of industrialization. One important reason for this was the government's emphasis on energy efficiency. China's per capita emission is still low (one-fourth of U.S. CO_2 emissions), but the large population and high speed of economic development have led to a large increase in energy demand and have been primary drivers of the recent acceleration in global carbon emissions (*7*). The rebound in carbon intensity in the last few years (see chart, below) was caused mainly by accelerated urbanization and industrialization (see photo, page 731). For instance, stimulated by a construction boom, steel production has increased from 140 to 419 million tons from 2000 to 2006, now accounting for 34% of world total. In 2006, 7.2 million cars were sold, compared with 1.2 million in 1999. However, about 23% of the CO_2 emissions are a result of producing goods exported to other countries (*8*).

The large population dictates that China cannot duplicate the energy-intensive Western model because of resource limitations. In 2006, China imported 47% of its crude oil and is projected to import 60% by 2020. Given escalating oil prices and concerns about energy security, China has no alternative but to focus on domestic resources.

China has one of the largest coal reserves in the world, and coal accounts for 67% of its primary energy use, compared with 24% for the world average (*9*). China is currently bringing two additional coal-fired power plants to the electric power grid every week (*10*). Although the government had set the goal of 20% reduction in energy intensity (energy consumption per unit GDP) in the 11th Five-Year Plan period (2006–10) (*11*), the goal looks unlikely



CO_2 emissions and carbon intensity for China from 1980 to 2006 (*1*–*3*).

PHOTO CREDIT: TANGCUN TOWN, SHANDONG, CHINA; WWW.ZCTC.GOV.CN

[1]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742, USA. [2]National Climate Center, Chinese Meteorological Administration, Beijing, China. [3]Research Center for Sustainable Development, Chinese Academy of Social Sciences, Beijing, China. [4]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China. [5]Department of Geography, University of Maryland, College Park, MD 20742, USA.

*Author for correspondence. E-mail: zeng@atmos.umd.edu

Published by AAAS

to be met, as energy intensity has decreased by only 1.23% in 2006. In a hypothetical scenario in which carbon intensity keeps pace with a GDP growth rate of 7%, by 2030, China would be emitting as much as the world as a whole is today (8 GtC/year). If China could limit carbon intensity at half of the economic growth rate, as was done in the 1980s–90s (*1*–*3*), a quadrupling of GDP in 20 years would still lead to CO_2 emissions of 3 to 4 GtC/year.

Rapid development leads to high energy demand and CO_2 emissions.

PHOTO CREDIT: JAY GREGG

Opportunities

The most effective near-term strategy is energy conservation and efficiency. Because infrastructure has a long lifetime, it is much more cost-effective to design and build from the ground up rather than retrofit afterwards. Many energy-saving measures can be highly effective, and long-term energy savings can substantially outweigh the initial investment. However, such measures are often not implemented because of market and institutional barriers. For instance, when offered the choice of paying a few percent extra for energy-efficient construction, the owners of a new building often choose not to, because of the burden on their budget and uncertainty about future savings.

International investment of carbon funds should be aggressively directed to the infrastructure buildup in China to prevent a legacy effect of inefficient technologies. For instance, funds can be used to subsidize low-interest loans to energy-efficient buildings or to pay for technology transfer. Research and development in technology, such as renewable energy, energy efficiency, and carbon sequestration, in the developed countries could be conducted in collaboration with Chinese partners, so that these technologies could be implemented as early as possible.

Rural development must avoid the pollution-heavy and energy-intensive route, instead, incorporating state-of-the-art technology suitable for the local circumstances. City planners should develop more efficient public transportation, reemphasize the role of bicycles, and use incentives and regulations to encourage electric bikes, buses, and cars; to reduce burgeoning traffic and air pollution problems; and to save energy.

The rate of development of renewable energy in China is even faster than that of coal-fired power plants. The 2020 targets for hydroelectric, nuclear, biomass, wind, and solar power are 300, 40, 30, 30, and 1.8 GW, respectively (*12*). Even though China is a top manufacturer of solar heaters, solar panels and wind turbines, core technologies are often not available to them. As a result, the high cost hinders rapid deployment of the most efficient technologies.

Avoided emissions and active carbon sinks deserve credit on the international carbon market. Over the past three decades, the Chinese taxpayers have supported several massive ongoing reforestation projects, such as the northern China project to prevent desertification, with a total area of 60 million ha of new forests. As a result, the country's forest cover increased from 12% in 1980 to 18.2% in 2005 and is projected to increase to 23% by 2020 and 26% by 2050 (*11*). In contrast, only a handful of reforestation projects in China of roughly 10,000 ha have been funded through the Clean Development Mechanism (CDM) under the UNFCCC's Kyoto Protocol, because of a 1% limit on reforestation's share in meeting the Kyoto target of each developed country, the complicated accounting and verification procedures, and the current low carbon market price. About half of CDM investment projects are in China, but they have not been effective at cutting CO_2 emissions (*13*). The Bali climate conference in December 2007 started a new round of negotiations, and substantial progress will have to be made in order to effectively help developing countries like China reduce CO_2 emissions.

China has actively participated in UNFCCC and the Kyoto Protocol and played an important role in the Intergovernmental Panel on Climate Change (IPCC). China and other developing countries are not subject to emission targets under the Kyoto Protocol. For China, this makes coal attractive as an energy source because it is domestically abundant and cheap. To prevent continued reliance on inefficient coal power, developers need a clear market signal that a climate policy, be it international or domestic, is a certainty in the future policy landscape. Under such a policy, inefficient coal plants will become a liability. The Chinese government could set internal emission targets and devise strategies to meet them. For example, a carbon tax in China could be used to fund research in energy efficiency, renewable energy, carbon sequestration, and prudent urban design. Such voluntary efforts would protect China's energy security, create careers for the masses of young and educated Chinese citizens, and also earn China political capital in international climate policy negotiations.

Despite the recent surge of worldwide attention in the climate change problem, its enormous scale and urgency are often underappreciated. The Chinese challenge is arguably the most difficult, and coal is the leading stumbling block. If China can face the challenge and seize the opportunities with the help of the international community, it could lead the world in sustainable development in the 21st century.

References and Notes

1. GDP from International Monetary Fund (www.imf.org/external/data.htm) converted to constant current U.S. dollars.
2. Carbon emission from Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge, TN (http://cdiac.ornl.gov/trends/emis/meth_reg.htm), with updates from (*3*).
3. Updates for 2005–06 from the Netherlands Environmental Assessment Agency (MNP) (www.mnp.nl/en/service/pressreleases/2007/index.html).
4. IPCC, *Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, Cambridge, 2007).
5. *China's National Assessment Report on Climate Change* (China Science Press, Beijing, 2007) (in Chinese).
6. C. H. Peng *et al.*, *Science* **316**, 546 (2007).
7. M. R. Raupach *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10288 (2007).
8. T. Wang, J. Watson, Tyndall Centre Briefing Note no. 23, October 2007; http://tyndall.webapp1.uea.ac.uk/publications/briefing_notes/bn23.pdf.
9. International Energy Agency (IEA), *World Energy Outlook* (IEA, Paris, 2007).
10. J. Katzer *et al.*, *The Future of Coal* (Massachusetts Institute of Technology, Cambridge, MA, 2007); http://web.mit.edu/coal/.
11. National Development and Reform Commission (NDRC), *China's National Climate Change Programme* (NDRC, Beijing, 2007); http://en.ndrc.gov.cn/newsrelease/P020070604561191006823.pdf.
12. NDRC, *China's Medium-to-Long-Term Renewable Energy Plan* (NDRC, Beijing, 2007) (in Chinese).
13. M. Wara, *Nature* **445**, 595 (2007).

14. We thank all who have contributed to the *National Assessment Report* (*5*), in particular Y. Luo.

10.1126/science.1153368

GENETICS

Dwarfism, Where Pericentrin Gains Stature

Mutations in a protein that functions in cell division result in human growth disorders, possibly by connecting DNA damage signaling and centrosome dysfunction.

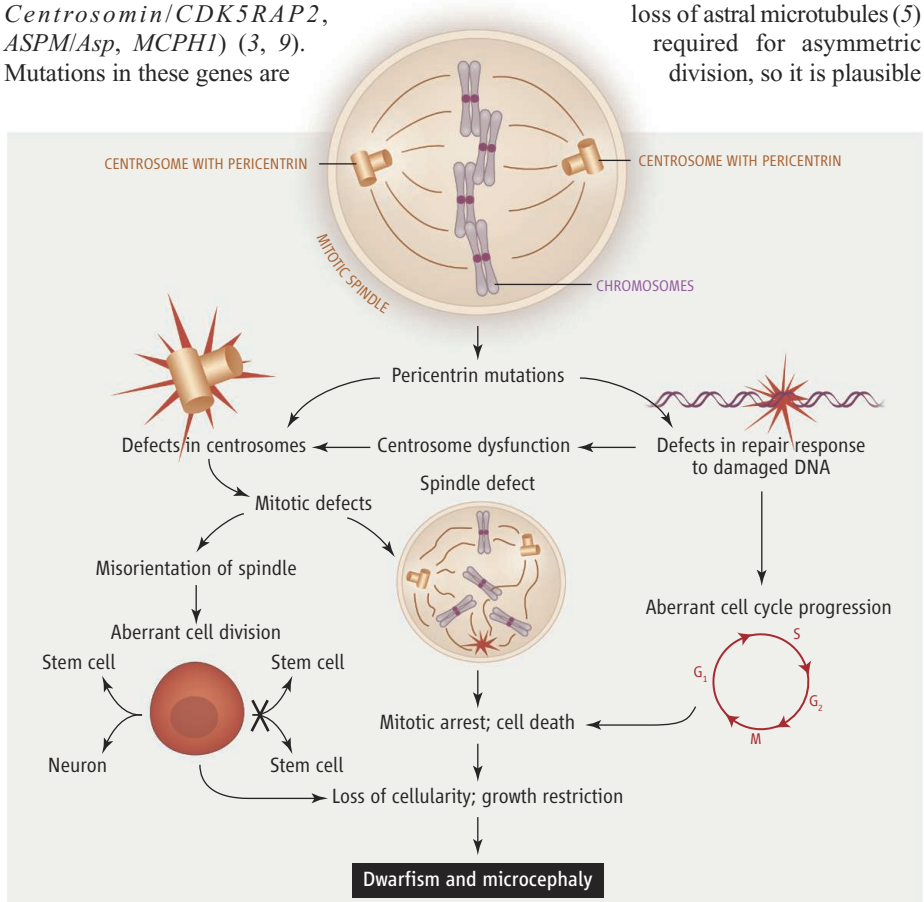
Benedicte Delaval and Stephen Doxsey

It is thought that the smallest human ever recorded was a sexually mature girl, 12 years of age, who was 20 inches tall and weighed 5 pounds (*1*). She was afflicted with a rare genetic disorder called Majewski osteodysplastic primordial dwarfism type II (MOPD II), characterized by short stature (dwarfism) and small brain size relative to age-matched individuals (microcephaly). Two research groups now report—Rauch *et al.* on page 816 of this issue (*2*) and Griffith *et al.* (*3*)—that mutations in the gene *PCNT*, which encodes the centrosome protein pericentrin, cause MOPD II and Seckel syndrome, two disorders that share small body and brain size but exhibit distinctive features.

The precise mechanisms underpinning these autosomal recessive disorders are unknown. Earlier work on pericentrin identified a role in cell division (mitosis) and thus the potential to modulate growth of the body and brain. Pericentrin is an integral component of the centrosome, an organelle that organizes the mitotic spindle for segregation of chromosomes during cell division and appears to influence cell cycle progression (*4*). Depletion of pericentrin in human cells and in budding and fission yeast (but not flies) induces loss of centrosome and spindle integrity, followed by cell death (*5*–*8*). It is easy to imagine how pericentrin depletion could lead to loss of cellularity and growth restriction.

Is this also true in MOPD II and Seckel syndrome? In fact, both disorders share cellular abnormalities consistent with loss of pericentrin, including centrosome and mitotic spindle defects. However, Rauch *et al.* and Griffith *et al.* propose that different pathways participate in the common phenotypes characteristic of the two disorders. Rauch *et al*. suggest that mitotic centrosome dysfunction results in loss of cellularity, whereas Griffith *et al*. implicate defective progression through mitosis due to an abnormal DNA damage signaling pathway. A remaining question is whether deficits in these two cellular pathways act separately or together to cause dwarfism and microcephaly. Independently, defects in these pathways could each contribute to impaired mitotic progression and cell death (see the figure). However, the two pathways may be connected by a common element, centrosome dysfunction. DNA damage has previously been shown to induce centrosome disruption in mitosis, leading to cell death (*4*). Thus, centrosome defects arising either indirectly from DNA damage or directly from loss of centrosome proteins could induce mitotic failure, cell death, and growth restriction.

Another potential pathway disrupted in pericentrin-associated disorders is suggested by mutations in genes that encode other centrosome proteins with functions similar to pericentrin (*SAS4*/*CENPJ*, *Centrosomin*/*CDK5RAP2*, *ASPM*/*Asp*, *MCPH1*) (*3*, *9*). Mutations in these genes are associated with primary microcephaly, a disorder characterized by extreme reduction in brain size but normal stature (except for mutations in *MCPH1* for the latter). Loss of cellularity is thought to result from aberrant spindle orientation during stem cell divisions, a process that requires centrosomes (*10*, *11*). Spindle orientation determines whether cell division will generate a neuron and a stem cell or two stem cells. Defects in this process decrease symmetric divisions that replenish stem cells (see the figure), and so deplete stem cell progenitors essential for brain growth (*11*). This phenotype provides a direct link between centrosome defects and microcephaly and could help explain the mutant *PCNT* phenotype. Depletion of pericentrin in cultured cells induces loss of astral microtubules (*5*) required for asymmetric division, so it is plausible



Pericentrin function in dwarfism with microcephaly. Centrosome defects, disrupted DNA damage pathway, and impaired cell division are potential contributors to cellular loss and growth restriction that characterize human dwarfism with microcephaly.

The authors are in the Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA. E-mail: stephen.doxsey@umassmed.edu

Published by AAAS

that *PCNT* mutations associated with MOPD II and Seckel syndrome perturb asymmetric cell division throughout the body—both in the embryo and in adult stem cell niches—resulting in decreased body size.

Further insights into the etiology of dwarfism, microcephaly, and related disorders will require comparative analyses to understand how centrosome and spindle defects, altered DNA damage signaling, and aberrant spindle orientation individually contribute to the disease phenotypes. It is also important to determine whether *PCNT* mutations, which may otherwise be lethal, are hypomorphic (partially functional) or are compensated for by the expression of other genes.

References

1. J. G. Hall *et al.*, *Am. J. Med. Genet. A* **130**, 55 (2004).
2. A. Rauch *et al.*, *Science* **319**, 816 (2008).
3. E. Griffith *et al.*, *Nat. Genet.* **40**, 232 (2008).
4. S. Doxsey *et al.*, *Annu. Rev. Cell Dev. Biol.* **21**, 411 (2005).
5. W. C. Zimmerman *et al.*, *Mol. Biol. Cell* **15**, 3642 (2004).
6. M. R. Flory *et al.*, *Cell Growth Differ.* **13**, 47 (2002).
7. D. A. Stirling *et al.*, *J. Cell Sci.* **109**, 1297 (1996).
8. M. Martinez-Campos *et al.*, *J. Cell Biol.* **165**, 673 (2004).
9. M. O'Driscoll *et al.*, *Cell Cycle* **5**, 2339 (2006).
10. M. G. Giansanti *et al.*, *Development* **128**, 1137 (2001).
11. J. L. Fish *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10438 (2006).

10.1126/science.1154513

PHYSICS

Amplifying a Tiny Optical Effect

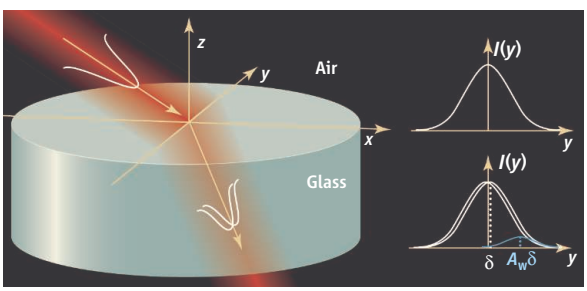
A controversial approach in quantum physics now appears capable of improving the sensitivity and precision of measurements.

K. J. Resch

How do you measure a transverse shift of a light beam to within a nanometer? Recently, physicists predicted (*1*) that light could experience such a shift similar to what happens to electrical currents in semiconductors through the spin Hall effect (*2*–*4*). On page 787 of this issue, Hosten and Kwiat (*5*) use "weak measurement," a controversial procedure from the foundations of quantum mechanics, to amplify this spin Hall effect in light (SHEL) rather than detecting it directly. In their experiment, they boost the tiny shift by a factor of 10,000, detecting it for the first time and characterizing it at the angstrom scale. This realizes one of the long-standing promises of weak measurement and demonstrates its potential in precision measurements.

In the spin Hall effect (*2*–*4*), an applied electric field induces a transversely flowing spin-polarized current. SHEL is a direct optical analog, and occurs when the electron spin and electric field are replaced by light polarization and a refractive index gradient, respectively (*1*). Light propagating in the *x*-*z* plane (see the figure, left) will experience a polarization-dependent shift in the *y* direction when it passes from one material to another with a different refractive index. For a glass-air interface and optical wavelengths, these shifts are just a few tens of nanometers. The small scale of the effect explains why it has escaped detection until now.

An unorthodox quantum measurement procedure was the key to seeing the effect. In quantum theory, any attempt to learn about an object disturbs it in an uncontrollable way. Interaction between a quantum system and a detector is essential for any measurement. Normally the interaction causes the measurement device "meter" (the part that is read to obtain the result) to undergo a shift substantially larger than its quantum uncertainty. But this is not the only way to make a measurement.



A light touch. (Left) SHEL occurs when a beam of light passes from a material with one refractive index to another as shown. If the interface is between nearly crossed linear polarizers, then the system is well described by weak measurement with a purely imaginary weak value. **(Top right)** A linearly polarized laser beam initially has a Gaussian intensity profile I_y along the *y* direction. **(Bottom right)** SHEL will shift each circular polarization component of the light by a small amount δ in opposite directions. Destructive interference in a weak measurement procedure boosts the transverse shift in the light by a large factor A_w over a directly measurable shift, as shown by Hosten and Kwiat.

Weak measurement refers to a three-step procedure invented by Aharonov *et al*. as an extension of conventional quantum measurement (*6*). The quantum system is first prepared in a well-defined initial state. Then, the measurement device is very weakly coupled to the system, such that the shift of the meter is much smaller than its quantum uncertainty. Finally, the meter position is recorded only when the quantum system is found in a specific final state, a process known as post-selection. The expected shift in the meter position is proportional to the "weak value" (a quantity dependent on the initial state, final state, and the nature of the coupling). Very large weak values can be achieved, and in these cases, the meter shift is dramatically larger than a directly measured shift (*6*). Effectively, destructive interference cancels the mean, leaving only an amplified signature of the interaction.

Weak measurement has been a controversial topic since its introduction. This is largely because the results of weak measurements can be arbitrarily large, even when standard quantum measurements are bounded; they do not even have to be real numbers. However, the issues are interpretational because weak measurement is firmly based on standard quantum mechanics. Furthermore, weak measurement was found and experimentally demonstrated to have a perfect analog in classical optics (*7*, *8*).

In contrast to previous weak measurement experiments, Hosten and Kwiat use it as a tool to measure a new phenomenon. SHEL is a natural weak coupling between circular polarization and the transverse position of an optical beam, which they use as a meter. The authors pass a linearly polarized laser beam (see the figure, top right) through a glass-air interface. If the incident angle is not perpendicular to the surface, SHEL shifts the circular polarization components of the beam by a small amount δ (see the figure, bottom right). To complete the weak measurement process, they measure only the light that passes through a second polarizer oriented almost 90° from the first. Only a small fraction of the light gets through both polarizers, because they are nearly crossed, but with a bright laser beam there is still sufficient signal to detect.

The author is at the Institute for Quantum Computing and the Department of Physics & Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. E-mail: kresch@iqc.ca

For linear polarizers, SHEL produces a purely imaginary weak value. Imaginary weak values do not shift the position, but rather the propagation angle of the light (*9*). Hosten and Kwiat developed a theory to describe the deflection of the optical beam under these conditions. Their theory predicts, and experimental results confirm, highly amplified shifts, about 10000 times δ (bottom right panel), a much larger effect than direct detection and larger amplification than even previously measured real weak values. This enhancement allowed the detection and characterization of SHEL over the full range of incident angles with angstrom precision.

SHEL is a very small effect in a standard glass-air interface, but it is predicted to be much more pronounced in photonic crystals (*1*). In such materials, SHEL may be a valuable tool for manipulating the angular momentum of photons in, for example, quantum information applications. Furthermore, by studying SHEL in clean optical systems, it may be possible to turn the tables and gain further insight into the spin Hall effect in semiconductors. In the first work on weak measurement, it was speculated that the technique could be useful in amplifying and measuring small effects (*6*). Now, 20 years later, this potential has finally been realized.

References

1. M. Onoda *et al.*, *Phys. Rev. Lett.* **93**, 083901 (2004).
2. S. Murakami *et al.*, *Science* **301**, 1348 (2003).
3. J. Sinova *et al.*, *Phys. Rev. Lett.* **92**, 126603 (2004).
4. J. Wunderlich *et al.*, *Phys. Rev. Lett.* **94**, 047204 (2005).
5. O. Hosten, P. Kwiat, *Science* **319**, 787 (2008); published online 10 January 2008 (10.1126/science.1152697).
6. Y. Aharonov *et al.*, *Phys. Rev. Lett.* **60**, 1351 (1988).
7. I. M. Duck *et al.*, *Phys. Rev. D* **40**, 2112 (1989).
8. N. W. M. Ritchie *et al.*, *Phys. Rev. Lett.* **66**, 1107 (1991).
9. A. M. Steinberg, *Phys. Rev. Lett.* **74**, 2405 (1995).

10.1126/science.1154149

IMMUNOLOGY

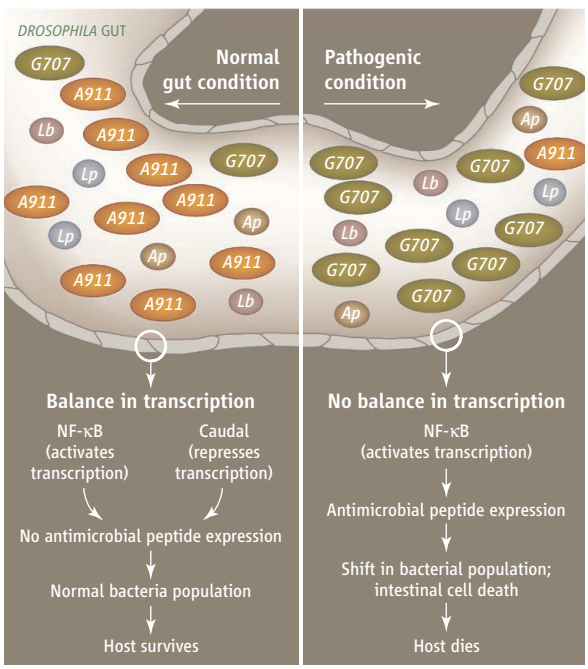
The Right Resident Bugs

A link between a transcription factor and control of immune responses in the fly gut opens the door to analyses of host-microbe mutalism.

Neal Silverman and Nicholas Paquette

The human gastrointestinal tract harbors ~500 distinct microbial taxa (*1*), comprising an estimated 10^{14} microbes. The proper maintenance of this microbial consortium is of great importance for health. Conversely, damage to the gut microbiota community is implicated in disease, such as inflammatory bowel disease (*2*). The immense complexity of gut flora and its complicated interactions with the immune system make the human gut a challenging experimental system. Recently, several groups have investigated the resident microbiota communities of insects, in particular the experimentally powerful fruit fly *Drosophila melanogaster* (*3*–*5*). These studies show that the insect intestinal microbiota, consisting of ~25 phylotypes with just a few dominant bacterial species, is much less complex than our own. On page 777 of this issue, Ryu *et al.* (*5*) exploit this limited microbial diversity and the genetic tools available in *Drosophila* to dissect the mutualistic relationship between the gut microbiota and their host.

Insects rely primarily on innate immune responses to control microbial infection. One of the best-studied mechanisms of immune protection in *Drosophila* is the inducible production of a battery of antimicrobial peptides. Production of these peptides is regulated by two signaling cascades—the Toll and immune deficiency pathways—that control transcription factor nuclear factor kappa B (NF-κB) homologs. In septic infection, the synthesis of antimicrobial peptides occurs primarily in the fat body (equivalent to the vertebrate liver). After oral infection with pathogenic microbes, expression of antimicrobial peptides can also occur in intestinal epithelial cells (*6*–*9*). However, under conventional culture conditions for flies, gut-specific expression of antimicrobial peptides is very low, even after ingesting nonpathogenic bacteria. Instead, reactive oxygen species are generated in intestinal epithelial cells to prevent growth of ingested, nonpathogenic microbes (*10*).



Gut microbes. Caudal inhibits the expression of antimicrobial peptide genes in the fly gut, even though the immune deficiency pathway is activated by resident gut microbes (**left**). In the absence of Caudal, antimicrobial peptides are produced, altering the composition of the bacterial population and resulting in apoptosis of the gut epithelium and increased mortality (**right**). *Lp, Lactobacillus plantarum; Lb, Lactobacillus brevis; Ap, Acetobacter pomorum.*

Ryu *et al.* investigated why antimicrobial peptide production is unaffected by resident intestinal microbiota in *Drosophila*, and the physiological consequence of this regulation. Surprisingly, bacteria normally resident in the gut activate the immune deficiency pathway in intestinal epithelial cells. Yet, this does not induce the expression of antimicrobial peptide genes.

Instead, the homeobox transcription factor Caudal, well known for its role in the development of the gastrointestinal tract (*11*), represses antimicrobial peptide gene expression (see the figure). Blocking Caudal expression in intestinal cells by RNA interference (RNAi) increased production of antimicrobial peptides, causing profound changes in the gut's bacterial population, particularly in two species. The *A911* strain of *Acetobacteraceae*, a dominant member of the gut microbial community (more than 10^5 bacteria per gut) in wild-type flies, was greatly reduced (to less than 10^3 bacteria per gut) in flies where Caudal expression was disrupted by RNAi (*Caudal-RNAi*). By contrast, the *G707* strain of *Gluconobacter*, a minor constituent of the gut flora in wild-type animals, increased (to more than 10^4 per gut) in the *Caudal-RNAi* flies. *A911* bacteria were also sensitive to a

The authors are in the Division of Infectious Disease, University of Massachusetts Medical School, Worcester, MA 01605, USA. E-mail: neal.silverman@umassmed.edu

synthetic antimicrobial peptide, Cecropin A, whereas *G707* was much less so. And expression of just one antimicrobial peptide (Diptericin or Cecropin A) in the gut of wild-type flies, by way of a transgene, caused a similar shift in the bacterial population.

Additionally, enhanced growth of *G707*, caused by production of antimicrobial peptides in the gut, was detrimental to the host. In the *Caudal-RNAi* flies, apoptosis of intestinal cells in the gut increased and fly survival decreased. These changes required the presence of *G707* bacteria, because apoptosis and survival returned to near-normal levels in germ-free animals (it is unclear how this change in bacterial population induces apoptosis). Whereas feeding germ-free animals *G707* bacteria induced cell death and mortality, feeding them "normal" microbiota (that are resident in wild-type animals, such as *A911*) did not induce cell death or changes in host survival. Moreover, *G707* fed to conventionally reared animals (with normal gut microbiota) did not induce any apoptosis. Indeed, germ-free animals first colonized with *A911* did not support the growth of *G707* and did not exhibit cell death after inoculation with *G707*.

The experiments by Ryu *et al.* elegantly demonstrate that the normal flora in the fly gut is sufficient to suppress the growth of pathogenic bacteria, a phenomenon referred to as colonization resistance. In humans, alterations in gut microbiota communities (such as that following antibiotic treatment) are theorized to lead to loss of colonization resistance and the expansion of minor gut microbial residents and other pathogens (*12*). This failure of colonization resistance has been linked to pathology induced by *Clostridium difficile* as well as infections in neutropenic patients. The data presented by Ryu *et al.* clearly establish the important role that microbiota play in their own proper maintenance, the ability of this microbial consortium to support and sustain health, and the critical role that properly regulated host immune responses play in supporting this microbial consortium.

References

1. P. B. Eckburg *et al.*, *Science* **308**, 1635 (2005).
2. F. Guarner, *Curr. Opin. Gastroenterol.* **21**, 414 (2005).
3. C. R. Cox, M. S. Gilmore, *Infect. Immun.* **75**, 1565 (2007).
4. C. Ren *et al.*, *Cell Metab.* **6**, 144 (2007).
5. J.-H. Ryu *et al.*, *Science* **319**, 777 (2008); published online 24 January 2008 (10.1126/science.1149357).
6. N. T. Nehme *et al.*, *PLoS Pathog.* **3**, e173 (2007).
7. A. Basset *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3376 (2000).
8. P. Liehl *et al.*, *PLoS Pathog.* **2**, e56 (2006).
9. K. Senger, K. Harris, M. Levine, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15957 (2006).
10. E. M. Ha *et al.*, *Science* **310**, 847 (2005).
11. J. A. Lengyel, D. D. Iwaki, *Dev. Biol.* **243**, 1 (2002).
12. U. Stiefel, C. J. Donskey, *Curr. Infect. Dis. Rep.* **6**, 420 (2004).

Published online 24 January 2008
10.1126/science.1154209
Include this information when citing this paper.

PHYSICS

From Complexity to Simplicity

A combination of positively and negatively charged current carriers may provide a key to understanding cuprate superconductors.

Sudip Chakravarty

HAMLET: Do you see yonder cloud that's almost in shape of a camel?

POLONIUS: By th' mass, and 'tis like a camel indeed.

HAMLET: Methinks it is like a weasel.

POLONIUS: It is backed like a weasel.

HAMLET: Or like a whale.

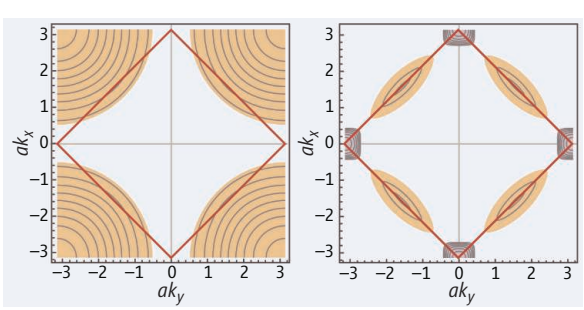
POLONIUS: Very like a whale.

—*William Shakespeare*

More than 20 years ago, Bednorz and Müller discovered superconductivity in copper oxides at remarkably high temperatures (*1*). Since then, physicists have struggled to understand the mechanisms at work. Recently, a set of experiments on cuprates in high magnetic fields (*2*–*6*) has completely changed the landscape of research in high-temperature superconductors (HTSs). In particular, the data suggest that the current carriers are both electrons and holes, when in fact the materials are "hole doped"—i.e., the current carriers should be positively charged. Moreover, the data cannot be reconciled with an important theorem about how electrons are organized in materials (*7*) unless one assumes that the signals arise from a combination of both holes and electrons. Until now, physicists have not been able to decide whether the cuprates, in Shakespeare's terms, are camels or whales; in fact, these experiments foreshadow a remarkable degree of simplicity in these complex materials.

The cuprates start out as insulators and become superconductors when doped with additional charge carriers. These so-called Mott insulators insulate by virtue of strong repulsive Coulomb interaction and need not break any symmetries in the lowest energy state, the ground state. A symmetry of a system is a transformation, such as a translation or a rotation, that keeps it unchanged. Such a symmetry is said to be broken, or spontaneously broken, if the system does not obey the symmetry of the underlying fundamental physical nature of the material; for example, a ferromagnet breaks the spin-rotational symmetry with its magnetization pointing in a definite direction. The notion of symmetry and broken symmetry finds many deep applications in physics.

Soon after the discovery of the cuprate superconductors, Anderson proposed (*8*) that their parent compounds begin as a featureless spin liquid that does not break any symmetries, called the resonating valence bond (RVB) state: "The preexisting magnetic singlet pairs of the insulating state become charged superconducting pairs when the insu-



In the pocket. Electron states in cuprates, with constant energy curves (black) plotted in momentum coordinates (ak_x, ak_y) in units of $\hbar/a$, where $\hbar$ is Planck's constant divided by 2π, and a is the lattice spacing. (**Left**) The Fermi surface separates the occupied states (light blue) from the unoccupied states (orange); the latter can act as positively charged carriers called holes. (**Right**) When the material is "underdoped," the Fermi surface may reconstruct, which forms two hole "pockets" (adding the four halves) (orange) and one electron "pocket" (adding the four quarters) (purple), as revealed by recent experiments.

The author is in the Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA. E-mail: sudip@physics.ucla.edu

lator is doped sufficiently strongly" (*8*). Unfortunately, experiments show that the insulating phase is a simple antiferromagnet in which the spins are arranged in antiparallel manner, that is, with a broken symmetry. The materials remain antiferromagnets for a range of doping, and then, after a sequence of not well understood states as a function of doping, they become superconductors.

How this plays out experimentally can be understood by looking at the Fermi surface, a fundamental concept in condensed matter physics. The Fermi surface differentiates the occupied electronic states from the unoccupied states (in coordinates of momentum rather than "real" space). Electrons fill the Fermi surface up to some highest occupied energy called the Fermi energy (see the figure). The excitations from the Fermi surface (e.g., when a current flows) are called Landau quasiparticles (quasi, because they are combinations or superpositions of real particles).

The new experimental work (*2*–*6*) yielded measurements of the oscillations that arise from energy levels created by imposing a magnetic field on the material (the Landau levels). As the magnetic field is increased, the highest fully occupied levels sweep past the Fermi energy, and the system periodically returns to itself, hence the oscillation in physical properties. The oscillations of the Hall resistance (*2*, *4*), capable of detecting the sign of the charge carriers, seem to show the presence of electron and hole pockets in the Fermi surface, suggesting that it undergoes some kind of reconstruction. This requires a global deformation of the Fermi surface, most likely a broken symmetry, and is probably a general feature of underdoped HTS materials.

One might complain that these high field measurements are still considerably below the upper critical field where superconductivity disappears (about 100 T or more) and are affected by the complex motion of vortices generated by the magnetic field. This may be true, but quantum oscillations in many superconductors are observed at fields as small as half the critical field, with the oscillation frequencies unchanged from the nonsuperconducting state (with an increased damping, however). It is also known that the quasiparticles of HTSs do not form Landau levels (*9*). Thus, it is very likely that the quantum oscillation experiments are accessing the normal state beyond the realm of superconductivity. But what kind of state? As the oscillations definitively point to both electron and hole pockets, it cannot be a conventional Fermi surface, rather one that has undergone a reconstruction due to a broken symmetry at variance with the RVB picture (*10*).

We may be finally beginning to understand these superconductors after two decades. The fly in the ointment is the lack of observation of electron and hole pockets in other measurements in hole-doped superconductors (in angle-resolved photoemission spectroscopy, for instance) that are also capable of measuring Fermi surfaces [see, however, the work on electron-doped materials (*11*)]. Missing so far in experiments are also the higher frequency oscillations that must arise from the hole pockets, not just the electron pockets (*4*). With further experimental work, we should be able to tell just what kind of animal we are dealing with.

References and Notes

1. J. G. Bednorz, K. A. Müller, *Z. Physik B* **64**, 189 (1986).
2. N. Doiron-Leyraud *et al.*, *Nature* **447**, 565 (2007).
3. A. F. Bangura *et al.*, http://arxiv.org/abs/0707.4461 (2007).
4. D. LeBoeuf *et al.*, *Nature* **450**, 533 (2007).
5. D. A. Yelland *et al.*, http://arxiv.org/abs/0707.0057 (2007).
6. C. Jaudet *et al.*, http://arxiv.org/abs/0711.3559 (2007).
7. J. M. Luttinger, *Phys. Rev.* **119**, 1153 (1960).
8. P. W. Anderson, *Science* **235**, 1196 (1987).
9. M. Franz, Z. Tesanovic, *Phys. Rev. Lett.* **84**, 554 (2000).
10. S. Chakravarty, H.-Y. Kee, http://arxiv.org/abs/0710.0608 (2007).
11. N. P. Armitage *et al.*, *Phys. Rev. Lett.* **87**, 147003 (2001).

12. Supported by NSF grant DMR-0705092. I would like to thank H-Y. Kee, E. Abrahams, N. P. Armitage, R. B. Laughlin, Z. Tesanovic, and J. Zaanen for important comments.

10.1126/science.1154320

CHEMISTRY

Taking a Selective Bite Out of Methane

A clever experiment enables the selective cleavage of a specific bond in a surface reaction.

C. Buddie Mullins and Greg O. Sitz

For more than a decade, scientists have attempted to use lasers to drive the selectivity of chemical reactions, with mixed success. On page 790 of this issue, Killelea and co-workers report a successful realization of this approach for a surface reaction (*1*).

For laser-driven selective chemistry to work, the laser must excite a particular vibrational mode in a molecule, and the energy must reside in that particular vibrational mode long enough for the molecule to dissociate and/or collide with another reactant molecule or surface. However, the energy in multi-atom molecules may be redistributed as a result of coupling between the vibrational modes in the molecule, and the excitation energy provided by the laser frequently may thus reside only transiently in the desired vibrational mode. As a result, demonstrating bond-selective chemistry of polyatomic molecules has proved harder to achieve than originally imagined (*2*).

Nevertheless, some success has been achieved for gas-phase reactants. Nearly two decades ago, Crim and co-workers described the first example of a bond-selective bimolecular reaction (*3*). When the authors selectively excited the OH or OD stretch in monodeuterated water (HOD) and then scattered the energetic molecules from H atoms, they observed two chemical reactions:

$$H + HOD \rightarrow OD + H_2 \text{ (path A)}$$

$$H + HOD \rightarrow OH + HD \text{ (path B)}$$

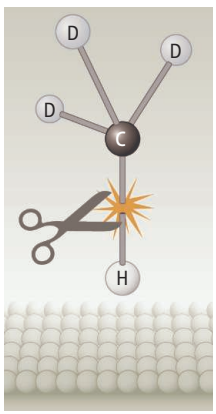
When the OH stretch was excited, path A occurred more than 100 times as frequently as path B; in contrast, excitation of the OD stretch produced reaction products almost exclusively via path B. The coupling between the OD and OH stretches is small, so that the excitation can be localized in a chosen vibration.

Bond-selective chemistry on solid surfaces is expected to be more difficult to achieve than that between gas-phase reactants, because energy can easily dissipate via the solid, and the coupling between the molecule and the surface is therefore often strong. Killelea and co-workers now greatly advance the field by reporting the C-H bond–selective dissociation in triply deuterated methane (CHD_3) on a Ni(111) surface. The dissociative adsorption of methane (CH_4), and hence the breaking of a C-H bond, on nickel is an elementary and rate-limiting surface chemical reaction used to produce molecular hydrogen from natural gas. Because of the importance of this reaction, there have been many studies

C. B. Mullins is in the Department of Chemical Engineering, University of Texas at Austin, Austin, TX 78712, USA. G. O. Sitz is in the Department of Physics, University of Texas at Austin, Austin, TX 78712, USA. E-mail: mullins@che.utexas.edu; gositz@physics.utexas.edu

of the reaction of methane with various metal surfaces (*4*–*10*).

In their study, Killelea and co-workers excited the C-H bond in a CHD_3 molecule using an infrared laser. They varied the translational energy of the molecules independently. In this case, the energized vibration is sufficiently long-lived for almost all vibrationally excited molecules to collide with the surface before deexcitation; coupling to other vibrational modes in the molecule is minimal. For impulsive (or direct) scattering of the molecule, the interaction time is comparable to a single vibrational period. But how can potential bond-selective chemistry be detected?



Cutting in the right place. Bond-selective chemistry using lasers has been difficult to achieve. Killelea and co-workers now show how the C-H bond in the triply deuterated methane molecule, CHD_3, can be cut selectively. Preparation of the surface before exposing it to the selectively excited CHD_3 molecules is key to detecting the selectivity.

Previous researchers simply measured the amount of carbon adsorbed to the surface from complete dissociation of the methane, but this method cannot be used to establish which bond was broken first. Instead, Killelea and co-workers exploit the fact that—if the excited molecules dissociate in a bond-selective manner—one would expect enhanced cleavage of the C-H bond in the excited CHD_3 (compared to the unexcited molecule), because this bond is in an excited state. The authors make use of two key studies conducted by Ceyer and co-workers (*6*, *7*) to devise their detection scheme.

Ceyer and co-workers showed that impinging translationally energetic methane on Ni(111) at 150 K led to stable adsorbed methyl (CH_3) groups and hydrogen adatoms on the surface (*6*). In a later study, they first placed deuterium (D) below the surface (by exposure to a D atom source with subsequent exposure to an energetic rare-gas beam to remove surface-adsorbed D) and then exposed the surface to the energetic methane beam. Subsequent heating of this surface produced only DCH_3; that is, the adsorbed H atom from methane dissociation did not recombine with the methyl group to form CH_4.

Using this knowledge, Killelea *et al.* created a Ni(111) sample populated mostly by subsurface D and held this sample at 120 K before impingement by either their CHD_3 beam with the excited C-H stretch or the unexcited CHD_3 molecular beam. If the C-H bond cleaved during impingement, a subsequent temperature-programmed reaction measurement would yield CD_4 (20 atomic mass units). In contrast, if a C-D bond in the CHD_3 molecule was broken, the subsurface D atom would react with the doubly deuterated surface methyl group (CHD_2) to yield CHD_3 (19 atomic mass units). Thus, by comparing the relative amounts of 20 and 19 atomic mass units produced, the efficacy of bond-selective dissociation could be determined. The relative yield of C-H bond cleavage in the excited molecules increases by more than a factor of 90 compared to that in the unexcited molecules—a clear demonstration of bond-selective surface chemistry.

These exciting results suggest other interesting experiments investigating bond-selective surface chemistry. What would be the fate of a molecule with an excited, long-lived vibrational mode that was physically adsorbed on the surface (and hence interacts much longer with the surface)? Physically adsorbed (that is, van der Waals–bound) species are often precursors to dissociative chemisorption in catalytic reactions (*8*–*10*). Would the solid surface perturb the physically adsorbed molecule sufficiently to induce additional redistribution of the energy within the molecule? Would the surface quench the energy too rapidly for reaction to occur via strong coupling to the solid? For molecular species covalently bound to the surface, the coupling is indeed too strong and the lifetime of the localized excitation is much too short; however, for physically adsorbed species (such as methane), the coupling will be much weaker, and bond-selective chemistry may indeed be possible.

References and Notes

1. D. R. Killelea *et al.*, *Science* **319**, 790 (2008).
2. R. N. Zare, *Science* **279**, 1875 (1998).
3. A. Sinha *et al.*, *J. Chem. Phys.* **92**, 6333 (1990).
4. C. T. Rettner *et al.*, *Phys. Rev. Lett.* **54**, 2716 (1985).
5. R. R. Smith *et al.*, *Science* **304**, 992 (2004).
6. M. B. Lee *et al.*, *J. Chem. Phys.* **85**, 1693 (1986).
7. A. D. Johnson *et al.*, *Science* **257**, 223 (1992).
8. D. C. Seets *et al.*, *J. Chem. Phys.* **107**, 3986 (1997).
9. D. C. Seets *et al.*, C. B. Mullins, *J. Chem. Phys.* **107**, 10229 (1997).
10. G. O. Sitz, C. B. Mullins, *J. Phys. Chem. B.* **106**, 8349 (2002).
11. We gratefully acknowledge the support of the NSF, the U.S. Department of Energy, and the Defense Threat Reduction Agency.

10.1126/science.1153906

MATERIALS SCIENCE

Toward Flexible Batteries

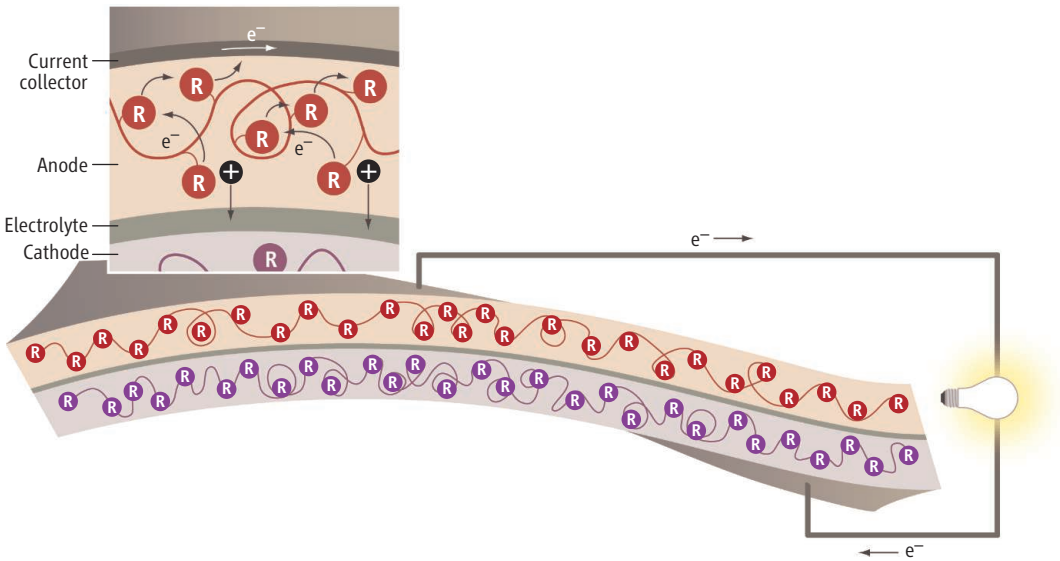
Hiroyuki Nishide and Kenichi Oyaizu

Flexible batteries are under development for use in bendable electronic equipment.

The design of soft portable electronic equipment, such as rollup displays and wearable devices, requires the development of batteries that are flexible. Active radio-frequency identification tags and integrated circuit smart cards also require bendable or flexible batteries for durability in daily use. Several routes toward the development of flexible batteries are being explored. Some involve batteries made mostly or entirely from plastic, with the added advantage of avoiding ignitable and toxic substances such as lithium and lead. An inorganic primary battery that can be bent like a piece of paper has been developed for disposable-card applications (*1*, *2*). However, primary batteries produce current by a one-way chemical reaction and are not rechargable; their usefulness of in portable electronic equipment is therefore limited. Rechargeable secondary batteries are generally used to power portable equipment. There have been recent efforts to make secondary lithium-ion batteries into thin films while maintaining their high energy capacity (*3*).

Making a bendable lithium-ion battery requires the development of soft electrode-active materials, such as metal oxide nanoparticles or nanocoatings for cathodes and lithium foil or nanocarbon materials for anodes (*4*, *5*). Virus-templated Co_3O_4 nanowires have been shown to improve the capacity of thin, bendable lithium-ion batteries (*6*). The charging/discharging process of batteries is generally dominated by the electron and counterion transport at the surface of the electrodes. By using nanostruc-

The authors are in the Department of Applied Chemistry, Waseda University, Tokyo 169-8555, Japan. E-mail: nishide@waseda.jp



Example of a flexible plastic battery. The R groups in the cathode and in the anode have different redox potentials. During the charging process, charge is stored by oxidizing R groups at the cathode and reducing R groups at the anode. The output voltage of the battery corresponds to the gap between the redox potentials. The curves connecting the R groups are polymer chains, which give flexibility. Many R groups are attached to the polymer chain, so that electrons can hop between neighboring R groups to produce the output current.

tured inorganic materials, the rates of the electron and counterion transport are increased. The nanostructured inorganic materials are also soft and bendable, helping in the development of bendable lithium-ion batteries. However, transport rates at inorganic electrodes are often limited by the slow kinetics of ion intercalation and migration in these materials (*7*). The transport properties have been improved somewhat by using nanoarchitectured electrodes with large surface areas (*8*).

Plastic batteries using organic electrodes have inherent advantages over lithium-ion batteries, because the organic materials are flexible and their properties can be tuned through chemical synthesis. Several avenues toward such batteries have been explored, and test batteries have been demonstrated.

Research on plastic batteries has a long history, which began with the discovery of the electric conductivity of doped polyacetylene in the late 1970s. However, early attempts to develop organic polymer batteries based on polyacetylene (*9*) did not lead to commercialization because of the chemical instability of the doped and virgin polyacetylene. Electrically conducting polymers—such as polyaniline, polypyrrole, and their derivatives—have also been examined as electrode-active materials, on the basis of their reversible electrochemical doping behaviors. However, no successful battery will emerge from this work, because the doping levels are insufficient, the resulting redox capacities are low, and the doped states are not chemically stable, leading to self-discharge and degradation of the rechargeable properties of the resulting batteries.

Another effort to making plastic batteries uses an electrolyte layer sandwiched between thin layers of polymers that have low conductivity but incorporate redox-active groups, with a view to increase the overall redox capacity of the battery (see the figure). In this case, the polymer backbones provide a matrix, rather than a conducting path, to interconnect innumerable redox sites for the hopping of electrons by a self-exchange mechanism, resulting in the storage and transport of charge in a homogeneous solid. However, concurrent chemical reactions such as chemical bond cleavage and formation, accompanied by the redox reaction of closed-shell molecules, generally result in an electrochemically irreversible reaction characterized by slow kinetics (*10*, *11*). Organic electrode reactions have to be reversible in order for organic batteries to be developed that can be rapidly charged and discharged with a large current.

Another approach to the development of organic electrode-active materials is based on the large redox capacity of aliphatic redox polymers, which are densely populated with pendant redox-isolated sites (*12*, *13*). Purely organic polymers based on open-shell molecules called radicals have also been studied. These radicals, such as nitroxides and galvinoxyls, allow fully reversible one-electron redox reactions featuring fast electrode kinetics, reactant recyclability, and high redox capacity. The radical polymers act as both cathode- and anode-active materials, because their redox potentials can be tuned by varying the organic substituents (*14*). These "radical polymer batteries" can be charged in less than 30 s, can generate burst power at rated voltages, and can be transformed into completely flexible, foldable, and semi-transparent batteries (*15*).

The technologies of lithium-ion batteries are regarded as mature, but they have limitations because of safety concerns (prompted by accidents involving ignition and explosion) and because it is very difficult to make highly flexible lithium-ion batteries. Compared with their inorganic counterparts, plastic batteries are safer, adaptable to both roll-to-roll and inkjet printing processes, and comparatively easy to dispose (they can be burned away without toxic gas and ash formation); furthermore, they can be fabricated from less-limited resources. Plastic batteries are intrinsically more bulky than lithium-ion batteries. However, if the batteries are sufficiently light, flexible, and environmentally benign, then bulkiness will not be a significant problem. We may then see the commercialization of flexible plastic batteries for use in electronic equipment.

References

1. R. Huvila, presentation during Session 6 at the *3rd Global Plastic Electronics Conference and Showcase*, Frankfurt, Germany, 29 October 2007.
2. See www.enfucell.com/Products.htm.
3. N. J. Dudney *et al.*, *J. Electrochem. Soc.* **154**, A805 (2007).
4. B. Scrosati, *Nat. Nanotechnol.* **2**, 598 (2007).
5. J. -M. Tarascon *et al.*, *Nature* **414**, 359 (2001).
6. K. T. Nam *et al.*, *Science* **312**, 885 (2006).
7. A. S. Aricó *et al.*, *Nat. Mater.* **4**, 366 (2005).
8. P. L. Taberna *et al.*, *Nat. Mater.* **5**, 567 (2006).
9. A. J. Heeger *et al.*, *Chem. Commun.* 317 (1981).
10. R. Berridge *et al.*, *J. Phys. Chem. B* **110**, 3140 (2006).
11. H. -K. Song *et al.*, *Adv. Mater.* **18**, 1764 (2006).
12. S. A. Miller *et al.*, *J. Am. Chem. Soc.* **126**, 6226 (2004).
13. J. M. Cooper *et al.*, *J. Am. Chem. Soc.* **126**, 15362 (2004).
14. H. Nishide *et al.*, *Electrochem. Soc. Interface* **14**, 32 (2005).
15. T. Suga *et al.*, *Chem. Commun.* 1730 (2007).

10.1126/science.1151831

INTRODUCTION

Reimagining Cities

CITIES ARE NOW HOME TO HALF OF THE WORLD'S 6.6 BILLION HUMANS. BY 2030, nearly 5 billion people will live in cities. This special issue explores the enormous implications of the mass embrace of city life. News articles offer a look at how cities are tackling specific problems, a set of Reviews and Perspectives examines trends and demographics arising from the urban transformation.

As Grimm *et al.* (p. 756) show, cities are hot spots of production, consumption, and waste generation. Already, according to the United Nations, cities are responsible for 75% of global energy consumption and 80% of greenhouse gas emissions. Without careful investment and planning, megacities (those with more than 10 million inhabitants) will be overwhelmed with burgeoning slums and environmental problems. There are advantages to city life, such as the relative proximity of health care (Dye, p. 766) and jobs. However, Mace (p. 764) describes continuing costs in terms of fertility, and Bloom *et al.* (p. 772) challenge a commonly accepted perception that urbanization fuels economic growth.

CITIES ONLINE

For an expanded view of this section, see **www.sciencemag.org/cities**

Cities have taken novel approaches to dealing with urbanization. A News article (p. 740) explores how the Chinese government is encouraging a variety of schemes, including the development of "eco-cities." Other News items highlight success stories, including Bogotá's reduction of traffic fatalities (p. 742), London's reduction of traffic jams (p. 750), and Mexico's efforts to alleviate urban poverty (p. 754).

The pace of urbanization is accelerating throughout the developing world (Montgomery, p. 761). One of the most pressing issues for these cities is the provision of clean water and sanitation. News articles feature three cities with different solutions: Durban (p. 744), Salvador, (p. 745), and Phnom Penh (p. 746). Sometimes, however, a lack of money and powerful lobbies can thwart the best intentions, as people in Kolkata have learned as they try to clean up their city's foul air (p. 749).

How will cities evolve? Batty (p. 769) shows that in spite of the apparently amorphous growth of urban sprawl, resilient patterns emerge. He advocates the use of complex systems analysis in future urban planning. Preparing for natural disasters, and recovering from them, will also challenge planners—especially because many of the world's largest cities lie on coasts and are vulnerable to flooding as the climate warms (p. 748).

Someday, cities may grow their own crops and raise their own livestock in vertical farms (p. 752). Next-generation hybrid cars could help cut greenhouse gas emissions (p. 750). A more distant dream is a "supercity" that relies on superconducting electricity cables and liquid hydrogen for its energy needs (p. 753). Futuristic concepts, perhaps, but the time has come for a radical rethink of our concept of cities and their place in the global environment.

–CAROLINE ASH, BARBARA R. JASNY, LESLIE ROBERTS, RICHARD STONE, ANDREW M. SUGDEN

CREDIT: GETTY

Cities

CONTENTS

News

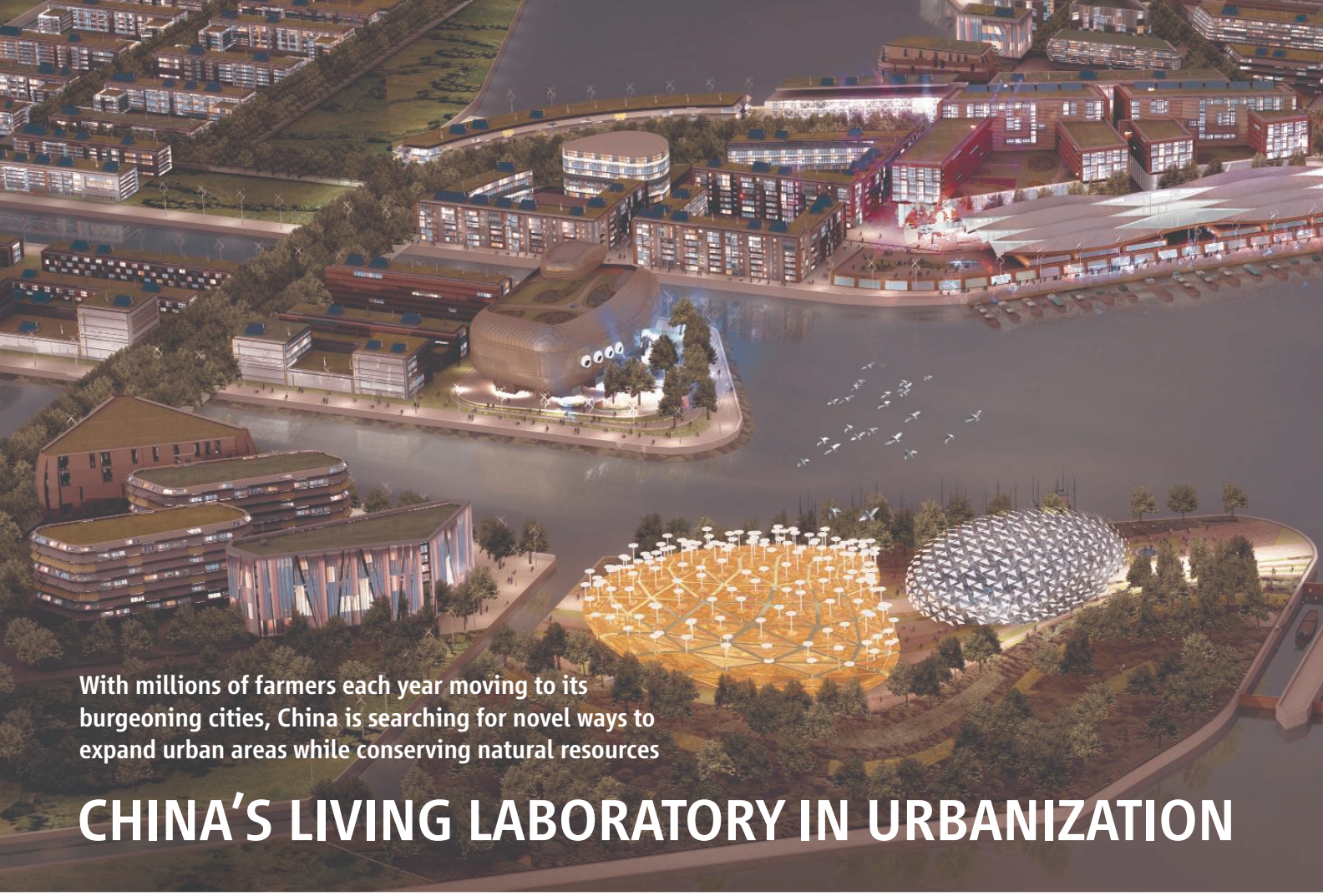
Reviews

Perspectives

See also related Editorial page 697

Science

Published by AAAS



With millions of farmers each year moving to its burgeoning cities, China is searching for novel ways to expand urban areas while conserving natural resources

CHINA'S LIVING LABORATORY IN URBANIZATION

CHONGMING ISLAND, CHINA—Standing in a sea of marsh grass at the eastern tip of Chongming Island, in the mouth of the Yangtze River, it's easy to forget that this wilderness lies within the boundaries of Shanghai municipality. Tidal mud flats, feeding grounds for migratory birds on the East Asian–Australasian Flyway, reach toward the East China Sea as far as the eye can see. A million shore birds pass through every year, including the endangered black-faced spoonbill. To the west, scattered sparsely across the 1041 square kilometers of Chongming, the world's largest alluvial island, are villages, paddies, and orchards.

Shanghai is about to burst another seam and spill onto this peaceful isle. A bridge-tunnel link scheduled for completion in 2009 will turn a torturous 3-hour car-and-ferry trip from downtown—just over the horizon to the south—into a 30-minute commute. And with well over 300,000 new residents each year swelling one of the world's biggest cities—Shanghai has more than 17 million inhabitants—development of Chongming's wide-open spaces is inevitable.

Shanghai is hoping to show that development can be environmentally responsible with the world's first "carbon neutral" city, in which carbon emissions would be completely offset by carbon absorption. Construction of Dongtan Eco-city will begin early this year on land adjacent to Chongming's wetlands. Dongtan's backers hope it will offer a new model that contrasts with China's haphazard urbanization of the past 2 decades. Some planners familiar with practices here, however, wonder if Dongtan's ambitious aims can be fully realized.

Dongtan is one of a half-dozen or so eco-cities on the drawing boards as Chinese leaders cope with one of the fastest urbanization rates in the world. The leadership now realizes that unchecked urban sprawl threatens the country's environment and security, says Niu Wenyuan, chief scientist of China's sustainable development strategy program and a counselor of the State Council. As a result, he says, the country is striving for three "zero net-growth rates": the population by 2020, urban energy consumption by 2035, and urban ecological degradation by 2050. "We still have a long way to go," Niu said at the first Xiamen International Forum on Urban Environment in Xiamen, China, last November.

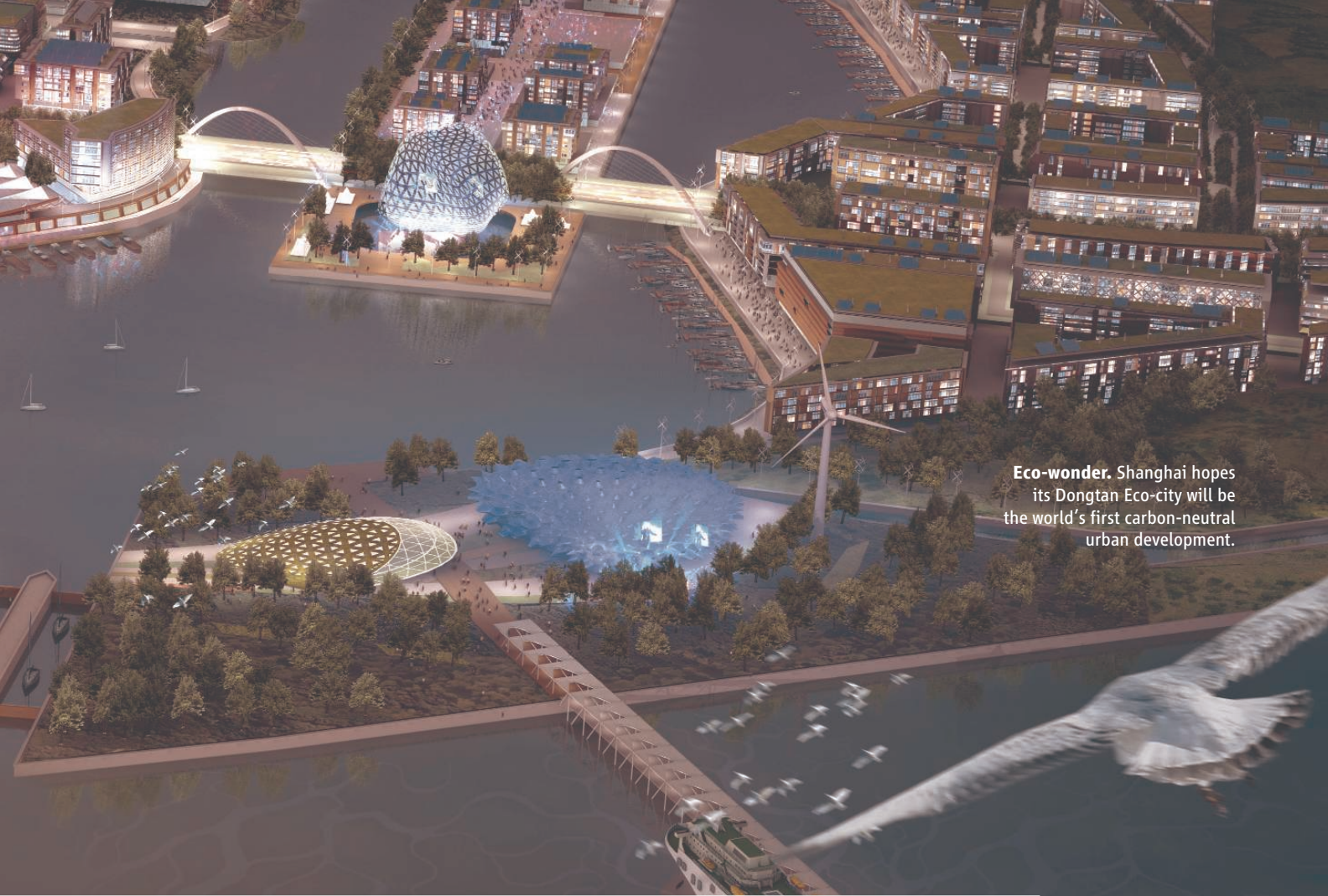
This may be China's last chance to get urbanization right, says Qiu Baoxing, vice minister of construction. "If China chooses the wrong [urbanization] model," he says, "it will [impact] the entire world."

Much of the developing world is urbanizing rapidly, but China's sheer numbers make the stakes here higher. China now has 670 cities, up from 69 in 1947 and 223 in 1980. According to United Nations statistics, China has 15 of the world's 100 fastest-growing cities with a population of a million or more (based on population growth between 1950 and 2000); India, next on the list, has eight. China has 89 cities with a population of a million or more. The United States has 37 and India 32.

The government estimates that 44% of China's population now lives in cities, but that figure does not include migrant workers registered as residing in rural areas. If they are included, "China's real urbanization rate is already around 50%," says Deng Wei, an urban economics specialist at Tsinghua University in Beijing. By 2020, some 60% of the population will live in cities, according to government estimates. Each year, about 12 million farmers move to cities, Niu says: "The biggest agrarian

CREDIT: ARUP

Published by AAAS



Eco-wonder. Shanghai hopes its Dongtan Eco-city will be the world's first carbon-neutral urban development.

society in the world is becoming the biggest urban society in the world."

The implications are enormous. "Urbanization concerns the use of resources, human lifestyles and culture, economic efficiency, modernization, the welfare system, everything," says Qiu, an expert in economics and urban planning. According to Zhao Jinhua of the Massachusetts Institute of Technology in Cambridge, China's cities weren't designed to accommodate breakneck growth, which leads to chronic problems such as water and housing shortages.

Throughout the world—China is no exception—city dwellers are typically wealthier, consume more, and produce more waste, including greenhouse gases, than people in rural areas. If China has not already done so, it will soon surpass the United States as the largest emitter of carbon dioxide. A large share comes from coal-fired power plants, but tailpipe emissions are an increasing contributor, especially in cities, Qiu says.

A generation ago, China's urbanites overwhelmingly relied on bikes and public transportation, Zhao says. But starting in the 1980s, haphazard planning spawned economic zones tailored for manufacturing but with minimal housing or shopping areas and bedroom communities with few job opportunities—all of which encouraged commuting by car. Then in the 1990s, dozens of "new towns" sprang up on the outskirts of cities, most "designed with the car as the dominant mode of transportation," says Zhao, who is also executive commissioner of the China Planning Network, an organization of Chinese and overseas scholars who study China's urbanization.

Well-intentioned development has exacerbated the problem. Beijing's ring roads, for example, were supposed to ease crosstown traffic but instead have accelerated sprawl and private-car use, Qiu says. In some areas, bicycle lanes and the median strips that once separated them from traffic were sacrificed to make more room for cars.

Well-planned cities could ameliorate these problems. That means "dense and diverse" cities, Qiu says. Packing more people per square kilometer makes public transportation more feasible, he says. Apartments use resources, including energy, more efficiently than detached houses. Diversity entails what planners call mixed-use—an intermingling of residential, shopping, and office areas that creates opportunities for walking or biking to shops or work. The construction ministry and local governments are also encouraging a nascent "green building" movement that seeks to

For the birds. Dongtan aims to protect an adjacent waterfowl refuge.

CREDIT: D. NORMILE/*SCIENCE*

Published by AAAS

make better use of energy, water, and materials to minimize a building's environmental impact throughout its life cycle. These trends are converging in plans for several eco-cities, the most notable being Dongtan.

From the outset, the Shanghai government, which owns the site, has viewed Dongtan as an "eco demonstrator" of urban development existing in harmony with the environment—even on an ecological treasure like Chongming. "This is not just about saving energy or saving water," says Roger Wood, a partner in the engineering consulting firm Arup in London that is in charge of Dongtan's master planning. "It is about a holistic approach that goes right through the social, governance, education, transportation, wastewater issues—all the things that actually make a community."

Dongtan will rise on a portion of an 86-square-kilometer strip of Chongming owned by the municipal government's Shanghai Industrial Investment Corp. The city wants housing for 10,000 residents completed in time for Shanghai's 2010 World Expo, which, appropriately, will explore the theme "Better City, Better Life." The goal of the start-up phase, scheduled for completion by 2020, is a community of 80,000, businesses providing 50,000 jobs, and shops, entertainment, and cultural amenities that offer residents everything they need in Dongtan, although it's expected that some people will commute to Shanghai and some nonresidents will work in Dongtan. Eventually, the Eco-city could be extended to cover 30 square kilometers and house half a million people.

The investment corporation instructed Arup to minimize the project's ecological footprint: the land and water areas needed to provide Dongtan's resources and absorb waste. Using established technologies, the planned ecological footprint could be less than half that of a comparable conventional city, Wood says. Buildings will be properly insulated and rely on low-energy lighting and appliances. A double-piping system will provide drinking water and treated wastewater to flush toilets and irrigate vertical farms (see p. 752). The initial target is that no more than 10% of Dongtan's trash will end up in a landfill; planners would like to eventually make it the world's first zero-waste city. Most Chinese cities dump about 90% of their waste and burn the rest.

A second requirement is that all energy consumed in Dongtan comes from renewable sources. Solar panels, wind turbines, and a biomass cogeneration plant, fueled by rice husks, will generate electricity for power, heating, and cooling. Husks, currently burned or dumped, will be collected from throughout the Yangtze delta.

Calming Traffic on Bogotá's Killing Streets

With humor, education, and tough laws, this Colombian city has dramatically reduced traffic injuries and deaths

LONG BRANDED AS ONE OF THE WORLD'S MOST DANGEROUS CITIES, Bogotá, Colombia, has won plaudits for cutting its murder rate by more than 70% during the past decade. But this city of 7 million people has received far less attention for a dramatic decline in a more common danger that plagues urban areas everywhere: traffic-related injuries and deaths.

With a combination of innovative education campaigns, an overhaul of its public transportation system, strict law enforcement, and redesign of streets and highways, Bogotá has made moving from place to place safer and more efficient. "In 1997, everything was a mess and we were losing the battle," says Dario Hildalgo, a transportation engineer from Bogotá who is now with the World Resources Institute in Washington, D.C. "To solve the problems, we needed a miracle. The miracle happened."

Mark Rosenberg, the former head of injury prevention at the U.S. Centers for Disease Control and Prevention in Atlanta, Georgia, says Bogotá is a model for the world. "Bogotá is not unique in having this problem, but it is unique in solving it," says Rosenberg, who now heads the nonprofit Task Force for Child Survival and Development in Decatur, Georgia.

In a 2004 report, the World Health Organization and the World Bank blamed 1.2 million deaths and some 50 million injuries each year on road crashes. For people between the ages of 10 and 24, traffic injuries are the leading cause of death worldwide. The report projected that without major changes, deaths and injuries would increase 65% by 2020. "We have interventions that work, and we know how to bring the rates down," says Rosenberg. "There's no other opportunity like this in public health. It's as good as the best vaccines. But we need the resources"—and political will.

Silent treatment. Bogotá used mimes to shame traffic violators.

PHOTO: COURTESY OF *EL TIEMPO*; SOURCE: ROAD ACCIDENT PREVENTION FUND

Bogotá Traffic Safety History

	Accidents	Deaths	Injuries
1998	52,764	914	21,053
1999	52,327	872	22,035
2000	48,337	823	22,035
2001	42,776	764	24,265
2002	41,615	604	22,289
2003	40,175	759	22,884
2004	43,000	666	24,532
2005	35,838	564	17,249
2006	35,585	553	17,815
2007*	31,083	486	15,029

* Through October.

Rosenberg, Hildalgo, and others laud the aggressive and creative efforts of mayors Antanas Mockus and Enrique Peñalosa, who alternately ran the city from 1995 to 2003. Mockus, a former mathematician and philosopher, famously painted zebra stripes at intersections and employed mimes to shame bad behavior, pretending to pull on vehicles, for example, that blocked crosswalks at red lights. "The city was a very funny place," says Francisco José Fernandez, head of the Road Accident Prevention Fund, a private group supported by a special tax on car insurance. But Mockus had serious aims and some decidedly unfunny interventions. For one, he fired approximately 2000 traffic police. "The police department that was working on traffic was very corrupt," says Fernandez, who served as secretary of transit when Peñalosa took over in 1998.

Peñalosa, an erstwhile journalist, built on Mockus's efforts. Bogotá hired an army of 1000 to confront pedestrians who ignored red lights, cracked down on drunk drivers, built bicycle-only lanes, installed new signals, and restricted each car's access to the city center to 2 workdays a week.

Published by AAAS

The plan also calls for all vehicles in Dongtan to have zero tailpipe emissions. That will be a stretch technologically, and it will require a mind shift in middle-class aspirations. Dongtan planners hope to reduce dependence on private autos with apartment buildings laid out in clusters so that all residents are within a 10-minute walk of a shopping center and public transportation, which could be pollution-free fuel-cell buses or electric light rail.

Cars running on fossil fuels cannot achieve zero tailpipe emissions, so conventional cars would have to be parked outside city limits. Dongtan residents who wish to drive in town will have to use hydrogen fuel-cell or electric vehicles. However, such vehicles that match the performance and affordability of conventional cars are years if not decades away. Zhao wonders if enough people will be willing to give up the dream of owning a car and a detached home. One unresolved issue likely to affect car use is whether the rail line connecting Chongming to downtown Shanghai will extend to Dongtan. Deng says previous new towns lacking good public transportation links ended up encouraging private-car use.

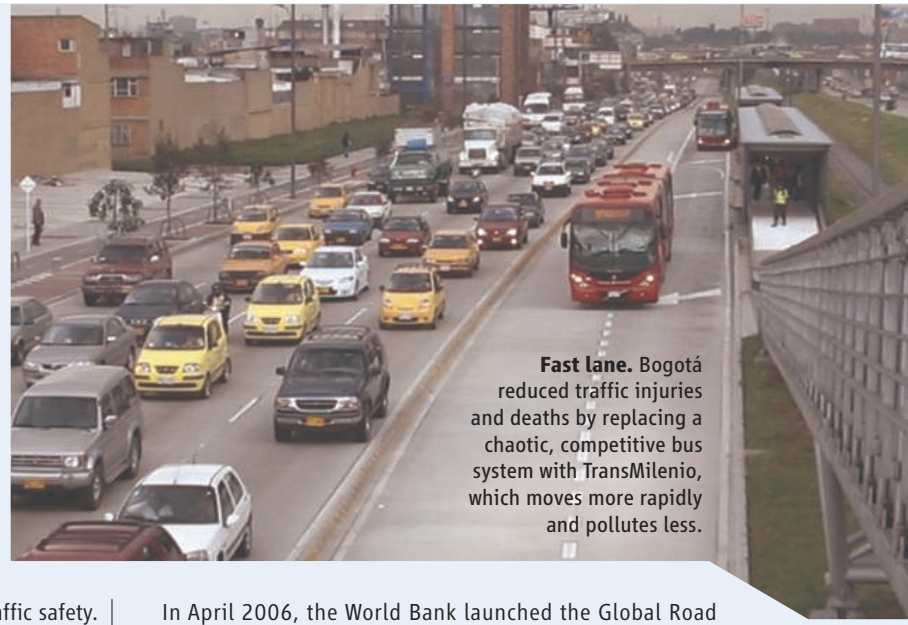
There are other concerns. Zhu Dajian, an economist who studies sustainability at Tongji University in Shanghai, says it will be a challenge turning Dongtan's impressive plans into reality without compromises: "The key issue is that the implementation is often out of the control [of the designers]." Zhu adds that although some of Dongtan's concepts and technologies could be put to use in other projects, he thinks it will be difficult to copy the model wholly because of Shanghai's financial and institutional support for Dongtan. (Officials at Arup say they are unable to disclose estimated costs or how the costs compare to those of a conventional new town.)

As they wait for Dongtan to materialize, planners welcome growing efforts to reduce energy and resource use, a trend that Qiu says will be furthered by several new national laws on planning and energy consumption. In addition, Shanghai, to alleviate traffic and promote mass transit, is considering a toll system on private cars entering downtown, similar to schemes in London and elsewhere (see p. 750).

China's urban planners realize that eco-cities, redevelopment projects, and green building efforts must be scrutinized to determine how well they enhance livability and reduce environmental costs, Qiu says. With so many cities growing so rapidly, China is already a laboratory for urbanization. Now it is poised to become an experiment in innovative urban planning as well.

–DENNIS NORMILE

With reporting by Richard Stone in Xiamen.



Fast lane. Bogotá reduced traffic injuries and deaths by replacing a chaotic, competitive bus system with TransMilenio, which moves more rapidly and pollutes less.

CREDIT: GETTY IMAGES

Perhaps most important, Peñalosa championed a new bus rapid transit system modeled after the widely celebrated one in Curitiba, Brazil. At the time, Bogotá relied on several bus companies that clogged the roads and vied for passengers. Peñalosa oversaw development of a bus rapid transit system called TransMilenio that has dedicated lanes. He also forced companies to work together by bidding for contracts and sharing revenue. Although Peñalosa met strong resistance, Hildalgo, who worked on TransMilenio, says the mayor told the bus companies, "I'm doing this with you or without you." When Peñalosa's term ended in 2001, Mockus was reelected and continued the traffic reforms.

Today, TransMilenio has only about 25% of its projected 388 kilometers in operation—funding shortfalls have slowed completion—but accounts for 18% of the transit trips in the city and moves 1.3 million passengers a day. The $750 million system has shaved about 15 minutes off the average trip, according to TransMilenio data, and has reduced emissions by replacing 1500 obsolete buses with a new fleet.

TransMilenio is one element in a broad push to improve traffic safety. Along the TransMilenio route, injuries plummeted from 18 a week in 1998 to four in 2002, notes Hildalgo. In Bogotá at large, accidents and traffic-related injuries and deaths all steeply dropped between 1998 and 2006 (see table, p. 742).

Bogotá proves that even cash-strapped cities can improve traffic safety, Rosenberg says: "Traffic deaths are not an essential consequence of growth and development."

Elsewhere, Sweden has paced the field with technologies—such as road dividers made of Mylar—and traffic-slowing strategies as part of Vision Zero, a project launched in 1997 to eliminate traffic deaths and injuries in that country. Traffic deaths, which already were low by international standards, by 2006 had dropped by 20%. "Probably the most important measure for bringing down the death rate is to build safer roads so when people make mistakes in driving they're not penalized with their lives, and they did this in Sweden," says Rosenberg. Norway and Australia now have similar programs.

In April 2006, the World Bank launched the Global Road Safety Facility to help low- and middle-income countries reduce traffic-related injuries and deaths. The fund hopes to spend $30 million per year, but to date, donors have contributed just $12 million. "It's nothing," complains Rosenberg.

As much progress as Bogotá has made, it, too, faces costly obstacles to further improving its traffic safety. The new mayor, Samuel Moreno Rojas, has promised to build a traditional rail system, an idea that the public likes but that some transportation experts worry will inevitably delay the completion of the much cheaper TransMilenio routes. And although use of private cars has dropped, motorcycles are increasingly popular and now are involved in 51% of all fatal crashes. "Motorcycles right now are a nightmare," says Fernandez. Improving motorcycle safety, he says, will require intensive courses for riders, more complex licensing tests, and stepping up enforcement—all of which cost money. "Bogotá is much more friendly now than it was 10 years ago," says Fernandez. "But there are a lot of things to do."

–JON COHEN

Published by AAAS

Clean water. A Zulu boy uses a new water dispenser to fill a bucket in one of Durban's semirural settlements.

By providing clean water and improved toilets in "township" settlements, Durban is tackling one of the remaining vestiges of apartheid

Durban's Poor Get Water Services Long Denied

DURBAN, SOUTH AFRICA—Swerving around a muddy puddle in his old Toyota, Lucky Sibiya cruises past a row of shacks in Cato Crest and stops in front of a postlike water dispenser, where a Zulu man is filling a 10-liter bucket. Nearby, a plastic roof tank is supplying water to the shack below, and, down the street, a woman is adjusting the water flow from a barrel-shaped tank perched on an old tire in front of her home. "That water is clean," says Sibiya, and for every household, 200 liters a day is free.

Sibiya advises communities such as Cato Crest—one of Durban's poorest neighborhoods—about how to get the most out of the city's eThekwini Water and Sanitation Unit. Fifteen years ago, when engineer Neil Macleod became head of the unit, Durban's water services reflected the apartheid divide that had split South Africans for decades. The wealthiest residents had First-World water service; the middle third had access to basic water and sewerage; and the poorest third in the slums and semirural areas drew water from muddy streams. Soon after majority rule began in 1994, indigent South Africans began to clamor for the services that they had long been denied. Macleod's department confronted the challenge of rapidly expanding water and sanitation services in "township" settlements while keeping its budget afloat.

At the time, a quarter of a million households in the Durban area had no access to clean water or sanitation. To jump-start improvements, Macleod got permission from the city in 1996 to provide a daily 200-liter water ration—a policy that became a national goal. Although water dispensed from standpipe posts remains free of charge, valves were installed to limit waste. Workers ran plastic piping into poor settlements and installed meters for household tanks. Families can have the spigot turn off at 200 liters or pay a metered rate beyond that limit.

The laborious pipe-laying took time, and sanitation lagged even further behind. A cholera outbreak in Durban in 2000, which killed more than 70 people and infected tens of thousands in poor neighborhoods, increased pressure on Macleod's unit to speed up water service and sanitation improvements.

Today, all but 120,000 of Durban's 3.5 million residents have access to clean water—at the very least, within a 200-meter walk. (A decade ago, some residents had to haul water as far as a kilometer.) All households should have water by the end of this year, says Macleod, but it will take 2 more years to make sure everyone has access to a proper toilet. "Water is relatively easy compared to the challenges of sanitation," says Macleod.

To tackle that problem, eThekwini Water is replacing the old "pit toilet" outhouses in many poor neighborhoods. At the time of the 2000 cholera epidemic, there were about 100,000 pit toilets in Durban, which posed disease-transmission problems when they were full. And they often were, because the hilly areas on the outskirts of the city are inaccessible to vacuum tankers that pump out deep pits. In 2003, Durban officials pledged to empty the communal pits every 5 years, and they started to research better options.

The best solution so far is urine-diversion (UD) double-pit toilets, which separate urine from feces to allow the latter to dry and decompose faster. UD toilets have shallower pits and are less costly to empty than conventional pit toilets. During the past few years, eThekwini has replaced nearly 60,000 pit outhouses with UD toilets. The utility also commissioned research into the health and environmental impact of the new toilets. Starting in 2006, eThekwini has given about $300,000 a year to the Pollution Research Group of the University of KwaZulu-Natal (UKZN) in Durban to study issues such as whether UD solid waste can be used as fertilizer. Early results are promising, says UKZN biologist Mike Smith.

Another project, funded by eThekwini and the World Health Organization, assessed the health benefits of better water and sanitation. For 12 weeks, public health workers surveyed more than 1300 households in Durban's poor areas—half with UD toilets

CREDIT: TEDDY GOUNDEN/ETHEKWINI WATER

Published by AAAS

and half without—questioning each household every 2 weeks and recording episodes of diarrhea, vomiting, worms, and skin infections. Preliminary results suggest a 30% reduction in diarrheal diseases among households with UD toilets compared with similar households using pit toilets, says Stephen Knight of UKZN's Nelson R. Mandela School of Medicine, who worked on the project with eThekwini's environmental health department and the Swedish Institute for Infectious Disease Control. Access to UD toilets helped avert an average of one diarrhea episode per person every 2 years, with the benefits of good sanitation three times greater for children under age 5 than for other age ranges, according to findings presented at a wastewater management conference last summer.

EThekwini has won wide acclaim and some criticism. Patrick Bond, director of UKZN's Centre for Civil Society, contends that eThekwini's guarantee of a small amount of free water disguises the fact that it has raised water prices for people who use more than the free 200 liters. Many poor people "can't afford to pay the high costs," Bond says. He also accuses the utility of disconnecting too many customers for not paying their water bills.

EThekwini's deputy head for customer services, Michael Singh, says water policy "has been focused on marginalized and poor communities" and its new debt-relief program forgives past debts if customers meet their monthly bills for 20 months in a row. Although Macleod concedes that water prices have risen, he claims that eThekwini has set a standard among South African cities: "We are on track now to give everyone access to a basic level of services—water and sanitation—within 2 years. Not many other cities on this continent can say that." **–ROBERT KOENIG**

PIPE DREAMS COME TRUE

A big investment in sewer connections in Salvador, Brazil, has led to a steep decline in diarrhea, a major killer of kids

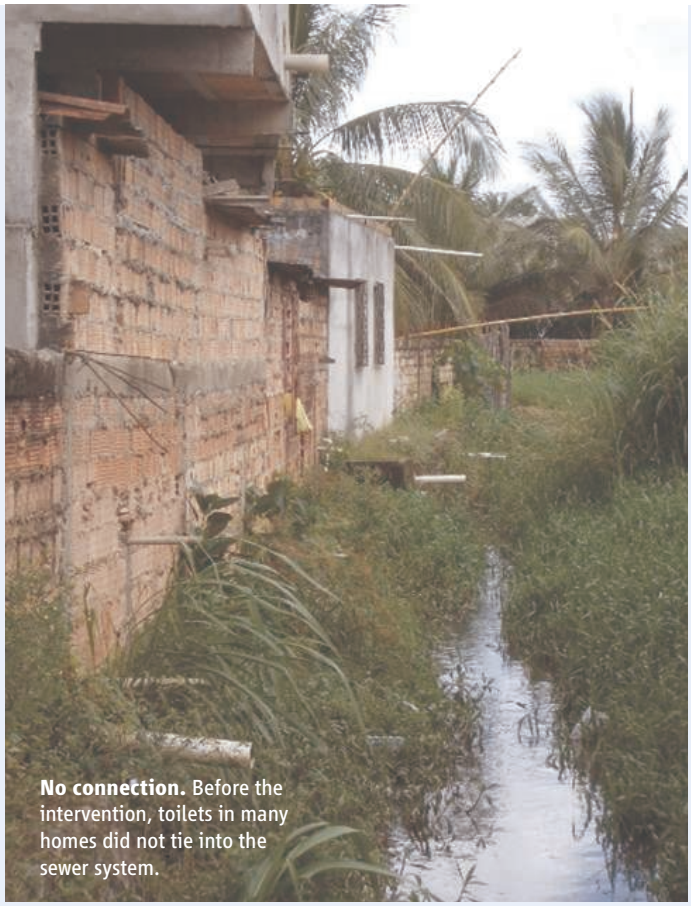
FECES HAPPEN. AND WHEN MILLIONS OF PEOPLE LIVE IN THE SAME city, a lot of it happens every day; and if it isn't properly disposed of, the health of the population goes down the toilet.

Throughout the world, more than 2 billion people lack proper sanitation, which is integrally tied to water supply for both personal hygiene and sewage systems. In all, 1.6 million children die each year from related diarrheal diseases, making it the third leading cause of mortality in children under age 15 in middle- and low-income countries. "Third World cities have a huge internal environmental problem created by fecal contamination that needs to be solved," says epidemiologist Mauricio Barreto of the Federal University of Bahia in Salvador, Brazil.

Project after project has demonstrated that cleaner water, proper sanitation, and hygiene education can improve the health of communities. Yet few studies have measured the impact of interventions citywide, and none has focused on sanitation, says Barreto. Now, an ambitious project in Salvador that he led claims to have done just that, documenting for the first time the health benefits of expanding sewer systems.

A decade ago in Salvador, 80% of the 2.5 million residents had flush toilets in their homes. Outside of wealthy neighborhoods, however, few toilets were linked to sewers; nearly three-fourths of the residents relied on septic tanks or, more commonly, flushed waste into creeks, streets, and the like. In 1996, the city received a loan from the Inter-American Development Bank and invested about $220 million in a project that laid 2000 kilometers of sewer pipes for more than 300,000 homes. "Very few cities in the world have made an investment in sanitation on this scale in such a short time," says Barreto. "We saw it as a great opportunity to see the effects on health."

As Barreto and co-workers explained in the 10 November 2007 issue of *The Lancet*, they recorded cases of diarrhea in 841 children before the intervention and in 1007 other children after the pipes were laid. The children, who were no more than 3 years old and had similar living conditions in 24 sentinel areas, were followed for more than 6 months by fieldworkers who came to their homes twice a week. Citywide, diarrhea dropped 22%, and it fell 43% in neighborhoods that had the highest diarrheal prevalence before the intervention. Neither the researchers nor the city provided hygiene education, and hygiene behavior did not explain differences in diarrhea prevalence. The study team concluded that the sewage hookups primarily prevented transmission of diarrhea by reducing exposure to feces in the "public domain"—that is, in open sewers. "[The benefit] wasn't dependent on whether your house had a connection," explains co-author Sandy Cairncross, a water engineer at the London School of Hygiene and Tropical Medicine in the U.K. "It was the extent of coverage of sewers in your neighborhood to which people could connect."



No connection. Before the intervention, toilets in many homes did not tie into the sewer system.

CREDIT: PATRICIA BORJA

Demographer Narayan Sastry of the University of Michigan, Ann Arbor, commends the researchers but questions whether the findings will apply elsewhere. He notes that other cities in Brazil have much higher connection rates to sewer systems and that it's easiest to see a dramatic impact in areas where none existed, as in some of the sentinel areas in Salvador.

Brazil's sanitation shortcomings pale, too, in comparison with many

Continued on page 746

Published by AAAS

Continued from page 745

countries in sub-Saharan Africa and south Asia that lack sewage systems and clean water supplies. In these less developed countries, simpler and cheaper interventions can have a big impact, says Albert Wright, a water and sanitation engineer and consultant in Woodbridge, Virginia. "The tools are there," says Wright, a native of Ghana who has worked at the World Bank. "It's just how you apply them."

Leaders in many cities in developing countries make the mistake of wanting to replicate the sewer systems of wealthier countries, Wright contends. "They think they must have conventional sewage like New York or Los Angeles, and they try to design a centralized system," he says. But a larger system means bigger pipes and higher costs. "They look at the high cost and have a sense of hopelessness, so they put the plans on the shelf," Wright says. These cities, he says, would be better off designing neighborhood sewer systems similar to the ones recently built in Bangkok, Thailand.

In some locales that can't afford sewage systems, basic, well-designed latrines can significantly improve sanitation. Wright points to work he does with WaterAID, a nonprofit that builds dry pits and pour-flush latrines that use pairs of pits. When one pit fills, the latrine diverts to the second. After 2 years, the contents of the first pit can be removed and used as fertilizer—at which point the pit is again ready for human waste.

Although experts disagree about how to get the most bang for the buck, there's wide consensus that there aren't enough available bucks. "Diarrhea kills more children than AIDS, tuberculosis, and malaria put together," says Cairncross. "Yet there isn't a global fund for diarrhea. All politicians would like an airport or a train station named after them but not a sewage plant." Also, there's often a disconnect between local health and water agencies, he says. Further complicating matters, private enterprises frequently run water and sanitation services. And as the Salvador study highlights, scant data exist to help municipalities make tough choices.

The United Nations has declared 2008 the International Year of Sanitation, which Barreto welcomes—albeit with a measure of skepticism. "The U.N. has been very, very shy when it comes to sanitation," he says, adding that it is no longer acceptable. "Societies have to make this a priority."

–JON COHEN

Rebuilt From Ruins, a Water Utility Turns Clean and Pure

A hard battle to root out corruption and renew devastated infrastructure has paid off in the capital of Cambodia

WHEN EK SONN CHAN WAS APPOINTED director of the Phnom Penh Water Supply Authority (PPWSA) in 1993, he faced challenges that were severe even for the developing world. A century-old infrastructure had fallen into disrepair during years of civil war and repression. Morale was low, training inadequate, and corruption rife. Stricken by the bloodletting of the Khmer Rouge, the utility was hemorrhaging water and money.

In little more than a decade, Chan turned PPWSA around. "It is an important success story," says Michael Rouse, who teaches water management and policy at the University of Oxford, U.K. Backed by political support from Cambodia's authoritarian government and many millions of dollars from international donors, Chan took several key steps. He raised water prices to help the utility become self-sustaining, stemmed the amount of water stolen or lost through leaks, and expanded services. "These are unusual achievements for urban water supply in developing cities," says Jennifer Davis, a water supply and sanitation expert at Stanford University in Palo Alto, California.

French colonists started Phnom Penh's first water utility in 1895. The utility grew 10-fold after the country gained independence in 1953, but neglect during a civil war in the 1970s took a toll on infrastructure. PPWSA suffered even more when the Khmer Rouge captured the city in 1975. Along with many of the country's intellectuals and professionals, water engineers were killed. Blueprints were lost, and much of the equipment destroyed. When Vietnamese troops ousted the Khmer Rouge in 1979, the utility limped back on line with few resources or experienced staff.

Corruption and cronyism were rampant. Water was free of charge but not freely available: Wealthier residents slipped PPWSA employees as much as $1000 to jury-rig illegal connections, which wasted water, says Chan. The government allowed PPWSA to start charging for water in 1986, but then meter readers took bribes and more than half the bills went unpaid. All told, the utility collected payment for just 28% of the water it supplied, which meant it couldn't keep up with maintenance.

Well-connected. Ek Sonn Chan has won praise for reforming the water utility in Phnom Penh.

CREDIT: ADB

Chan was an unlikely savior. He was trained as an electrical engineer, but the Khmer Rouge forced him and other professionals to work in the rice fields. The Khmer Rouge killed his entire family. After the regime fell, Chan got a job as a butcher in Phnom Penh and then entered the civil service. He worked his way up to director of municipal commerce and, in 1993, was tapped to lead PPWSA. "The demand for change came from Ek Sonn Chan himself, and he saw the potential for reform to improve performance," says

Wouter Lincklaen Arriens, a water-resources specialist at the Asian Development Bank (ADB) in Mandaluyong City, Philippines.

Chan started with a yearlong evaluation during which staffers rooted out illegal connections. He then had workers install more water meters—in 1993, just 9% of connections were metered; within 3 years, the utility had reached 85%. It wasn't easy because most residents still expected to get their water for free. Chan held public meetings and went door to door to explain that the fees—then, 10 cents per 1000 liters—would improve the system. Once, Chan recalls, a military officer who objected to paying for water pulled a gun. It was a "harrowing experience," he says. Chan left but contacted the military police and then returned with an armed squad to disconnect the illegal tap.

Prime Minister Hun Sen's support was key to raising prices. "You've got to have a government, whether national or local, that has the courage to make these changes," Rouse says. He credits PPWSA for spreading out the increases over 7 years, which eased the pain.

Streamlining the bureaucracy wasn't easy, either. But in 1996, the government granted the utility autonomy, a crucial step that allowed Chan to promote talented employees, give performance bonuses, and deal with slackers. PPWSA became more efficient; the number of staff per 1000 connections dropped from 22 in 1993 to four in 2005.

These kinds of improvements impressed donors. With $118 million from the Japan International Cooperation Agency, the World Bank, ADB, and others, Chan started repairing the infrastructure, and he installed a computer system that tracks flow and pressure to help detect leaks. Chan also added 950 kilometers of new piping. The system now delivers water 24 hours a day; adequate pressure helps prevent sewage and other contaminants from seeping into the pipes. PPWSA provides water to all of the 500,000 people who live in Phnom Penh's four inner districts and about 60% of the metropolitan area.

The work isn't over. The utility has subsidized the hookup fee for some 15,000 poor families, but many others draw from wells that are not all safe or buy expensive water from private vendors. As more poor people migrate to the city—the population of Phnom Penh is expected to double by 2034 to 3 million—the service area will need to expand fivefold to 500 square kilometers. Chan estimates the price tag at $300 million. Sanitation, the province of another agency, will become an ever-bigger problem, as will pollution dumped into the rivers upstream by the city's factories.

But Phnom Penh has already made enough progress to share its experience with other cities. PPWSA is now training staff at a utility in Vietnam's Binh Duong province. "Remember that PPWSA started out in very poor condition, worse than many other utilities, yet was able to achieve phenomenal improvements over a short period," says Lincklaen Arriens. "It is clear that there are lessons for other cities."

–ERIK STOKSTAD

Digging deep. The utility stretched its dollars by installing much of the new distribution system itself.

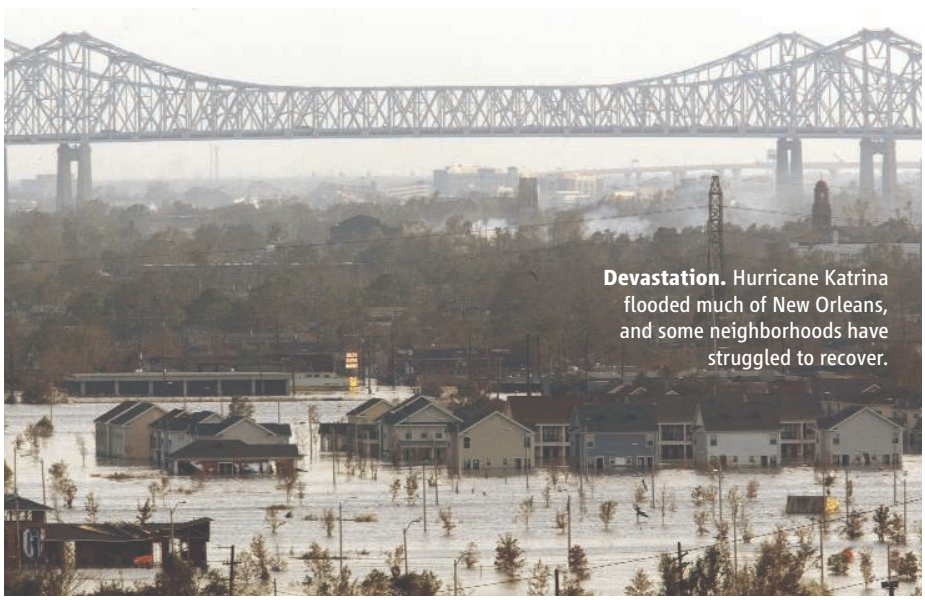
CREDIT: ADB

Published by AAAS

LIVING IN THE DANGER ZONE

The siren call of storm-battered coasts and other beautiful but hazardous human nesting areas has compelled urban planners to gird for the worst

ON A FRIDAY MORNING IN RAVAGED NEW Orleans, Louisiana, Joe Brown learned just how fiercely people value their homes. Along with several dozen other disaster experts, the veteran urban planner had been recruited by the Urban Land Institute in Washington, D.C., to develop a rebuilding plan for the city, which had been devastated by Hurricane Katrina in August 2005. At a town hall meeting that October, at the Sheraton Hotel on once-bustling Canal Street, Brown, who leads the San Francisco design and planning firm EDAW, shared his vision, formulated after interviewing 300 people over 2 weeks. Much of New Orleans was salvageable, he said, to nods and murmurs of approval. But about a quarter of the city lay in utter ruin and remained at high risk of flooding. Brown displayed diagrams that suggested turning some blocks, for the time being, into open space.



Devastation. Hurricane Katrina flooded much of New Orleans, and some neighborhoods have struggled to recover.

Reaction was swift and harsh. A council member accused Brown of aiming to "replace these fine neighborhoods with fishes and animals," he recalls. A couple of audience members rose up and declared, "All we want to do is get back to our homes." The planners were startled. "We got shock and amazement to what, to us, were fairly obvious truths," Brown says.

Although rebuilding a broken city can ease trauma, it often goes against research suggesting that reconstruction ought to proceed differently, or not at all, on land acutely vulnerable to disasters. "Cultural, social, emotional, and psychological" factors are "more important than physical safety and ecological determinism" to displaced residents, Brown says. "People believe you must always be able to recolonize." Persuading residents to resist this imperative is nearly impossible, he says.

Residents of New Orleans are not alone in their dogged determination to place themselves in harm's way. According to a report last August from the Government Accountability Office (GAO), nearly half the U.S. population lived in counties that had declared flood disasters at least six times between 1980 and 2005, and 29% made their home in a county hit by at least one hurricane in that time. Large swaths of the western United States are at risk of wildfires, such as those that emptied parts of southern California last October. People are willing to gamble by building homes on earthquake fault lines, in landslide zones, and along tornado alleys. "Population trends are increasing the nation's vulnerability to these risks," the GAO report noted dryly.

The collision of science and human psychology frustrates urban planners as they try to insulate communities from danger. One approach is to gird cities for the blow, for example, by toughening building standards and construction permits. When a disaster does strike, communities are rarely abandoned; in most cases, picking up the pieces means modifying how a city is rebuilt. "You're working at it backwards in a way," says Ken Topping, a city planning consultant and lecturer at California Polytechnic State University, San Luis Obispo. "Sometimes you have to wait for a crisis to get the idea across" to planners and residents that how and even where they live should change.

Hurricane Katrina prompted Topping and colleagues to consider the impact of levy failures as they drafted a California state disaster-mitigation plan. Levies were built in the Central Valley to protect cropland from floods. Now they are essential to shielding communities. If levies were to fail, catastrophic effects could include disruption of the drinking water supply to southern California, the mitigation plan's authors concluded, echoing past warnings.

One region where it has proved especially tough to modify development is along the Florida coast. "The attractions and the amenity of being on the water are very, very strong and overcome a lot of rational understanding of the risk," says urban planner Steven French of the Georgia Institute of Technology in Atlanta. In 1996, after Hurricane Opal tore up the Florida panhandle, Robert Deyle, who studies coastal planning at Florida State University in Tallahassee, surveyed 35 counties and 158 municipalities about their postdisaster plans. Only 12% had one in place.

Although planners may hold little sway with Floridians, insurance companies are modifying behaviors as they become less inclined to cover homeowners in exposed areas. Ceres, an investor coalition that focuses on sustainability, has reported that Allstate Insurance let 120,000 policies lapse in Florida in 2006, after canceling 95,000 the year before. Another major insurer, State Farm, declined to renew 39,000 windstorm policies in 2006. In a sign of how dire the situation has become, the Citizens Property Insurance Corp., set up by Florida legislators in 2002 as the insurer of last resort, is now the state's biggest property insurer. It has raised premiums by as much as

CREDIT: A. J. SISCO/CORBIS

CREDIT: RAVEENDRAN/AFP/GETTY IMAGES

Faced with political opposition, Kolkata lags behind other Indian cities in tackling auto emissions

Choking on Fumes, Kolkata Faces a Noxious Future

ON A TYPICAL WEEKDAY IN KOLKATA, A CITY OF 14 MILLION PEOPLE IN the Indian state of West Bengal, the streets are clogged with vehicles and the air is thick with exhaust. Soot coats the leaves of trees along the sidewalks. The skin tingles. Snot turns black. Every few weeks, when a strike shuts down businesses and traffic for an entire day, citizens see a silver lining: They can walk through downtown without handkerchiefs pressed to their noses.

Officials admit that air pollution in India's former colonial capital poses a serious health risk to its inhabitants. The annual average concentration of nitrogen oxides in many residential areas, for example, is nearly double the regulatory threshold of 60 micrograms per cubic meter. "The winter months are very serious," says Dipak Chakraborty, chief scientist of the West Bengal Pollution Control Board (WBPCB). "Elderly people with respiratory illnesses are severely affected."

Unlike other Indian metropolises such as New Delhi and Mumbai, where court-mandated measures to control auto emissions sparked recent improvements in air quality, Kolkata's air quality appears to be deteriorating. Environmental groups accuse the West Bengal government of doing little to address the problem. "Pollution is simply not an issue for them," says Sunita Narain of the nonprofit Centre for Science and Environment in New Delhi. State officials counter that they have forced many factories out of the city and required them to switch from coal to oil or natural gas. But they admit that a lack of resources and resistance from powerful taxi and bus unions—which wield considerable influence over the state's communist government—have prevented them from tackling auto emissions.

In 2005, WBPCB recommended that Kolkata's taxis, three-wheelers, and buses convert to cleaner fuels such as liquid petroleum gas and compressed natural gas, as New Delhi and Mumbai began doing 5 years ago. But vehicle operators argued that they could not afford to make the switch and threatened to strike. The proposal languished. "The government does not wish to become unpopular with the transportation lobby," acknowledges an official.

The board also recommended that, as Delhi and Mumbai have done, Kolkata begin phasing out all commercial vehicles made before 1990, when emissions standards were practically nonexistent. The government issued an order to that effect in May 2005, but bus and taxi unions challenged it in Calcutta High Court. In December 2007, when state lawyers failed to appear in court, judges voided the order. "The ban was a farce," says Anjali Srivastava, a chemist with the National Environmental Engineering Research Institute (NEERI) in Kolkata. "It was clear that the state government had no intention of following through."

Another obstacle for Kolkata and other Indian cities is an allegedly corrupt system for emissions checks. All vehicles are required to have an annual inspection at a private testing station. The certificates issued by these stations are often unreliable, says Chakraborty. "There is a lot of manipulation, and certificates are even issued without an inspection," he says, adding that government proposals to require testing stations to use tamper-proof technology have so far failed.

India's auto boom, fueled by a rising economy, is likely to aggravate Kolkata's woeful air quality and reverse some gains in Delhi and Mumbai, environmental groups predict. With manufacturing giant Tata set to debut a $2500 car to the Indian market by late 2008, tens of millions of automobiles are expected to be added to the country's congested streets in the next few years. Even though new vehicles emit far less pollutants, their sheer number could add up to high levels of air pollution, says Narain. An added concern, she says, is that a third of all new cars run on diesel, a more polluting fuel than gasoline. "Manufacturers want to sell diesel cars because diesel is cheaper," explains Narain, who wants the government to raise taxes on new vehicles and on diesel and invest more heavily in mass transit systems.

Rakesh Kumar, a scientist at NEERI in Mumbai, is opposed to any move that would put vehicles out of reach of the middle class. "Why should cars only be owned by rich people?" he asks. Instead, he wants cities to improve bus and train services to enable people to curtail their use of personal vehicles. This is already the case in Mumbai, Kumar says, which has the best public transportation of any Indian city.

Kolkata's priority should be to run public vehicles on cleaner fuels, Chakraborty says. Toward that end, the state government plans to spend $1 million this year to subsidize the cost of switching several hundred three-wheelers and taxis to liquid petroleum gas. That will not be enough, he concedes: Eventually, the government will have to phase out older vehicles and stem the numbers of new vehicles. "Unless we take drastic steps," Chakraborty says, "the problem will continue to get worse." **–YUDHIJIT BHATTACHARJEE**

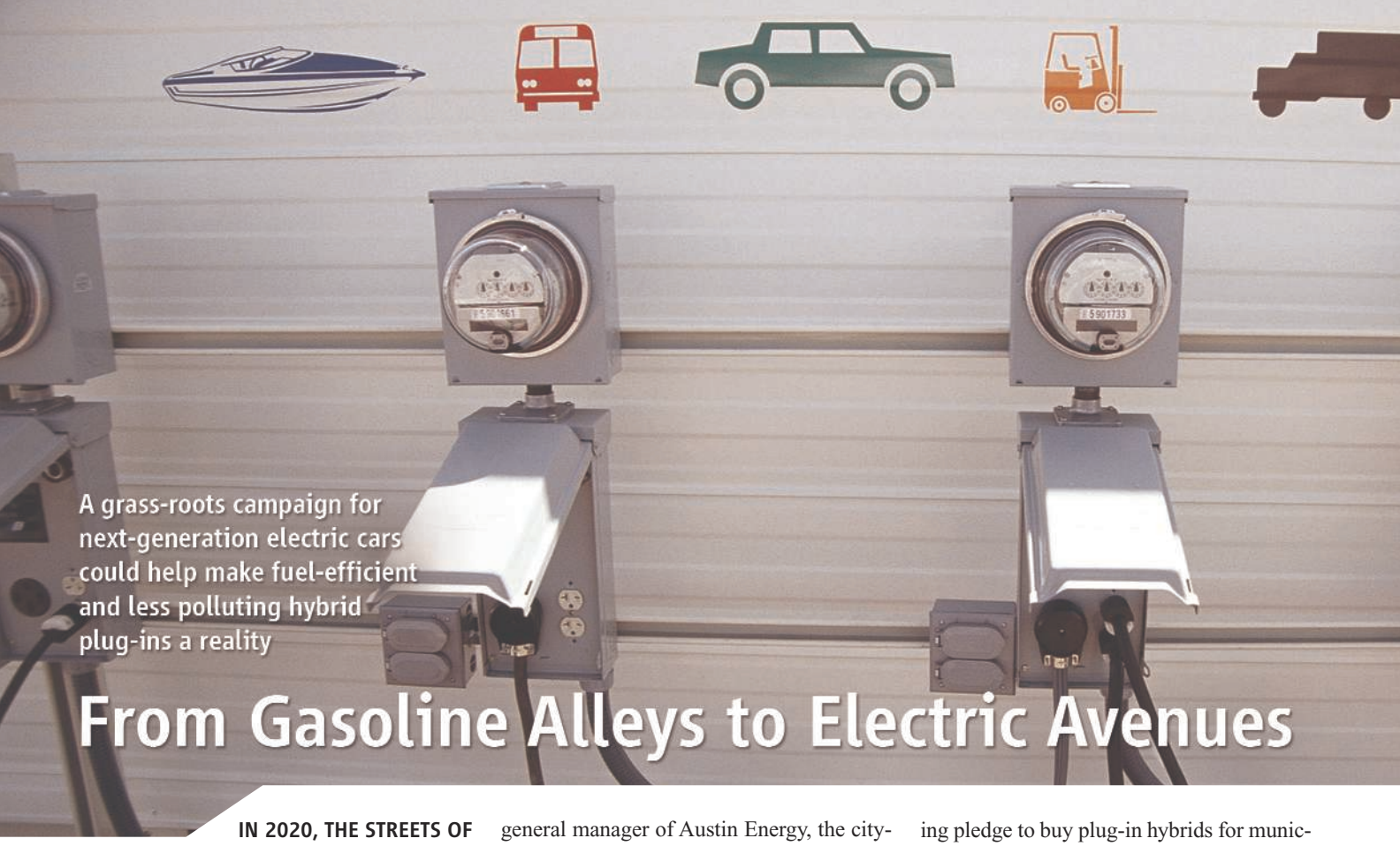
150% in the last 2 years. Rising premiums may price some residents out of hurricane zones, says French.

On a few occasions, the government has said enough is enough. Consider Brownwood, Texas. The tony Gulf Coast subdivision had been subsiding for years and was highly susceptible to flooding when Hurricane Alicia hit in 1983. "We didn't have very many homes standing," recalls A. Jean Shephard, who had lived in Brownwood since 1954. The Federal Emergency Management Agency offered to buy out the 400 or so wrecked houses. "I can't tell you how it felt" to leave, says Shephard, who now lives in Paris, Texas.

Brownwood lay abandoned for years. In the early 1990s, a special planning committee chose to turn the ghost town into a nature preserve with wetlands and a butterfly garden that now occupy more than 160 hectares.

Brownwood is a stark exception; New Orleans is the norm. Reconstruction there has proceeded fitfully. The most vulnerable areas are a patchwork of rebuilt homes and decaying shells—in Brown's words, a "shantytown of semiabandonment," and exactly the reincarnation he'd hoped to avoid. **–JENNIFER COUZIN**

Published by AAAS



A grass-roots campaign for next-generation electric cars could help make fuel-efficient and less polluting hybrid plug-ins a reality

From Gasoline Alleys to Electric Avenues

IN 2020, THE STREETS OF Austin, Texas, won't seem much different from the way they are today. Sidewalks and bicycles will be used much as they are now: seldom. The automobile will continue to be central to daily life. Most of the city's residents will commute to work, take children to Little League games, and pick up groceries on the way home by car. But there will be one big difference: Drivers will stop far less frequently at gas stations, and the air will be considerably cleaner. That, at least, is the way that Roger Duncan, deputy general manager of Austin Energy, the city-owned utility, sees it.

"We have a vision," says Duncan. He's spearheading a nationwide grass-roots effort to popularize the next-generation hybrid vehicle. Like today's hybrids, such as Toyota's popular Prius, they have dual gasoline and electric engines. But whereas current hybrids recharge their battery packs only during driving, plug-ins can also be recharged from the electrical grid by plugging into wall sockets. Last year, Duncan persuaded dozens of cities to sign a nonbinding pledge to buy plug-in hybrids for municipal fleets when they come on the market, as early as 2010.

Plug-ins "are most valuable in an urban environment," says Duncan, who notes that electric engines don't waste energy by idling at stoplights or in traffic jams. Less gasoline burned means less pollution: A 2006 study by the Electric Power Research Institute (EPRI) and the Natural Resources Defense Council predicts "small but significant improvements in ambient air quality" if half of U.S. drivers adopt plug-ins by 2050. Even

UNCLOGGING URBAN ARTERIES

LONDON—In 2003, soon after London authorities slapped a tax on each vehicle entering the city center, traffic volume fell 15%, and drivers spent 30% less time in gridlock, according to the city's Transport for London. Commuters were delighted, and once-virulent opposition to the fee, now £8 ($16) a day, subsided.

Congestion charges are a big hit in London and Stockholm, which adopted a similar tax in 2007. Other cities are expected to follow suit. New York City plans to implement a charge this year, and politicians in Shanghai, China, and Sydney, Australia, are debating the idea. "As congestion becomes worse in other major cities, it becomes more likely that charges will be put in place," says Graeme Craig, Transport for London's director of congestion charging.

Congestion is a bane of urban life. The average U.S. commuter spends 38 hours stuck in traffic, according to a study released last September by the Texas Transportation Institute at Texas A&M University in College Station. A century ago, economists suggested that road taxes would alleviate this ill. Since then, proposed schemes have included highway tolls, daily parking charges for cordoned areas, and graduated pricing based on time of day and kilometers traveled. Officials in Singapore first implemented a congestion charge in 1975; interest in such charges picked up considerably after London used the fee to untangle its downtown grid.

The tax is not a panacea. "You can't introduce it by itself and expect it to solve the problem," says Philip Blythe, a civil engineer at Newcastle University in the U.K. Charging schemes require detailed modeling of a city's layout, travel patterns, and public transportation, as well as occasional tweaks to maintain benefits. Still, some critics question whether the benefits will last. Transit time on London streets has slowed since the initial improvements; officials attribute this to construction projects.

CREDIT: GETTY IMAGES

Published by AAAS

if only a fifth of U.S. drivers buy plug-ins, the study found, that would lead to much higher energy efficiency, averting the emission of hundreds of millions of tons of carbon dioxide each year.

Current plug-in prototypes are more than twice as efficient in gasoline consumption as today's gas hybrids, says James Francfort of Idaho National Laboratory in Idaho Falls. No wonder, then, that many experts are itching to see the cars on the road. Plug-in hybrids, says California-based electric car advocate Chelsea Sexton, are "the killer app" of car technology.

Duncan became acquainted with hybrid plug-ins after Austin's city council in 2004 asked him to brainstorm green-energy solutions for the progressive university town. The council latched onto the technology the next year, authorizing $1 million for rebates for future plug-in hybrid purchasers. Duncan persuaded dozens of Austin government offices, businesses, and nonprofits to commit to soft orders: nonbinding pledges to car companies to buy plug-ins once available. Then in 2006, Austin created the Plug-In Partners national campaign, which has brought together 77 cities—from Alameda, California, to Wenatchee, Washington—and more than 100 organizations and businesses to set up similar programs and pledges. "Carmakers aren't going to make plug-in hybrids just for Austin," says Duncan.

Last month, Toyota announced that it will build a plug-in for sale by 2010, matching a General Motors (GM) target. Meanwhile, DaimlerChrysler is testing a prototype plug-in van, the Dodge Sprinter, which it hopes will compete with models that Japanese companies are developing. It's unclear whether Duncan's grass-roots campaign will gel into a potential market for plug-in hybrids, cautions GM's Larry Nitz. Still, says EPRI's Mark Duvall, "support for the technology has really grown in the last year." In Washington, D.C., the Plug-In Partners campaign has teamed with green groups and defense-minded organizations concerned about U.S. dependence on foreign oil to push for federal research and tax incentives.

There are several technical wildcards, such as how the larger battery packs—four times larger than those of the Prius—will withstand the rigors of city driving, how many recharging cycles they'll endure, and what changes to the U.S. power grid will be needed to take on a heavy charging load. "Our [industry's] research effort is really just starting out," says Francfort, whose team studies batteries, fuel use, and car performance.

Austin is chipping in some research as well. Its municipal fleet includes two Priuses equipped with extra batteries that can be recharged using a standard plug. Duncan says the city is tracking mileage performance, battery life, recharging capacity, and durability. And Internet giant Google is reviewing proposals for $10 million in technology grants it hopes to hand out this year. One priority is creating a computer protocol for cars to communicate their charging needs to an upgraded power grid able to manage a city full of plug-ins.

In Duncan's vision of Austin, plug-ins are a two-way street. During the daytime, commuters would leave them plugged in and allow the city grid to draw electricity from the cars during afternoon peak demands. "We'd buy [electricity] back," says Duncan, "instead of having to build another power plant." That could help keep Austin the way its residents like it: pretty much as it is today. **–ELI KINTISCH**

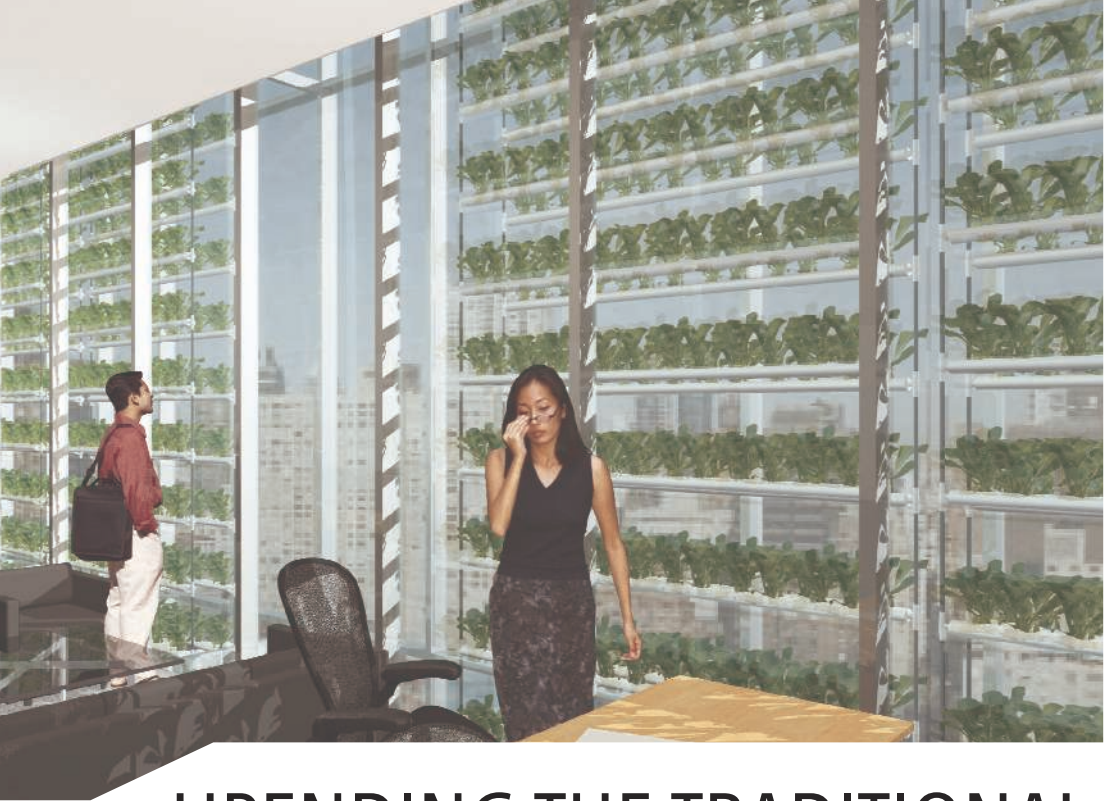
As other cities mull a congestion tax, Blythe emphasizes that there is no one-size-fits-all approach. For example, freeway traffic tends to be a bigger problem than downtown traffic in Los Angeles and other cities in the western United States, says David Brownstone, a transport economist at the University of California, Irvine. Some cities may not be ready for congestion charges. In Beijing, "we need to construct more roads first, and we need to provide better public transportation," says Yang Xinmiao, an engineer at Tsinghua University in Beijing.

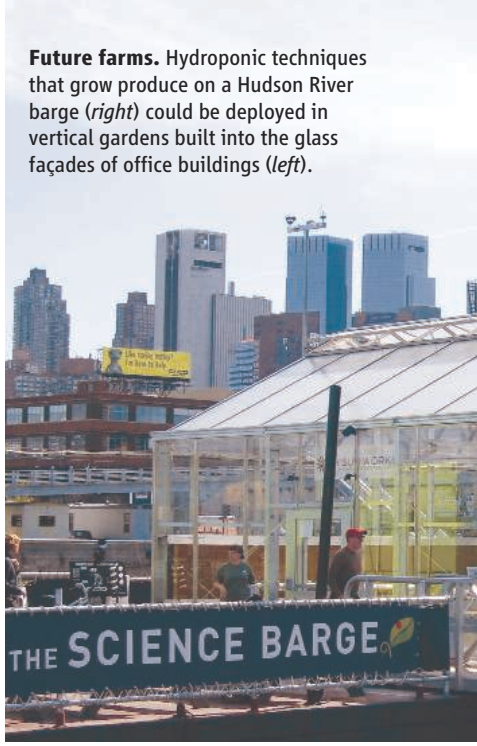
Hoping to build on its success, officials in London will use its congestion charge scheme to turn the screws on less eco-friendly cars. Later this year, authorities plan to raise the charge for high-polluting vehicles—those that emit more than 225 grams of carbon dioxide per kilometer—to £25. Blythe sees this trend eventually leading to a carbon-trading scheme for transport. "At some stage, we may have our own personal carbon allowance for travel," he predicts. "But first we have to get the basics right."

–ELIZABETH QUILL

CREDITS (TOP TO BOTTOM): AUSTIN ENERGY; MARTYN GODDARD/CORBIS

Published by AAAS





Future farms. Hydroponic techniques that grow produce on a Hudson River barge (*right*) could be deployed in vertical gardens built into the glass façades of office buildings (*left*).

UPENDING THE TRADITIONAL FARM

Cities are taking over farmland. Could they someday take over the job of farming, too?

IN A HIGH-TECH ANSWER TO THE "LOCAL food" movement, some experts want to transport the whole farm—shoots, roots, and all—to the city. They predict that future cities could grow most of their food inside city limits, in ultraefficient greenhouses.

"Vertical farms," proponents say, could produce more food using a fraction of the resources that traditional farms consume. The lives of millions of people may depend on it. Dickson Despommier, a parasitologist at Columbia University and an avid proponent of vertical farming, calculates that with projected population increases, the world will need 1 billion more hectares of arable land by 2050—roughly the area of Brazil and far more land than will be available.

Researchers are now putting prototypes of intensive urban farms to a real-world test. The basic concept is an evolution, not a revolution, of greenhouse technology. Greenhouses "can grow any crop anywhere at any time—at a cost," says Gene Giacomelli of the Controlled Environment Agriculture Center at the University of Arizona in Tucson, which has built a $450,000 greenhouse for researchers who winter at the South Pole. Well-designed greenhouses use as little as 10% of the water and 5% of the area required by farm fields, says Theodore "Ted" Caplow, executive director of the engineering company New York Sun Works in New York City, which designs energy-efficient urban greenhouses. "We are removing that footprint from the countryside," he says, and reducing pressure on habitats and depleted soils.

Urban indoor farms can't do it all. Growing grains such as wheat, corn, and rice indoors does not save as many resources as growing vegetables and fruits indoors, says Caplow, and most trees grow too slowly to make greenhouse orchards pay off. Some of the more ambitious concepts for vertical farms will require technological breakthroughs in lighting and energy consumption. And initially, at least, urban produce will likely be more expensive than that grown at conventional farms and shipped to a city.

But as oil prices rise, greenhouse economics look more favorable, Giacomelli says. "All our cheap food is based on cheap transportation, cheap water, and cheap energy for nitrogen-based fertilizer," he says.

One approach that could be implemented quickly is rooftop greenhouses. In a demonstration of what can be grown on New York City's roofs, Caplow's company last summer built and operated the Science Barge, a floating greenhouse on the Hudson River that used solar power and recycled water to grow fruits and vegetables. New Yorkers eat 100 kilograms of fresh vegetables on average per year, Caplow says, and the rooftops of New York City would provide roughly twice the needed space to supply the entire city. New York Sun Works is now installing a demonstration greenhouse on top of a New York City school that would serve as a teaching area and supply produce to its cafeteria.

A more ambitious concept is farming the facades of office buildings. Double-glass facades are already popular among architects as an energy-saver, allowing winter sun in while insulating against noise and heat loss. In the summer, most double facades have built-in shades to keep the interior cool. Hydroponic gardens could provide that shade, Caplow says. Vertical conveyor belts could cycle plants to the lower floors in time for harvest. "The systems we are designing are what we can actually do today," Caplow says.

Gazing further into the future, Despommier and his students are refining the idea of skyscraper farms. They estimate that a 30-story farm on one city block could feed 50,000 people with vegetables, fruit, eggs, and meat. Upper floors would grow hydroponic crops; lower floors would house chickens and fish that consume plant waste. Heat and lighting would be powered by geothermal, tidal, solar, or other renewable energy sources. Nitrogen and other nutrients would be sieved from animal waste and perhaps from the city sewage system. "That's where a significant fraction of your fruits and vegetables are going," into sewage, Despommier says. "You have to close the loop." Eventually, he says, hydroponic greenhouses could

CREDITS (LEFT TO RIGHT): KISS AND CATHCART, ARCHITECTS; NEW YORK SUN WORKS AND ARUP; WWW.NYSUNWORKS.ORG

Published by AAAS

CREDIT: ANDREAS RENTZ/GETTY IMAGES

also be a boon for the developing world. In tropical regions, they could make use of ample sun, conserve water, and give worn-out soils a rest. Ideally, they would also provide a way to safely turn human waste into plant food, he says.

Such ideas are inspiring, says Jan Broeze, an agricultural scientist at the University of Wageningen, the Netherlands. But "you need large technological breakthroughs" in lighting and waste processing to realize them. In 2001, Broeze, Peter Smeets, and their colleagues proposed a six-story urban farm called Deltapark at Rotterdam harbor that would recycle water and nutrients and use excess heat from nearby buildings. The agricultural ministry supported Deltapark, but the project was abandoned after the press criticized it for being "too industrialized." Now Broeze is working on several projects that link greenhouses with livestock producers to recycle waste and reduce energy consumption. And he and other Dutch scientists are working with colleagues in India and China to design urban farms in several cities. The biggest project is part of Dongtan Eco-city, near Shanghai (see p. 740).

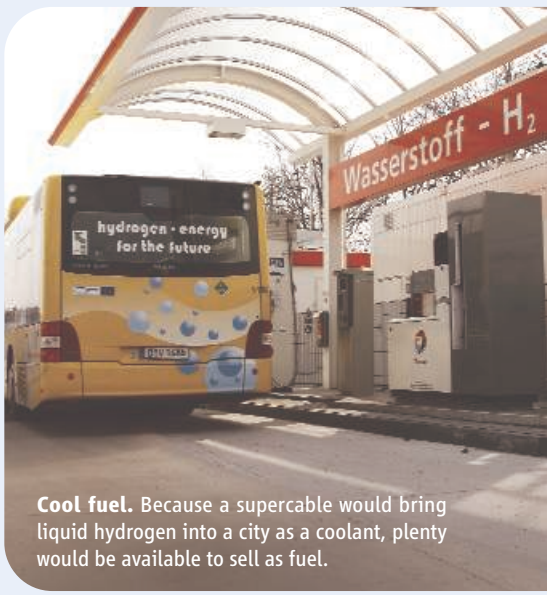
One goal of Dongtan is to grow enough food to replace lost productivity as farmland is urbanized, says Peter Head, director of Arup, a design company leading the project. "The big question is whether it is economically viable," he says. Head predicts that the lessons learned in China will propel a fundamental shift in the world's approach to agriculture. "It isn't a matter of whether we think it would be nice to do urban farming or not," he says. "It's a matter of whether we are going to survive."

–GRETCHEN VOGEL

Imagining a City Where (Electrical) Resistance Is Futile

IF YOU WERE BUILDING A CITY IN THE 21ST CENTURY, HOW WOULD YOU POWER AND FUEL IT? Pipe in electricity by copper cable and haul in gasoline by road, as is done today? Paul Grant thinks not. He envisions using superconducting electrical cables and liquid hydrogen to energize a metropolis while emitting little or no carbon.

Grant's "SuperCity" is popular with proponents of nuclear energy because it relies heavily on the next generation of nuclear plants. But the hard work of fleshing out the concept and developing the cables has only just begun. "The basic research still needs to be done. Can this thing really be built?" asks electrical engineer Thomas Overbye of the University of Illinois, Urbana-Champaign, who has organized a couple of workshops on the related idea of a SuperGrid to wire a whole continent. Not everyone is convinced. "We should sort out the problems with the energy system we have" before inventing new ones, says Robert Socolow of the Energy Group at Princeton Environmental Institute. Grant concedes that at the moment, SuperCity remains a utopian vision. "It's about the energy society we should be looking at 50 years from now," he says.



Cool fuel. Because a supercable would bring liquid hydrogen into a city as a coolant, plenty would be available to sell as fuel.

While working at IBM's Almaden Research Center in northern California during the 1980s, Grant pioneered the development of high-temperature superconductors, complex oxides that carry electricity with zero resistance at temperatures that are low but much higher than those for earlier materials. Unlike their predecessors, high-temperature superconductors can be cooled using easily obtained liquid nitrogen. In 1993, after more than 40 years with IBM, Grant moved down the freeway to the Electric Power Research Institute (EPRI) in Palo Alto, where he applies his skills to solving energy problems, such as improving transmission. In 1999, Grant was challenged to come up with a "wild idea" to present at a U.S. Department of Energy (DOE) meeting on superconductivity. His brainstorm was to combine superconducting cables with a hydrogen economy.

Today's electricity grids lose about 7% of power to resistance, so superconducting cables would boost efficiency—if the lines could be supercooled. Grant proposed pumping liquid hydrogen into a pipe in which the superconducting cable runs down the middle, creating a "supercable" that carries both electricity and hydrogen into a city. Fuel stations could tap into the coolant to power electric fuel-cell cars. Hydrogen could also be burned for domestic heating and cooking, circulated au naturel for air conditioning, or converted into electricity during peaks in demand.

The SuperCity requires nuclear power to generate electricity, Grant argues, because renewable energy sources such as biofuels and wind or solar farms take up valuable land and degrade the environment. "You cannot beat nuclear for its [small] footprint and [high] power density," he says. Upcoming nuclear reactor designs, known as generation IV, will operate at high temperatures and produce hydrogen as a byproduct of electricity generation. Grant's SuperCity doesn't dismiss other energy sources entirely, however. Every roof could be covered with solar cells, and waste could be burned to generate power. Such alternative sources could supply about 10% of a city's energy demand, he says.

Because of the huge investment sunk into conventional technology, it may take decades to realize the entire SuperCity vision. Supercables integrated into the electricity grid could happen sooner. "It would have immediate appeal," says Steven Eckroad of EPRI, which has carried out cable-design studies. Overbye adds: "The concept really needs a big funder [such as DOE] to step up and say, 'This is the way we're going to go.' But that hasn't occurred yet."

Grant, now retired from EPRI but still consulting, is undaunted. "If we were starting from scratch on another planet, this would be the way to do it," he says.

–DANIEL CLERY

Published by AAAS

Money—With Strings—to Fight Poverty

Cash incentives for education and health care are a new tool for helping the urban poor in Mexico and New York

IN 1995, THE PESO'S DEVALUATION plunged Mexico into its worst economic crisis in half a century. For years, the government had helped those in poverty by subsidizing the price of staples such as tortillas and milk. But that meant that the rich and middle class who didn't need help got it anyway, wasting scarce government resources. Mexico needed to target the most destitute.

So Deputy Finance Minister Santiago Levy and others proposed a radical idea: Scrap the food subsidies and use census data and surveys to identify the poorest households and give them cash instead. Not handouts: To get the money, families would have to keep their children in school and get regular health checkups. Payments would go only to mothers, because research showed that they were less likely than fathers to squander it. Children would get nutritional supplements. And the program would include treatment and control groups, just like a clinical trial. The long-term goal was to pull the next generation out of poverty.

Other Mexican officials thought it was "crazy" at first, Levy says—like "bribing" parents. But 10 years later, the "conditional cash transfer" (CCT) program is credited with higher school enrollment and taller, healthier children. Outside researchers regard Mexico's CCT program as an innovative approach to reducing Third World poverty. As a way to boost education level, which influences income, CCTs are "one of the most effective programs," says economist Alan Krueger of Princeton University. The giant social experiment has been adopted widely in cities including Rio de Janeiro and Istanbul and has just debuted in New York City.

Levy, an economics professor at Boston University who returned to Mexico in 1992, says he and others set up the CCT program in recognition of the "strong coupling" between health, nutrition, education, and lifetime earnings. Benefits range from $11 per month for children in 3rd grade up to $69 for a girl finishing high school, with attendance verified by school records. Mothers also get cash for buying food and nutritional supplements for babies when they attend family health checkups at free clinics.

The program, called Progresa, built in outside evaluation because "credibility was essential," says Levy. The International Food Policy Research Institute (IFPRI) in Washington, D.C., and Mexican academics set up a study design with 320 villages randomly assigned as treatment groups and 186 control villages.

After 18 months, the comparisons showed clear gains. For instance, the program raised junior high school enrollment 20% for girls and 10% for boys; children under age 6 had 12% fewer bouts of illness; early prenatal health visits rose 8%; and toddlers grew 1 centimeter more per year. The results, published in the *Journal of the American Medical Association* in 2004 and in economics journals, persuaded the next government in 2001 to rename the program Oportunidades and extend it to cities. It now covers 5 million families, or one-quarter of Mexico's population.

In urban areas, the health benefits have been similar, but school enrollment hasn't climbed as much, perhaps because of the greater job opportunities for youth compared with rural areas, notes economist Susan Parker of Spectron Desarrollo, a research organization in Mexico City. One disappointment in both rural and urban areas is that test scores haven't risen. "The deeper issue is the quality of education and teachers," says Levy, now at the Inter-American Development Bank.

Sound investment. Mexico pays poor families to keep their children in school.

Selected Results for Oportunidades in Mexican Cities

- ↑ **13%** 7-year-old boys and ↑ **8%** 7-year-old girls enrolling in school
- ↓ **24%** 16- to 19-year-old boys who drop out of school
- ↑ **52%** preventive health visits for children under 6 years old
- ↓ **6 days** sick days/year per family
- ↑ **1 cm** height for children aged 6 to 24 months in 2002
- ↑ **0.57 kg** weight for infants under 6 months in 2002

Note: Enrolled families were compared with matched families who were eligible but did not sign up or did not live in a participating area.

PHOTO AND SOURCE: OPORTUNIDADES

Published by AAAS

About 20 other countries in Latin America and elsewhere have adopted CCT programs, says economist Laura Rawlings of the World Bank. In urban Turkey, for example, the program has raised the attendance of girls in secondary school by 6%, says Michelle Adato of IFPRI. In the World Bank's view, CCTs are "not a silver bullet" for eliminating poverty, says Rawlings. But they are, she says, "a promising tool."

Perhaps the most surprising imitator is New York City. This fall, after visiting Mexico City and the nearby city of Toluca, Mayor Michael Bloomberg launched Opportunity NYC, a privately financed program that will pay families for achieving certain schooling, health, and work targets. About 5000 families (including controls) identified from lists of children receiving free school lunches have enrolled in a 2-year pilot study. Families can get $50 per month for a child who has a 95% attendance rate at high school and as much as $200 for a preventive health care visit; one of the biggest awards, $600, is for passing a high school achievement test.

The program has drawn plenty of flak. The notion of paying for better test scores has raised concerns about the kind of rote learning that may encourage. A more basic question, notes Krueger, is whether the payments are big enough to motivate the New Yorkers, who although poor are well-off compared with the average Oportunidades family subsisting on $100 a month. "It's right to be skeptical," says sociologist James Riccio of MDRC, a nonprofit research organization in New York City that is evaluating the program. Still, he says, like Mexico's CCT program, "what's most impressive is how rigorously it's going to be tested. In the end, it will be informative."

As New York City samples these waters, Mexican researchers are now investigating the program's long-term effects. Last year, they began collecting data on the education level, jobs, and marital status of the first generation of boys to benefit from the program. But as Levy points out, improved health and more schooling may not be sufficient to ensure that Oportunidades children fare better than their parents, unless accompanied by economic growth and better schools. "My concern is that people will begin to think this is a panacea," says Levy. Still, even if Mexico's CCT program alone cannot wipe out urban poverty, it is attacking some of its causes at their roots.

–JOCELYN KAISER

Replacement plan. Converting dirt floors like this one to concrete yields health benefits.

CREDIT: JACK FIELDS/CORBIS

BUILDING ON A FIRM FOUNDATION

IMPROVING THE HOMES OF SLUM DWELLERS IS ONE OF THE OLDEST STRATEGIES FOR tackling urban poverty in the developing world. But the health benefits of these approaches have rarely been put to a controlled test. In one such effort, researchers found striking improvements in child health from a simple measure: giving poor Mexican families a concrete floor.

Eight years ago, the government of the state of Coahuila in northern Mexico began offering households with dirt floors $150 worth of free concrete—enough to cover most rooms in a house. Trucks delivered the wet concrete, and families spread it after it was poured. To assess the benefits of the program, called Piso Firme (Firm Floor), state officials approached economist Paul Gertler of the University of California, Berkeley, who was part of a team evaluating Mexico's new poverty payments program (see main text).

Gertler and colleagues compared households in the twin cities of Torreón, Coahuila, and Gómez Palacio/Lerdo in the state of Durango, which has similar socioeconomics but did not have a Piso Firme program. Two to 4 years after the Torreón families got their concrete floors, Gertler's team interviewed 2755 mothers in both cities about their family's health, took stool and blood samples from children less than 6 years old, and gave the children vocabulary tests.

"It turned out to be much more interesting than I expected," Gertler says. His team calculated that in homes converted from all-dirt floors to all-concrete floors, children had 78% fewer parasitic infestations (one in three children in the control group had parasites compared with 7% in the treatment group). The children had also had half as many diarrhea episodes in the past month, an 81% drop in anemia, and a 36% or better improvement on cognitive tests. Mothers also reported less depression and more satisfaction with their lives. "The benefits are incredibly impressive," says epidemiologist Nancy Padian of the University of California, San Francisco, who has studied economic interventions to improve women's reproductive health in Africa and recently began collaborating with Gertler.

Gertler cautions that in rural areas lacking clean water, the effects might not be so dramatic. Still, the program's success spurred Mexico to adopt a national Piso Firme program. Although it's "sort of a no-brainer" that concrete floors are better than dirt floors, says Padian, the study demonstrates the value of low-tech ways to ease urban poverty. "We need to capitalize more on the no-brainer solutions," she says.

–J.K.

Published by AAAS

REVIEW

Global Change and the Ecology of Cities

Nancy B. Grimm,[1*] Stanley H. Faeth,[1] Nancy E. Golubiewski,[2] Charles L. Redman,[3] Jianguo Wu,[1,3] Xuemei Bai,[4] John M. Briggs[1]

Urban areas are hot spots that drive environmental change at multiple scales. Material demands of production and human consumption alter land use and cover, biodiversity, and hydrosystems locally to regionally, and urban waste discharge affects local to global biogeochemical cycles and climate. For urbanites, however, global environmental changes are swamped by dramatic changes in the local environment. Urban ecology integrates natural and social sciences to study these radically altered local environments and their regional and global effects. Cities themselves present both the problems and solutions to sustainability challenges of an increasingly urbanized world.

Humanity today is experiencing a dramatic shift to urban living. Whereas in 1900 a mere 10% of the global population were urban dwellers, that percentage now exceeds 50% and will rise even more in the next 50 years (Fig. 1). More than 95% of the net increase in the global population will be in cities of the developing world, which will approach the 80% urbanization level of most industrialized nations today (*1*). In addition, individual cities are growing to unprecedented sizes, with nearly all of these new megacities (>10 million, by convention) in the developing world (Fig. 1). Economic growth and demographic changes will accompany growth in urban populations, especially in populous China and India, producing ever-greater demands on services that nearby and distant ecosystems provide.

Ecologists shunned urban areas for most of the 20th century, with the result that ecological knowledge contributed little to solving urban environmental problems. Recently, however, increasing numbers of ecologists have collaborated with other scientists, planners, and engineers to understand and even shape these ascendant ecosystems. With the advent 10 years ago of National Science Foundation–funded urban research programs in the United States, which built upon but differed from earlier efforts (see references in section 1 of the supporting online material), urban ecology also has begun to change the discipline of ecology. Urban ecology integrates the theory and methods of both natural and social sciences to study the patterns and processes of urban ecosystems. Evolving conceptual frameworks for urban ecology view cities as heterogeneous, dynamic landscapes and as complex, adaptive, socioecological systems, in which the delivery of ecosystem services links society and ecosystems at multiple scales (*2*–*5*).

Urban ecologists seek commonalities among city ecosystems, an understanding of how context shapes the socioecological interactions within them, and their role as both drivers and responders to environmental change. Here, we focus on five major types of global environmental change that affect and are affected by urban ecosystems (Fig. 2): changes in land use and cover, biogeochemical cycles, climate, hydrosystems, and biodiversity. We argue that cities themselves represent microcosms of the kinds of changes that are happening globally, making them informative test cases for understanding socioecological system dynamics and responses to change.

Land-Use and Land-Cover Change Accompanying Urbanization

The unprecedented rates of urban population growth over the past century have occurred on <3% of the global terrestrial surface, yet the impact has been global, with 78% of carbon emissions, 60% of residential water use, and 76% of wood used for industrial purposes attributed to cities (*6*). Land change to build cities and to support the demands of urban populations itself drives other types of environmental change (Fig. 2).

Urban dwellers depend on the productive and assimilative capacities of ecosystems well beyond their city boundaries—"ecological footprints" tens to hundreds of times the area occupied by a city—to produce the flows of energy, material goods, and nonmaterial services (including waste absorption) that sustain human well-being and quality of life (*7*, *8*). At the same time, large urban agglomerations are fonts of human ingenuity and may require fewer resources on a per capita basis than smaller towns and cities or their rural counterparts (*9*) (see references in section 2 of the supporting online material; figs. S1 and S2 and table S1).

Even in ancient times, the excessive demands of a highly stratified urban elite led to degradation of productive landscapes and the collapse of otherwise successful societies (e.g., salinization in 3rd millennium BCE Mesopotamia) (*10*). Although exacerbated by recent globalization trends, centuries ago the demands of European consumers led to deforestation of colonial lands and more recently, demand for beef from countries of the Western Hemisphere has transformed New World tropical rainforests into grazing land.

It is also at the regional scale that land-use changes driven by and resulting from population movement are most apparent. Perceived opportunities in growing urban centers and lack of opportunities in rural settings, resulting from degraded landscapes and imbalanced economic systems, have made the migrations since the second half of the 20th century the greatest human-environmental experiment of all time (*11*). In China alone, 300 million more people likely will move to cities, transforming their home landscapes and continuing an already unbelievable juggernaut of urban construction (*12*). Shortages of construction materials such as metals, coal, cement, and timber are likely to constrain China's urbanization in the long term, however, and exert pressure on growth of infrastructure globally (*13*).

Urbanization leads to increased patch fragmentation and diversity (*14*), which may be expressed as more edges (i.e., interfaces between distinct land-cover types) or smaller patch sizes (e.g., urban, residential, and desert land-use patches averaged 20, 100, and 650 ha, respectively, in central Arizona) (*15*). Urban land use often leaves a legacy of impact in the ecological characteristics of a landscape. In the city of Phoenix, for example, formerly agrarian lands exhibit unique soil biogeochemical properties after 40 years (*16*), and other locations in the region still reveal agricultural legacies after centuries (*17*).

A much-debated urban-planning assumption holds that the form of cities follows the function of land-use patterns, leading to a diversity of land-use arrangements (*18*). However, a recent study of four Chinese cities found convergent urban form in shape, size, and growth rates despite varying economic and political drivers (*19*). Land-use policies (i.e., zoning, master plans, growth boundaries) help determine urban form and its impact, but a long-term study of the Seattle region found that growth-management efforts to increase housing densities within growth boundaries had the unintended consequence of encouraging low-density housing sprawl in rural and wildland areas just beyond those boundaries (*20*).

Urban ecology at the local scale centers on the relationships among urban design and construction, ecosystem services delivered in the new system, responses of people and their institutions to evolving opportunities, and actions that drive further change in the system (*2*, *3*, *5*). The "edge" of the city expands into surrounding rural landscape, inducing changes in soils, built structures, markets, and informal human settlements, all of which exert pressure on fringe ecosystems. These peri-urban environments are the glue that link

[1]School of Life Sciences, Arizona State University, Tempe, AZ 85287–4501, USA. [2]New Zealand Centre for Ecological Economics, Private Bag 11 052, Palmerston North, New Zealand. [3]School of Sustainability, Arizona State University, Tempe, AZ 85287–3211, USA. [4]CSIRO Sustainable Ecosystems, Canberra ACT 2601, Australia.

*To whom correspondence should be addressed. E-mail: nbgrimm@asu.edu

core cities in extended urbanized regions. Indeed, urban planner Robert Lang has suggested that cities are no longer independent but represent a limited number of dominant megapolitan regions across the globe—coalitions of urban centers and increasingly built-up intervening regions (*21*). The next frontier in urban ecology is to understand urbanization in the context of biophysical, economic, or political settings. Continental or global comparisons among cities might productively be based on this megapolitan concept.

Altered Biogeochemical Cycles in Cities and Their Regional-to-Global Effects

Urban areas are both responsible for, and respond to, changes in biogeochemical cycles (Fig. 2). The concentration of transportation and industry in urban centers means that cities are point sources of CO_2 and other greenhouse gases, which affect Earth's climate, as well as trace gases such as NO, NO_2, O_3, SO_2, HNO_3, and various organic acids (*22*, *23*). Regionally, air pollution in particular influences nutrient cycling and primary production in adjacent, exposed ecosystems. The disproportionate location of cities along rivers and coastlines makes these areas important contributors to eutrophication.

Wastes generated in cities and entering air and water transport affect biogeochemical cycles from local to global scales, with the extent of influence depending on the vectors by which materials are carried away from their source. For example, the 20 largest U.S. cities each year contribute more CO_2 to the global atmosphere than the total land area of the continental United States can absorb (*24*). The concept of urban metabolism analogizes a city to an organism that takes in food and other required resources and releases wastes to the environment (*8*, *25*). Scientists debate the appropriateness of the metabolism analogy (*25*), but its greatest utility has been in quantifying the longitudinal trends in consumption and waste generation of expanding cities (*26*). This and other studies show large increases over two decades in the throughput of materials such as the food-waste stream, import and solid-waste accumulation or decomposition of paper and plastics, and tremendous growth in demand for building materials. In Beijing, for example, total carbon emitted from solid-waste treatment increased by a factor of 2.8 from 1990 to 2003 (*27*).

Pollution generation by cities is of increasing concern when urbanization outpaces societal capacity to implement pollution-control measures. For example, in the United States, emissions controls somewhat counterbalance the increased driving distances resulting from urban sprawl (*28*); however, increased coal burning and automobile use accompanying economic expansion in some Chinese cities have had serious air-pollution consequences (*29*). Nutrient loads from rapidly urbanizing regions to rivers and coastal ecosystems in the developing world show large increases where sewage treatment is lacking or inadequate (*30*). However, although urbanization and economic expansion outpace environmental controls in the developing world, waste from the most affluent cities remains a primary driver of altered biogeochemical cycles globally.

Cities themselves show symptoms of the biogeochemical imbalances that they help to create at coarser scales. For example, cities experience high acid and N deposition and elevated atmospheric concentrations of CO_2, CH_4, and O_3, which can produce both growth-enhancing and growth-inhibiting effects on organisms (*31*). Elemental mass balances can frame this problem, because they identify potential excesses of inputs over outputs and likely sinks within the urban landscape (*8*, *22*, *32*). Cities are hot spots of accumulation of N, P, and metals (*8*, *33*) and, consequently, harbor a pool of material resources. Could high-nutrient, treated wastewater substitute for commercial N fertilizers to supply crops and lawns with nitrogen, for example? In Phoenix, using nitrate-rich groundwater to irrigate fields could reduce needed fertilizer by >100 kg/ha (*34*). A small (but growing) proportion of the copper extracted globally is recycled, yet increasing the reuse and recycling of copper and other metals would do much to stem the rapid rise in demand from sources increasingly difficult to extract (*33*). Such reuse also would alleviate problems of metal accumulation in soils (*35*).

Human management of urban landscapes is often highly heterogeneous within cities, depending on the financial resources to purchase plants, fertilizer, and even water, land cover (including impervious surfaces), and the relevant organiza-

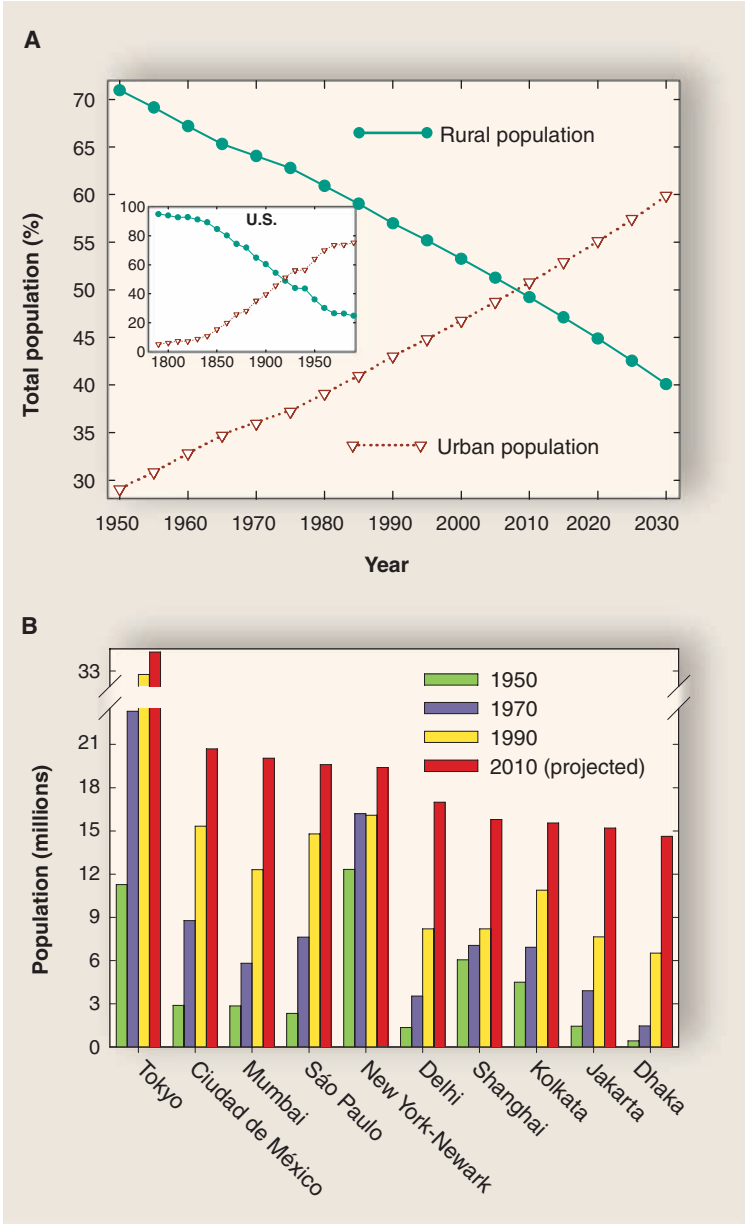


Fig. 1. (**A**) Change in world urban and rural population (%) from 1950 to 2030 (projected); plotted from data in (*1*). Inset shows comparable data for the United States from 1790 to 1990; plotted from data in (*73*). (**B**) Change in population of the 10 largest urban agglomerations from 1950 to 2010 (projected), ranked from left (largest) to right by their projected population size in 2010: Tokyo, Japan; Ciudad de México, Mexico; Mumbai, India; São Paulo, Brazil; New York–Newark, USA; Delhi, India; Shanghai, China; Kolkata, India; Jakarta, Indonesia; Dhaka, Bangladesh. Data are from (*1*).

tional level at which management is applied (e.g., household, neighborhood, city). For example, soil-nutrient concentrations across desert metropolitan regions can vary considerably because of legacy factors mentioned previously, as well as urban structure (impervious land cover) and landscape choices (lawns, tree cover, etc.) (*36*). Certain features of streams are more effective than others in retaining nutrients (*37*). For some atmospheric pollutants, localized variation in human behavior is less important than the collective, temporal behavior of the population—for example, in driving habits that produce daily or weekly cycles of particulate, CO_2, NO_x, or O_3 plumes (*38*).

Urbanization and Climate Change

Undoubtedly, urban centers, especially those in the developed world, are the primary source of greenhouse-gas emissions and thus are implicated in global climate change. Yet, the top-down influence of global climate change on cities may be overshadowed by local changes in climate that accompany urbanization (Fig. 2): increased minimum temperatures and sometimes reduced maxima, reduced or increased precipitation, and weekly cycles.

The best-documented example of anthropogenic climate modification is the urban heat island (UHI) effect: Cities tend to have higher air and surface temperatures than their rural surroundings (*39*), especially at night. Several characteristics of urban environments alter energy-budget parameters and can affect the formation of the UHI. These include land-cover pattern, city size (usually related to urban population size), increased impervious surfaces (low albedo, high heat capacity), reduced areas covered by vegetation and water (reduced heat loss due to evaporative cooling), increased surface areas for absorbing solar energy due to multistory buildings, and canyon-like heat-trapping morphology of high-rises. The UHI is a local phenomenon with negligible effect on global climate (*40*), but its magnitude and effects may represent harbingers of future climates, as already-observed temperature increases within cities exceed the predicted rise in global temperature for the next several decades. Kalnay and Cai (*41*) estimated that urbanization and other land-use changes accounted for half of the observed reduction in diurnal temperature range and an increase in mean air temperature of 0.27°C in the continental United States during the past century. By comparison, downtown temperatures for the United States have increased by 0.14° to 1.1°C per decade since the 1950s (*42*). Research on the effects of elevated temperature on remnant ecosystems (e.g., parks and open space) within cities, particularly when other variables are controlled [e.g., (*31*)], may contribute much to our ability to predict how ecosystems will respond to global climate change (*43*).

UHI affects not only local and regional climate, but also water resources, air quality, human health, and biodiversity and ecosystem functioning (*42*). Urban warming in hot climates exerts heat stress on organisms, including humans, and may influence water resources by changing the surface-energy balance, altering not only heat fluxes but also moisture fluxes near the surface. UHI may induce the formation of photochemical smog and create local air-circulation patterns that promote dispersion of pollutants away from the city. In warm regions (and summertime of cooler regions), urban warming greatly increases energy consumption for cooling. For example, about 3 to 8% of electricity demand in the United States was estimated to be used to compensate for UHI effects (*42*), representing another indirect feedback to global climate change. One way to mitigate the UHI effect is by increasing vegetation cover and albedo (*39*), but this strategy is a trade-off requiring greater water use, especially in arid regions.

Although local temperature changes may exert greater influence on urban ecosystems than global temperature increases at present, other aspects of regional and global climate change pose risks to cities. In particular, coastal cities would be exposed to rising sea level and any increased hurricane frequency caused by climate change. Thus, one important aspect of achieving urban sustainability is strengthening our ability to respond to the changing relation between urbanization and climate. For cities to effectively respond to global climate change, both mitigation and adaptation strategies—and economic markets for them—will be required.

Human Modifications of Hydrologic Systems

Throughout history, cities have sprung up along rivers and deltas, precisely because of the available water. Seldom are these waterways left unmodified. Within cities, water is intricately linked to not only domestic use but also industrial processes, adequate sanitation, and protection from natural disasters (floods, hurricanes, and tsunamis). Thus, humans have modified hydrosystems to meet a large array of oft-conflicting goals. Designed or altered streams, rivers, flood channels, canals and other hydrosystems serving urban areas neither replicate the aquatic ecosystems they replace nor preserve the ecosystem services lost (except for those, like flood convey-

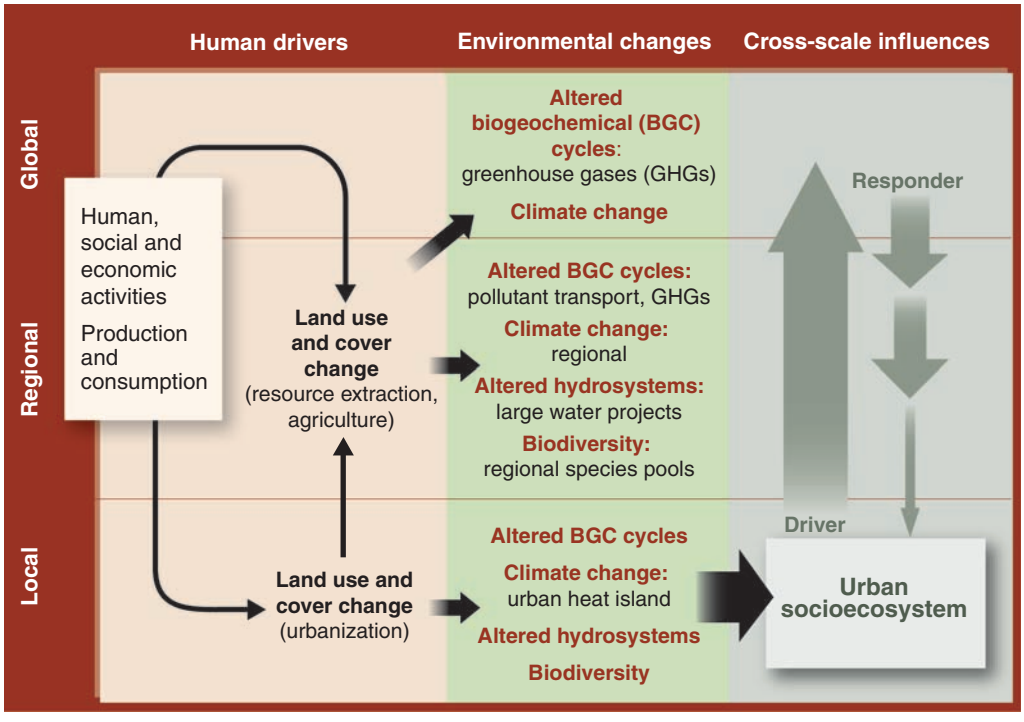


Fig. 2. Framework showing urban socioecosystem (lower right) as a driver of (upward arrows) and responder to (downward and horizontal arrows) environmental change. Land change to build cities and support their populations drives local to global alterations of biogeochemical cycles, climate, hydrosystems, and biodiversity. Large local environmental changes are greater than those that filter down from global environmental change (horizontal black arrow). Not all possible interactions and drivers are shown.

ance or water delivery, for which they are designed). Consequently, there are few model systems with which to compare these highly altered environments [e.g., (*44*)]. Some have called for restoration of streams in urban areas (*45*), while others advocate study and management of such designed ecosystems as unique ecosystems, with a view to optimizing services to urban populations (*46*). Among such services we would include flood protection, habitat for a diverse aquatic biota, nutrient retention, and a sense of place.

Among the most important modifications that affect streams in urban areas is increased impervious cover, which changes hydrology and funnels accumulated pollutants from buildings, roadways, and parking lots into streams. Point-source pollution has been dramatically reduced by regulation in the United States, but remains a serious issue in many developing countries (*47*). Industrial discharges, as well as sewage, contaminate rivers and lakes. Stormwater infrastructure systems in newer cities are separate from wastewater discharges, but the two streams are mixed in older European and American cities, creating acute pollution events in recipient systems. Both storms and low flow-discharge from cities contribute to localized or even regional pollution downstream, especially from pesticides and persistent organic pollutants.

The changes in chemical environment, exposure to pollutants, simplified geomorphic structure, and altered hydrographs of urban streams combine to create an urban stream "syndrome" of low biotic diversity, high nutrient concentrations, reduced nutrient retention efficiency, and often elevated primary production (*48*, *49*). Other ecosystem functional attributes respond less consistently to urbanization, perhaps because the extent and form of hydrologic alteration vary tremendously among urban areas. Countering the urban stream syndrome may require abandonment of the ideal of a "restored" stream in favor of a designed ecosystem. Successful, ecologically based designs of novel urban aquatic ecosystems are becoming more common and exemplify stream-floodplain protection, retrofitting of neighborhood stormwater flowpaths, and use of low-impact stormwater/water capture systems as creative solutions to urban stormwater management (figs. S3 to S5).

Biodiversity Changes in Cities

Within cities, urbanization and suburbanization usually reduce both species richness and evenness for most biotic communities [e.g., (*48, 50*)], despite increases in abundance and biomass of birds (*51*) and arthropods (*52*). Because the urban footprint extends far beyond municipal boundaries, urbanization may also reduce native species diversity at regional and global scales (Fig. 2). For example, urban sprawl in northern latitudes appears related to declines in abundances in some migratory birds in southern latitudes (*53*). Two exceptions to this pattern are notable: (i) Plant species richness and evenness both often increase in cities relative to wildlands (*54*–*56*), probably owing to the highly heterogeneous patchwork of habitats, coupled with human introductions of exotic species and preferences for species with few individuals of each in landscaped yards. (ii) Bird species richness may peak at intermediate levels of urbanization because of increased heterogeneity of edge habitats (*57*).

Humans often directly control plant richness, evenness, and density. Individual human and institutional choice do not directly control most other functional groups of species (herbivores, predators, parasites, omnivores, detritivores) or their trophic interactions (*52*), except for select pest species and intentionally introduced, domesticated herbivores and predators (e.g., cats). Human-dictated urban plant communities, often based on socioeconomic status, form the template for these other functional groups of species. Proposed mechanisms for changes in richness and evenness include increased rate and seasonal variability in productivity (*58*), relaxed predation on the dominant species (*59*), increased competitive abilities of some urban species (*60*), or increased parasite pressure on less successful urban species (*61*). These hypotheses are not mutually exclusive. Certain species may become better urban competitors because they are released from natural enemies.

Urbanization also alters the species composition of communities. Within cities, biological communities are often dissimilar to surrounding communities as urban species become reshuffled into novel communities (*56*). For example, bird communities often shift to more granivorous species at the expense of insectivorous species (*51*), and arthropod communities may shift from more specialized to more generalist species (*62*). Soil nematode diversity does not vary between rural and urban riparian soils, but functional composition changes to fewer predaceous and omnivorous species in urban than in rural soils (*63*). At the global scale of diversity, McKinney (*64*) argued that cities are great homogenizing forces, where some "urban-adapted" species become common in cities worldwide, and a subset of native species, usually species adapted to edges, become locally and regionally abundant at the expense of indigenous species. This homogenization of terrestrial and aquatic communities via urbanization proceeds at different rates in different geographic areas depending on human population growth and species composition (*65*).

The urban environment is a powerful selective force that alters behaviors, physiologies, and morphologies of city-dwelling organisms (*66*). Anthropogenic changes that are both direct (e.g., built structures, habitat modification and fragmentation, wildlife feeding) and indirect (e.g., altered temperatures, productivity, and light; noise and air pollution) (*67*) may cause short-term changes in phenotypes of urban-dwelling organisms [e.g., (*68*)]. In the longer term, urban environments act as a potent evolutionary force on population genetics and life-history traits of urban species (*68*). Human organisms are not immune to selective action of the urban environment. Social structure and interactions, physiology and health, morphology (e.g., increased obesity), and even long-term changes in genetics of human urban residents may be associated with urban living [e.g., (*67*)].

Given that urban land use and its footprint will continue to expand worldwide, the prognosis for maintaining diversity and function of biological communities and their associated ecosystem services within and near cities seems dire. However, intensified conservation efforts to preserve existing natural or semi-natural habitats or to reconstruct habitats within or near cities may ameliorate these biological changes (*69*). Introduction of nonnative species combined with the UHI may in some cities actually enhance ecosystem services, such as soil mineralization (*70*). Furthermore, reconciliation ecology (*69*), where habitats greatly altered for human use are designed, spatially arranged, and managed to maximize biodiversity while providing economic benefits (*57*, *69*, *70*) and ecosystem services (*64*, *71*), offers great promise that ecologists will be increasingly called upon to help design and manage new cities and reconstruct older ones (fig. S6). Cities offer real-world laboratories for ecologists to understand these fundamental patterns and processes and to work with city planners, engineers, and architects to implement policies that maximize and sustain biodiversity and ecosystem function. With an ever-increasing fraction of humans living in or near cities, these are the biological communities that humans experience—human connections and encounters with urban nature have supplanted experiences with natural biodiversity (*64*). Paradoxically, these human experiences with nonnative, global "homogenizers" (*72*), such as pigeons, may be essential for conserving global biodiversity in complex, human-modified environments.

Prospects

Cities are concentrated centers of production, consumption, and waste disposal that drive land change and a host of global environmental problems. Locally, they represent microcosms of that global environmental change and offer opportunities for enriching both ecology and global-change science. We know that the totality of human activity occurs on a biophysically constrained planet, and urban ecology can elucidate the connections between city dwellers and the biogeophysical environment in which they reside. As our ecological footprint expands, so should our perception of issues of the greater scales beyond us, and of the broader impacts of our individual and collective life-styles, choices,

and actions. Thus, our hope is that cities also concentrate the industry and creativity that have resided in urban centers throughout much of human history, making them hot spots for solutions as well as problems. Urban ecology has a pivotal role to play in finding those solutions and navigating a sustainable urban future.

References and Notes

1. United Nations Population Division, *World Urbanization Prospects: The 2005 Revision* (United Nations, New York, 2006).
2. N. B. Grimm, J. M. Grove, S. T. A. Pickett, C. L. Redman, *Bioscience* **50**, 571 (2000).
3. S. T. A. Pickett, M. L. Cadenasso, J. M. Grove, *Ecosystems (N. Y., Print)* **8**, 225 (2005).
4. S. T. A. Pickett *et al*., *Annu. Rev. Ecol. Syst.* **32**, 127 (2001).
5. S. L. Collins *et al*., "Integrated Science for Society and the Environment: A Strategic Research Initiative" (Miscellaneous Publication of the Long Term Ecological Research Network, available at www.lternet.edu/planning/).
6. L. R. Brown, *Eco-Economy: Building an Economy for the Earth* (Norton, New York, 2001).
7. C. Folke, A. Jansson, J. Larsson, R. Costanza, *Ambio* **26**, 167 (1997).
8. J. P. Kaye, P. M. Groffman, N. B. Grimm, L. A. Baker, R. V. Pouyat, *Trends Ecol. Evol.* **21**, 192 (2006).
9. L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kuhnert, G. B. West, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7301 (2007).
10. C. L. Redman, in *Human Impact on Ancient Environments*, C. L. Redman, Ed. (Univ. of Arizona Press, Tucson, 1999), pp. 127–158.
11. F. A. B. Meyerson, L. Merino, J. Durand, *Front. Ecol. Environ.* **5**, 182 (2007).
12. J. Fernandez, *J. Ind. Ecol.* **11**, 99 (2007).
13. L. Shen, S. Cheng, A. J. Gunson, H. Wan, *Cities* **22**, 287 (2005).
14. M. L. Cadenasso, S. T. A. Pickett, K. Schwarz, *Front. Ecol. Environ.* **5**, 80 (2007).
15. M. Luck, J. G. Wu, *Landscape Ecol.* **17**, 327 (2002).
16. D. B. Lewis, J. P. Kaye, C. Gries, A. P. Kinzig, C. L. Redman, *Glob. Change Biol.* **12**, 703 (2006).
17. J. M. Briggs *et al*., *Front. Ecol. Environ.* **4**, 180 (2006).
18. T. McGee, in *East West Perspectives on 21st Century Urban Development: Sustainable Eastern and Western Cities in the New Millennium*, J. Brotchie, P. Newton, P. Hall, J. Dickey, Eds. (Ashgate, Aldershot, UK, 1999), pp. 37–52.
19. K. C. Seto, M. Fragkias, *Landscape Ecol.* **20**, 871 (2005).
20. L. Robinson, J. P. Newell, J. A. Marzluff, *Landsc. Urban Plan.* **71**, 51 (2005).
21. R. E. Lang, A. C. Nelson, "Beyond the metroplex: Examining commuter patterns at the "megapolitan" scale" (Working paper of the Lincoln Institute of Land Policy, Cambridge, MA, 2007).
22. D. E. Pataki *et al*., *Glob. Change Biol.* **12**, 2092 (2006).
23. M. J. Molina, L. T. Molina, *J. Air Waste Manage. Assoc.* **54**, 644 (2004).
24. M. A. Luck, G. D. Jenerette, J. G. Wu, N. B. Grimm, *Ecosystems (N. Y., Print)* **4**, 782 (2001).
25. M. Fischer-Kowalski, *J. Ind. Ecol.* **2**, 61 (1998).
26. K. Warren-Rhodes, A. Koenig, *Ambio* **30**, 429 (2001).
27. Y. Xiao, X. M. Bai, Z. Ouyang, H. Zheng, F. Xing, *Environ. Monit. Assess.* **135**, 21 (2007).
28. M. E. Kahn, *J. Policy Anal. Manage.* **19**, 569 (2000).
29. S. Q. Zhao *et al*., *Front. Ecol. Environ.* **4**, 341 (2006).
30. United Nations Development Program, *Beyond Scarcity: Power, Poverty and the Global Water Crisis. Human Development Report 2006* (United Nations, New York, 2006).
31. J. W. Gregg, C. G. Jones, T. E. Dawson, *Nature* **424**, 183 (2003).
32. P. M. Groffman, N. L. Law, K. T. Belt, L. E. Band, G. T. Fisher, *Ecosystems (N. Y., Print)* **7**, 393 (2004).
33. T. E. Graedel *et al*., *Environ. Sci. Technol.* **38**, 1242 (2004).
34. In 1998 in Arizona, 114×10^6 m^3 of groundwater was applied to 98,542 ha of fields. If groundwater has a nitrate concentration of 10 mg N/liter, this represents 113 kg/ha of N added to fields.
35. X. D. Li, C. S. Poon, P. S. Liu, *Appl. Geochem.* **16**, 1361 (2001).
36. J. P. Kaye *et al*, *Ecol. Appl.* **18**, 132 (2008).
37. P. M. Groffman, A. M. Dorsey, P. M. Mayer, *J. N. Am. Benthol. Soc.* **24**, 613 (2005).
38. R. S. Cerveny, R. C. Balling, *Nature* **394**, 561 (1998).
39. T. R. Oke, in *Applied Climatology: Principles and Practices*, A. Perry, R. Thompson, Eds. (Routledge, London, 1997), pp. 273–287.
40. K. E. Trenberth *et al*., in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, D. Q. S. Solomon, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, H. L. Miller, Eds. (Cambridge Univ. Press, Cambridge, UK, 2007).
41. E. Kalnay, M. Cai, *Nature* **423**, 528 (2003).
42. E. G. McPherson, in *The Ecological City: Preserving and Restoring Urban Biodiversity* R. H. Platt, R. A. Rowntree, P. C. Muick, Eds. (Univ. of Massachusetts Press, Amherst, 1994), pp. 151–171.
43. M. M. Carreiro, C. E. Tripler, *Ecosystems (N. Y., Print)* **8**, 568 (2005).
44. J. L. Stoddard, D. P. Larsen, C. P. Hawkins, R. K. Johnson, R. H. Norris, *Ecol. Appl.* **16**, 1267 (2006).
45. E. S. Bernhardt, M. A. Palmer, *Freshw. Biol.* **52**, 738 (2007).
46. B. Chocat, P. Krebs, J. Marsalek, W. Rauch, W. Schilling, *Water Sci. Technol.* **43**, 61 (2001).
47. X. M. Bai, P. Shi, *Environment* **48**, 22 (2006).
48. M. J. Paul, J. L. Meyer, *Annu. Rev. Ecol. Syst.* **32**, 333 (2001).
49. C. J. Walsh *et al*., *J. N. Am. Benthol. Soc.* **24**, 706 (2005).
50. M. L. McKinney, *Bioscience* **52**, 883 (2002).
51. J. F. Chace, J. J. Walsh, *Landsc. Urban Plan.* **74**, 46 (2006).
52. S. H. Faeth, P. S. Warren, E. Shochat, W. A. Marussich, *Bioscience* **55**, 399 (2005).
53. I. Valiela, P. Martinetto, *Bioscience* **57**, 360 (2007).
54. J. M. Grove *et al*., *Ecosystems (N. Y., Print)* **9**, 578 (2006).
55. D. Hope *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8788 (2003).
56. P. G. Angold *et al*., *Sci. Total Environ.* **360**, 196 (2006).
57. J. M. Marzluff, *Urban Ecosyst.* **8**, 157 (2005).
58. E. Shochat, W. L. Stefanov, M. E. A. Whitehouse, S. H. Faeth, *Ecol. Appl.* **14**, 268 (2004).
59. J. C. Gering, R. B. Blair, *Ecography* **22**, 532 (1999).
60. J. M. Anderies, M. Katti, E. Shochat, *J. Theor. Biol.* **247**, 36 (2007).
61. D. E. Burhans, F. R. Thompson, *Ecol. Appl.* **16**, 394 (2006).
62. N. E. McIntyre, J. Rango, W. F. Fagan, S. H. Faeth, *Landsc. Urban Plan.* **52**, 257 (2001).
63. M. A. Pavao-Zuckerman, D. C. Coleman, *Appl. Soil Ecol.* **35**, 329 (2007).
64. M. L. McKinney, *Biol. Conserv.* **127**, 247 (2006).
65. J. D. Olden, *J. Biogeogr.* **33**, 2027 (2006).
66. E. Shochat, P. S. Warren, S. H. Faeth, N. E. McIntyre, D. Hope, *Trends Ecol. Evol.* **21**, 186 (2006).
67. K. J. Navara, R. J. Nelson, *J. Pineal Res.* **43**, 215 (2007).
68. J. Partecke, E. Gwinner, *Ecology* **88**, 882 (2007).
69. M. L. Rosenzweig, *Win-Win Ecology* (Oxford Univ. Press, Oxford, 2003).
70. A. J. Hansen, R. DeFries, *Ecol. Appl.* **17**, 974 (2007).
71. P. Kareiva, S. Watts, R. McDonald, T. Boucher, *Science* **316**, 1866 (2007).
72. R. R. Dunn, M. C. Gavin, M. C. Sanchez, J. N. Solomon, *Conserv. Biol.* **20**, 1814 (2006).
73. U.S. Census Bureau, in www.census.gov/population/censusdata/table-4.pdf (2007).

74. Support from the NSF grants DEB-0423704 (to the Central Arizona–Phoenix LTER), DEB-514382 (to N.B.G.), and BCS-0508002 (to C.L.R. and J.W.) is gratefully acknowledged. We thank students in the Urban Ecological Systems class at Arizona State University, S. Collins, an anonymous reviewer for comments that improved the manuscript, S. Deviche for drafting fig. S6, and H. Palmaira for help with Figs. 1 and 2.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5864/756/DC1
SOM Text
Figs. S1 to S6
Table S1
References

10.1126/science.1150195

PERSPECTIVE

The Urban Transformation of the Developing World

Mark R. Montgomery

Sometime in the next 20 to 30 years, developing countries in Asia and Africa are likely to cross a historic threshold, joining Latin America in having a majority of urban residents. The urban demographic transformation is described here, with an emphasis on estimates and forecasts of urban population aggregates. To provide policy-makers with useful scientific guidance in the upcoming urban era, demographic researchers will need to refine their data sets to include spatial factors as well as urban vital rates and to make improvements to forecasting methods currently in use.

By 2030, according to the projections of the United Nations (UN) Population Division (*1*), each of the major regions of the developing world will hold more urban than rural dwellers; by 2050 fully two-thirds of their inhabitants are likely to live in urban areas. The world's population as a whole is expected to undergo substantial further growth over the period, almost all of which is expected to take place in the cities and towns of poor countries. The total urban population of these countries was estimated by the UN Population Division to have been 1.97 billion persons in the year 2000, but that total is projected to increase to 3.90 billion by 2030 and further to 5.26 billion by 2050. This will be an enormous change in both relative and absolute terms. The urban demographic transformation influences and is influenced by four allied trends in economic development worldwide: globalization, which binds cities to each other through international networks; the decentralization of governments of poor countries, which is placing greater responsibilities on local and municipal governments (*2*); evolving international development strategies to fulfill the Millennium Development Goals, which explicitly recognize urban as well as rural poverty (*3*); and the urban implications of global climate change, which is likely to put large coastal city populations at risk from flooding, storm surges, and other extreme weather events (*4*, *5*).

In spite of its centrality to economic development, the urban transformation of poor countries has somehow largely escaped the attention of the demographic research community. When it has focused on urbanization, the demographic literature has tended to overstate the role being played by very large cities and has underemphasized the importance of small- and medium-sized cities. The literature has also given insufficient weight to urban natural increase versus rural-to-urban migration as a source of city population growth.

Department of Economics, Stony Brook University, Stony Brook, NY 11794, USA. Poverty, Gender, and Youth Program, Population Council, New York, NY 10017, USA. E-mail: mmontgomery@notes.cc.sunysb.edu

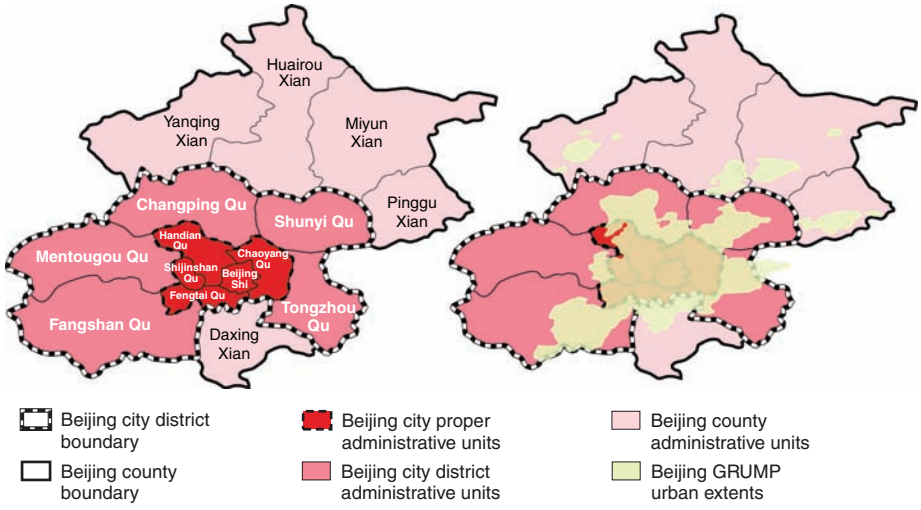


Fig. 1. Administrative Districts of Beijing Province, China.

There are policy opportunities here that warrant far more attention than has been given to date (*6*). In turning belatedly to urbanization, demographers have brought one important issue to the forefront: the poor performance of the methods currently being used to forecast city and urban growth in developing countries (*2*, *7*).

Urban and City Definitions

Much of what is known about the demography of the urban transition stems from research conducted by the UN Population Division, which since the 1970s has been the sole source of internationally comparable city and urban estimates and projections. The main challenge for its analysts is that of heterogeneity: The national definitions that yield these data vary substantially across countries and over time.

Beijing offers one example of the difficulties (*8*, *9*). For the year 2000, the population of "Beijing" was reported by Chinese authorities to be 11.5 million people. But depending on how the city boundary is drawn, the estimate could have been as low as 8.5 million had the boundary encompassed only the administrative units of the city proper (depicted in darkest shading in the left-hand side of Fig. 1). The official definition also includes the populations of surrounding city districts, which contain more rural than urban residents but are functionally linked to the city proper. Multiple social, economic, administrative, and political judgements come into play in the formulation of such city definitions, and it is not obvious that the adoption of any single definition is advisable. Although the conventional urban-rural distinction still retains value, a consensus is emerging that future classification schemes will need to reserve a place for third categories and degrees of urban-ness, as well as the rural and urban ends of the spectrum (*10*–*12*).

Although an international database that would allow for consideration of multiple definitions is not yet in hand, a template for this work has been developed by the Global Rural-Urban Mapping Project (GRUMP), which combines detailed administrative boundary data with urban and rural population counts for all countries and which has supplemented these data with imagery derived from remote sensing and other geographically coded sources (*13*, *14*). A sample of the GRUMP results can be seen in the right-hand side portion of Fig. 1, in which the irregularly shaped areas depicted in light shading indicate where the urban concentrations of population are located in Beijing province as determined by satellite observation of night-time lights (*15*). In this case, the physical proximity of lights might serve as a proxy for social, economic, and administrative interaction. New measures such as these may aid in the development of alternative urban and city definitions (*16*, *4*, *17*–*19*).

The Urban Population Transition

The following analysis has its basis in an October 2006 update of the UN Population Division's

cities database (*1*). Although the units in which city and town populations are recorded vary a good deal and systematic biases have plagued urban and city population forecasts based on these data, there is little disagreement about the broad patterns and trends.

During the period 2000–2024, the world's total population is projected to grow by 1.76 billion persons, with some 86% of this growth expected to take place in the cities and towns of developing countries (Fig. 2A). These near-term prospects stand in sharp contrast to what was experienced from 1950 to 1974, an era when rural growth still exceeded urban. The projections suggest that relatively little additional rural growth will occur in developing countries (an increase of some 190 million rural dwellers in total from 2000 to 2024) and that the UN anticipates that the rural populations of more-developed countries will continue to decline.

Among the major regions of developing countries, Asia now holds the largest number of urban dwellers and will continue to do so (Fig. 2B). By 2025, Africa will have probably overtaken Latin America in terms of urban totals, moving into second place among the regions. (The urban population of developing Oceania is also shown, but with only 1.92 million urban residents as of 2000 and 6.47 million urban dwellers projected for 2050, the totals for this region are hardly perceptible.)

In the 1950s, 1960s, and well into the 1970s, regional urban growth rates (Fig. 2C) approximated 4% per annum, although declines were already making an appearance in Latin America. Had the growth rates of this early era been sustained, the urban populations of the three regions would have doubled roughly every 17 years. By the year 2000, however, urban growth rates had fallen considerably in each of the three major regions. As Fig. 2C indicates, further growth rate declines are forecast for the first few decades of the 21st century, with urban Latin America projected to approach a state of zero growth. Much as with population growth rates overall in developing countries, the urban growth rates in force before 2000 are substantially higher than the rates that were seen during comparable historical periods in the West, with the difference being due to lower urban mortality in present-day populations, stubbornly high urban fertility in some cases, and an built-in momentum in urban growth that stems from the distinctive age and sex structures bequeathed by in-migration of young adults and past population growth (*2*). Even if the projected downward trends in growth rates come to pass, by 2050 urban growth rates in Africa would remain about 2% per annum, a rate that would double the urban population of that region in 35 years.

In each of the developing regions, the urban percentage is advancing in a seemingly inexorable fashion, and by 2030 urban majorities are projected to emerge in both Asia and Africa. Despite what is often assumed, when compared with the historical experience in Western countries, these decade-to-decade changes in urban percentages—sometimes termed the pace of urbanization—are not especially large (*2*). The literature exhibits some confusion on this point, often failing to distinguish rates of urban growth, which are rapid by historical standards, from the pace of urbanization, which falls well within the historical bounds.

What has no historical parallel is the emergence of hundreds of large cities, especially in Asia and Latin America, which each have several cities above 10 million in population. This remarkable feature of the urban transition has attracted a great deal of interest and seems to have fostered the impression that most urban residents in the developing world live in huge urban agglomerations. In fact, of all urban residents in cities of 100,000 and above in the developing world, only about 12% live in megacities, i.e., about 1 in 8 of

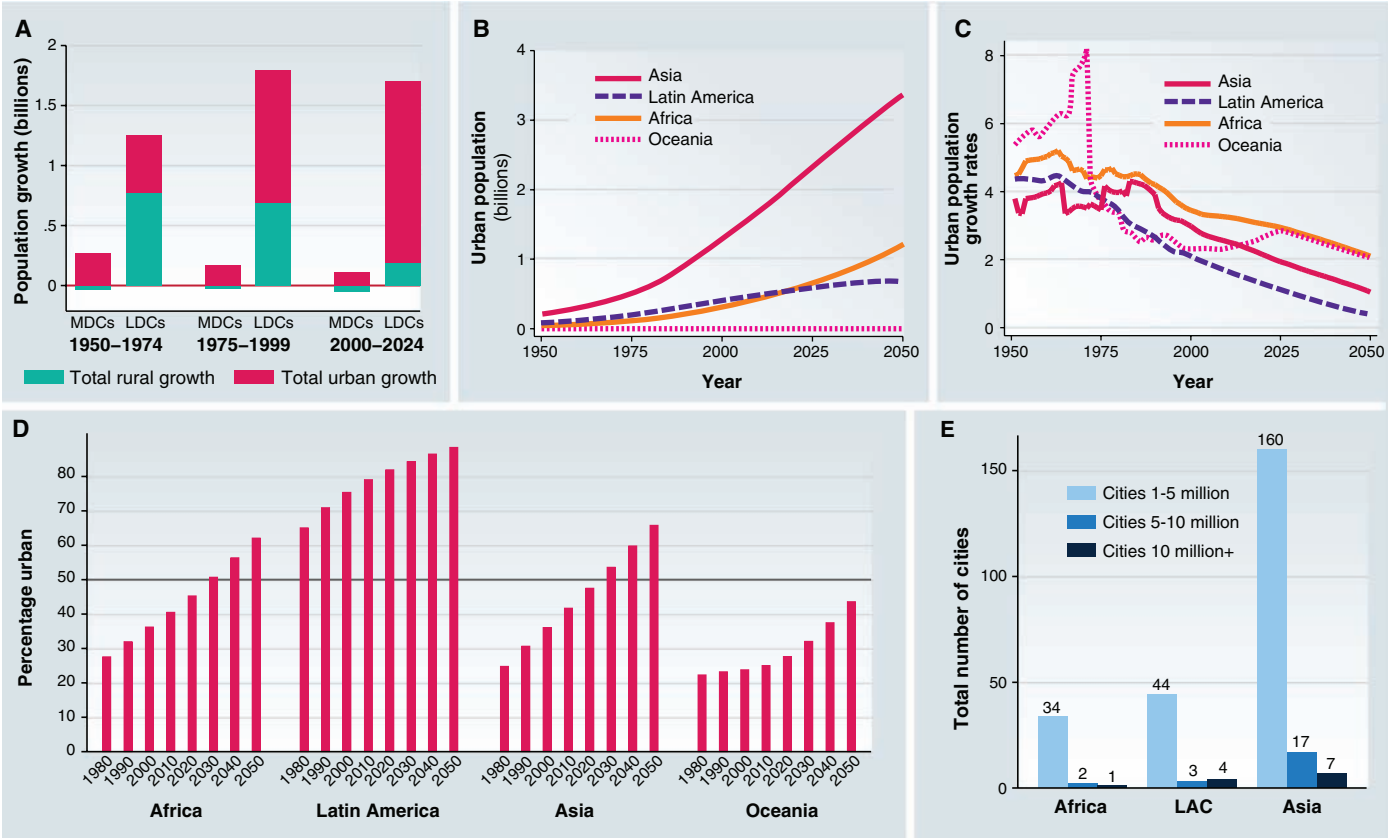


Fig. 2. (**A**) Urban population growth in more-developed countries (MDCs) and less-developed countries (LDCs), 1950–2024. (**B**) Total urban population by region in developing countries. (**C**) Growth rates of total urban population by region in developing countries. (**D**) Increasing percentage urban in developing countries. (**E**) Number of cities of 1 million residents or more in developing countries in 2000 by region.

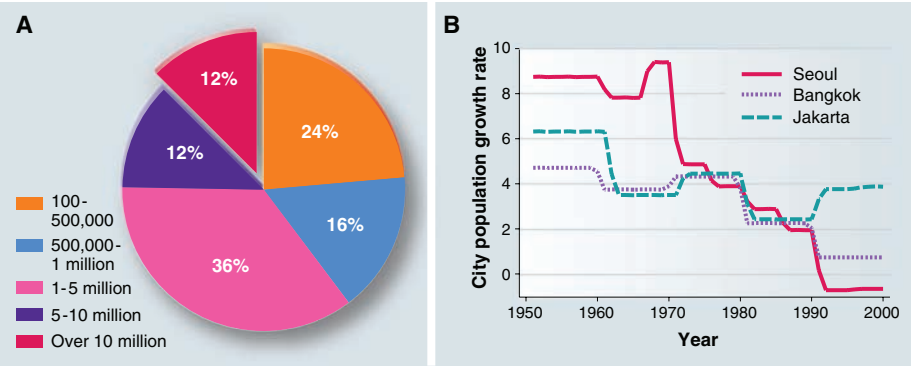


Fig. 3. (**A**) Distribution of urban population by city size in developing countries in 2000. (**B**) City growth rates for Seoul, Bangkok, and Jakarta, 1950–2000.

urban residents (Fig. 3A). Smaller cities are generally less well provided with basic services than large cities, such as improved sanitation and adequate supplies of drinking water (*2*). Rates of fertility and infant and child mortality in small cities can be little different from the rates prevailing in the countryside. Their municipal governments seldom possess the range of expertise and managerial talent found in the governments of large cities. Yet in an era of political decentralization, these smaller cities are increasingly being required to shoulder substantial burdens in service delivery and take on a larger share of revenue-raising responsibilities (*2*). Given all this, it is surprising how often small cities have been neglected in policy discussions (*20*).

The empirical record suggests that various social and spatial feedback mechanisms cause large cities to exhibit declining rates of population growth, as illustrated by the cases of Jakarta, Seoul, and Bangkok (Fig. 3B). In offering explanations, urban economists emphasize how increases in city size drive up rents and the many costs of congestion, discouraging prospective migrants and encouraging business relocation. Urban geographers stress the difficulties of locating and measuring the growth of large cities, noting that faster population growth at an urban periphery, which may not necessarily be recorded in growth rate statistics, often accompanies slower growth in the city center.

Another plausible explanation that receives far too little attention is that city growth rates are driven down over time by declines in urban fertility rates. Research by the UN Population Division, based on a sample of countries providing two or more national censuses, allows urban population growth rates to be divided into a natural urban growth component (the difference between urban birth and death rates) and a residual that combines net migration with spatial expansion (*21*). The details are complicated and the sample of countries small; nevertheless, the results are strikingly at odds with the usual perception of the sources of urban growth. In developing countries, about 60% of the urban growth rate is attributable to natural growth; the remaining 40% is due to migration and spatial expansion. Recently, a very similar rule was established for India over the 4 decades from 1961 to 2001, with urban natural growth again accounting for about 60% of the total [p. 32 of (*22*)]. Not surprisingly given its low fertility levels, the tight controls that kept migration in check until the 1980s, and the subsequent unleashing of migration, China presents something of an exception. There, the UN's estimate puts the contribution of natural urban growth at about 40% of the growth rate total.

Many developing-country policy-makers have expressed greater concern about rates of city growth in their countries than about national population growth, and they have not infrequently acted on such concerns with aggressive tactics aiming to expel slum residents and repel rural-to-urban migrants (*6*). It is therefore surprising how little attention has been paid to a growth-rate policy of a very different character: urban voluntary family-planning programs. Over the past half-century, such programs have compiled an impressive record across the developing world in facilitating fertility declines and reducing unwanted fertility. Empirical analysis of developing-country city growth and fertility suggests that when national total fertility rates decline by one child, this is associated with a decline of nearly 1 percentage point in city population growth rates for that country (*23*). Family-planning programs offer an effective and humane alternative to the ineffective and brutalizing measures that have been applied all too often.

Table 1. Urban population forecast errors for the year 2000 [from (*2*)].

	Mean percentage forecast errors		
	1980–2000	1990–2000	1995–2000
East Asia and Pacific (EAP)	3.9	26.7	−2.8
EAP excluding China	18.4	9.8	−0.4
Latin America and Caribbean	19.8	5.4	−0.9
Middle East and North Africa	13.3	6.8	8.5
South Asia	27.2	19.7	2.7
Sub-Saharan Africa	21.8	23.4	5.5
Level of development			
Low	23.1	18.3	3.2
Lower middle	6.9	26.1	−1.3
Lower middle excluding China	25.6	9.9	3.7
Upper middle	12.8	8.9	0.8

Forecasting City Growth

The performance of the UN urban and city population forecasts leaves much to be desired: As Table 1 indicates, they have consistently projected growth rates (and thus population sizes) that are too high (*2*, *7*). The mean percentage forecast errors are large for the 20-year- and 10-year-ahead forecasts; for example, the 20-year-ahead forecast for Latin America, made in 1980, proved to be 19.8% too high when the region's 2000 urban population was finally counted. The tendency to overproject is not evident in the UN's forecasts of total population at the national level, and it persists despite the insertion of an algorithm in the city forecasting model designed to slow projected growth rates as city size increases. Diagnosing the source of these errors is difficult given that the UN's method makes no use of fertility or mortality rates (which the UN projects in separate exercises) and has not yet incorporated spatially disaggregated data such as shown for Beijing (Fig. 1). Alternative forecasting methods are now being actively explored (*7*, *23*).

If the details remain in doubt while these scientific issues await resolution, at least the broad outlines of future urbanization can be perceived from the UN figures, as can the items in the research agenda that urgently need attention if demographic data and methods are to provide useful scientific guidance. Perhaps the greatest need on the demographic front is to ensure that the censuses regularly fielded by developing countries are analyzed at the level of small geographic units and the results placed in the hands of the local and municipal governments that will need to make use of such data to effectively plan for the pace and spatial distribution of future growth. Remote-sensing methods can serve as a valuable supplementary tool, if not in estimating population as such, then in monitoring the spatial spread of city populations in the intercensal periods.

References and Notes

1. United Nations, *World Urbanization Prospects: The 2005 Revision Population Database* (Department of Economic and Social Affairs, Population Division, United Nations, New York, 2005).
2. Panel on Urban Population Dynamics, *Cities Transformed: Demographic Change and Its Implications in the Developing World*, M. R. Montgomery, R. Stren,

B. Cohen, H. Reed, Eds. (National Academies Press, Washington, DC, 2003).
3. United Nations, *The Millennium Development Goals Report 2007* (United Nations, New York, 2007).
4. G. McGranahan, D. Balk, B. Anderson, *Environ. Urban.* **19**, 17 (2007).
5. R. Nicholls *et al.*, Ranking port cities with high exposure and vulnerability to climate extremes: Exposure estimates, OECD Environment Working Papers No. 1, Environment Directorate, Paris (2007).
6. UNFPA, *State of World Population 2007: Unleashing the Potential of Urban Growth* [United Nations Population Fund (UNFPA), New York, 2007].
7. P. Bocquier, *Demogr. Res.* **12**, 197 (2005).
8. K. W. Chan, Y. Hu, *China Rev.* **3**, 49 (2003).
9. K. W. Chan, *Eurasian Geogr. Econ.* **48**, 383 (2007).
10. T. G. McGee, in *The Extended Metropolis: Settlement Transition in Asia*, N. Ginsburg, B. Koppel, T. G. McGee, Eds. (Univ. of Hawaii Press, Honolulu, HI, 1991), pp. 3–25.
11. T. Champion, G. Hugo, Eds., *New Forms of Urbanization: Beyond the Urban–Rural Dichotomy* (Ashgate, Aldershot, UK, 2004).
12. T. Champion, "Where do we stand? Lessons from the IUSSP Working Group on Urbanization 2006," Center for Urban and Regional Development Studies, Newcastle University, Newcastle-upon-Tyne, UK (2006); http://eprints.ncl.ac.uk/file_store/nclep_411141727771.pdf (accessed 2 January 2008).
13. D. Balk, F. Pozzi, G. Yetman, U. Deichmann, A. Nelson, in *Proceedings of the Urban Remote Sensing Conference of the International Society for Photogrammetry and Remote Sensing* (International Society for Photogrammetry and Remote Sensing, Tempe, AZ, 2005).
14. Global Rural-Urban Mapping Project (GRUMP), alpha version, Socioeconomic Data and Applications Center (SEDAC), Columbia University, Palisades, NY. The joint effort involves the Center for International Earth Science Information Network (CIESIN), Columbia University; the International Food Policy Research Institute (IPFRI); the World Bank; and the Centro Internacional de Agricultura Tropical (CIAT) (2005). Data available at http://sedac.ciesin.columbia.edu/gpw/.
15. C. Small, *Int. J. Remote Sens.* **26**, 661 (2005).
16. S. Angel, S. C. Sheppard, D. L. Civco, *The Dynamics of Global Urban Expansion* (Transport and Urban Development Department, World Bank, Washington, DC, 2005).
17. C. Small, F. Pozzi, C. D. Elvidge, *Remote Sens. Environ.* **96**, 277 (2005).
18. K. C. Seto, M. Fragkias, *Landscape Ecol.* **20**, 871 (2005).
19. A. Schneider, K. C. Seto, D. Webster, *Environ. Plann. B* **32**, 323 (2005).
20. UN-Habitat, *Meeting Development Goals in Small Urban Centres* (UN-Habitat and Earthscan, London, 2006).
21. N. Chen, P. Valente, H. Zlotnik, *Migration, Urbanization, and Development: New Directions and Issues*, R. E. Bilsborrow, Ed. (UNFPA, New York, 1998), pp. 59–88.
22. K. Sivaramakrishnan, A. Kundu, B. Singh, *Handbook of Urbanization in India: An Analysis of Trends and Processes* (Oxford Univ. Press, New Delhi, 2005).
23. M. R. Montgomery, D. Balk, *Global Urbanization in the 21st Century*, E. L. Birch, S. M. Wachter, Eds., The City in the 21st Century (Univ. of Pennsylvania Press, Philadelphia, 2008).
24. Thanks are due to H. Zlotnik, Director of the UN Population Division; T. Buettner, assistant director and chief, Population Studies Branch; and G. Heilig, chief of the Division's Estimates and Projection Section, for making available the October 2006 version of the UN's cities database. I also thank D. Balk of Baruch College, V. Mara of CIESIN, L. Forouzan of Baruch, and S. Henning of the UN Population Division for preparing Fig. 1, in which they were guided by the work of K. W. Chan of the Department of Geography, University of Washington, Seattle. The work reported here was supported in part by an award from the William and Flora Hewlett Foundation to the Population Council.

10.1126/science.1153012

PERSPECTIVE

Reproducing in Cities

Ruth Mace

Reproducing in cities has always been costly, leading to lower fertility (that is, lower birth rates) in urban than in rural areas. Historically, although cities provided job opportunities, initially residents incurred the penalty of higher infant mortality, but as mortality rates fell at the end of the 19th century, European birth rates began to plummet. Fertility decline in Africa only started recently and has been dramatic in some cities. Here it is argued that both historical and evolutionary demographers are interpreting fertility declines across the globe in terms of the relative costs of child rearing, which increase to allow children to outcompete their peers. Now largely free from the fear of early death, postindustrial societies may create an environment that generates runaway parental investment, which will continue to drive fertility ever lower.

Reproducing in cities has been seen as a difficult enterprise ever since cities were created. Cities were, and often still are, havens of disease and crime, characterized by costly and crowded housing, transitory communities, and a lack of kin support among inhabitants. Indeed, the notion that the "country" is the best place to rear children dates back to the Romans (*1*). Urban dwellers faced higher rates of infant mortality than their rural compatriots right up to the early 20th century (*2*), such that most cities were only maintained by a constant influx of migrants. Infanticide and infant abandonment were used as means of controlling family size, and even if rescued, the prospects for foundlings were precarious (*3*). The famous 18th-century naval captain and philanthropist Thomas Coram, when home from his seafaring duties, was distressed by the number of abandoned babies he observed on the streets of London, which inspired him to set up England's first foundling hospital in 1741 (*4*). Unfortunately, similar institutions in other cities could not sustain such levels of philanthropic support in the face of rising numbers of foundlings.

An alternative to such drastic and cruel measures to defray the costs of child rearing is not to have so many children in the first place. The strategy of limiting fertility through techniques such as delaying marriage and sexual abstinence within marriage were not limited to city life; we know they were first used by rural groups, ranging from African pastoralists relying on slow breeding livestock such as camels (*5*) to preindustrial German farmers trying to maintain the integrity of their farms by reducing the numbers of offspring claiming the inheritance (*6*). The industrial revolution and its resulting urbanization seem ultimately to have led to a population-wide desire to curb fertility by these techniques. In the late 19th and early 20th centuries, fertility in Britain plummeted in 70 years, from families of six to eight children being common to families of more than four children becoming rare. This fertility decline has since become a global phenomenon, albeit with many local differences. Family sizes shrunk first throughout Europe and then in the Americas and Asia, and by the end of the 20th century, this process finally began in earnest in Africa (*7*). There is a broad association between declining fertility and declining mortality across countries, although the latter does not necessarily precede the former across individual states or regions, and fertility has continued its downward trend in the West, long after mortality rates were greatly reduced. Here, I have taken an evolutionary perspective to explain some of the mechanisms that might be underpinning this phenomenon.

The fertility decline in Africa is happening fastest in urban areas. In Ethiopia (Fig. 1), a difference of nearly four children between family sizes in the capital city (Addis Ababa) and in rural areas is the highest such difference seen anywhere (*8*, *9*). Although HIV/AIDS can depress fertility to some extent (*10*), this cannot fully explain the urban fertility decline, which started before the HIV epidemic took hold. The rural-urban difference is not specific to Ethiopia (even in Europe, where the rural-urban division is somewhat blurred, fertility is still higher in rural than in urban areas). But examining a case where the difference is so dramatic highlights what underlies this trend. Contraceptive services

Department of Anthropology, University College London, Taviton Street, London WC1H 0BW, UK.
E-mail: r.mace@ucl.ac.uk

are now readily available in much of urban Africa, but that cannot provide a general explanation, because the fertility declines in Europe predated the existence of modern contraceptives. Other underlying causes must be at work.

Like many cities, Addis Ababa has grown rapidly, with many recent migrants whose parents or grandparents might be living in distant rural areas. Housing is crowded and costly, often forcing people to travel long distances to work. Unskilled labor is paid very poorly, and many unskilled women work as domestic servants or sex workers (two professions not easily compatible with motherhood). Among the urban poor, marriage rates are low and divorce rates are high as individuals struggle to support themselves, let alone generate surplus means that could support a family.

Fig. 1. A camel bringing sacks of charcoal from the countryside into Addis Ababa, Ethiopia: an African city showing marked declines in family size. [Photo by R. Mace]

From an evolutionary demographic perspective, it is not surprising that the poorest in Addis Ababa are having the greatest difficulty building families. But this trend is not typical of historical fertility declines, where generally the wealthy were the first to reduce their fertility. Taking Ethiopia as a whole, the very poorest communities are rural, but for those living on farms, the additional costs of raising a family are low; so, as in most countries, the rural poor are outreproducing the relatively better-off city dwellers. Yet the increasing shortage of farmland, combined with heavy agricultural work loads, lower levels of public services, fewer opportunities for education, few means to escape poverty, and the associated risk of hunger, continues to push migrants into the city (where, unlike their 19th-century equivalents during European industrialization, they do not suffer increased infant mortality). Migrants clearly believe that they will be better off in the city, despite the much higher cost of living and thus the reduced chance of marriage (*8*). Indeed, higher urban body mass indexes and lower urban infant mortality rates imply generally better access to food and medical facilities in an urban environment (*11*).

It is a basic tenet of evolutionary ecology that individuals with more resources generally have more offspring. So the negative relationship between wealth and fertility that commonly emerges, at least in large heterogeneous populations like nation-states, presents something of a challenge for evolutionary demographers to explain. However, in homogeneous subpopulations, wealth usually does correlate positively with family size (*12*). This difference in direction between local and global trends can occur if subpopulations each have different costs associated with raising children, and the levels of parental investment per child (such as food, care, education, or material wealth to give them) vary as a result (Fig. 2). Paradoxically, a decrease in the cost of children can lead not only to larger families but also to greater poverty and higher malnutrition, because each child can survive on less food or with lower levels of inherited wealth (*13*, *14*); having a larger family maximizes reproductive success but goes hand in hand with lower levels of parental investment per child. In contrast, when the cost of family building is raised, the opposite occurs: More has to be invested in each offspring to enable them to go on to reproduce, and simulations have shown that this can lead to a more prosperous society (*14*).

Industrialization and urban living enable new professions to emerge, some of which are only available to those invested with considerable capital or training. In Britain, wealthy professionals were among the first to reduce their fertility, although different groups reduced their fertility at different times: for example, factory workers in cotton mills lowered their fertility far earlier than did coal miners. Historian Simon Szreter summed up this diversity of declining fertility rates in Britain and elsewhere as being driven by the "perceived relative costs of child rearing" (*15*). Whether perceived relative costs are equivalent to actual costs is a moot point; but, in essence, historical and evolutionary demographers are converging on similar explanations for demographic change. The cost of raising a child includes enabling it to compete with its peers—for marriage partners, for jobs, or for the means to support a family—and if that competition increases costs, then basic evolutionary ecology predicts that optimal fertility will decline (*16*). Education introduced a new mechanism through which children could compete for future employment opportunities. School also adds pressure on parents to present adequately fed and dressed offspring for public scrutiny. Culturally acceptable levels of parental investment then rise; but is this process based on real changes in costs and benefits, or is it simply based on a cultural shift that has swept around the world in the 20th century, as many demographers believe (*17*)?

That a trend could be adopted globally by chance stretches credulity: Some common process must underlie the ever-increasing perceived relative costs of child rearing. Evolutionary theory provides us with mechanisms, well described in the biological literature, through which competition for mates can generate runaway selection leading to ever more costly courtship displays (*18*–*20*). Models have shown that runaway processes can also occur in the evolution of cultural traits, either using models based on biased modes of cultural transmission (*21*) or on the ultimate value of the trait (*22*). In post-industrial urban societies over the long term, reproductive strategies based on having few, higher quality offspring could be more successful than strategies based on having more, lower quality offspring (*22*).

Transfers of resources from parents to offspring are key to understanding human life-history evolution (*23*). In wealth-owning societies, siblings compete with each other for their parents' material and intellectual resources (*24*–*26*). If parental investment is a key influence on children's future success, and the ability to invest effectively in children is heritable (and cultural traits

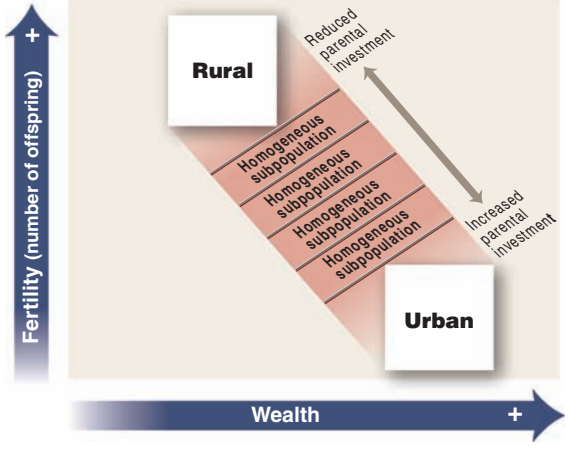


Fig. 2. Schematic diagram of how different levels of parental investment per child can generate positive relationships between fertility and wealth in subpopulations (where each of the diagonals represents a different subpopulation within a larger society) but a negative relationship between wealth and fertility over a large heterogeneous population. Levels of parental investment per child may be highest in urban areas.

such as wealth and status are usually highly heritable), then it is possible that runaway cultural selection has occurred in preferred levels of investment in each child (*27*), driving the quantity/quality trade-off further in the direction of offspring quality. Hence, I argue that the emergence of postindustrial life, now largely free from the fear of early mortality, seems to have generated conditions under which a runaway process of ever-escalating levels of investment in our children continues to drive fertility ever lower.

References

1. L. Betzig, in *Human Nature: A Critical Reader*, L. Betzig, Ed. (Oxford Univ. Press, New York, 1997).
2. J. A. Birchenall, *World Dev.* **35**, 543 (2007).
3. C. Panter-Brick, M. T. Smith, Eds. *Abandoned Children* (Cambridge Univ. Press, Cambridge, 2000).
4. S. B. Hrdy, *Mother Nature: Maternal Instincts and How They Shape the Human Species* (Vintage, London, 2000).
5. E. A. Roth, *Am. Anthropol.* **95**, 597 (1993).
6. E. Voland, R. I. M. Dunbar, *J. Biosoc. Sci.* **29**, 355 (1997).
7. B. Cohen, *World Dev.* **26**, 1431 (1998).
8. E. Gurmu, R. Mace, *J. Biosoc. Sci.*; published online 11 January 2008 (10.1017/S002193200700260X).
9. A. Sibanda, Z. Woubalem, D. P. Hogan, D. P. Lindstrom, *Stud. Fam. Plann.* **34**, 1 (2003).
10. S. Gregson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14586 (2007).
11. Central Statistical Agency, *Ethiopia Demographic and Health Survey, 2005* (ORC Macro, Addis Ababa, 2006).
12. B. S. Low, in *Adaptation and Human Behavior: An Anthropological Perspective*, L. Cronk, N. Chagnon, W. Irons, Eds. (Aldine de Gruyter, New York, 2000), pp. 323–343.
13. M. A. Gibson, R. Mace, *PLoS Med.* **3**, e87 (2006).
14. R. Mace, *Philos. Trans. R. Soc. London Ser. B* **353**, 389 (1998).
15. S. Szreter, *Fertility, Class, and Gender in Britain, 1860–1940* (Cambridge Univ. Press, Cambridge, 1996).
16. D. Lack, in *Evolution as a Process*, J. Huxley, A. C. Hardy, H. B. Ford, Eds. (Allen and Unwin, London, 1954), pp. 143–156.
17. J. Bongaarts, S. C. Watkins, *Popul. Dev. Rev.* **22**, 639 (1996).
18. R. A. Fisher, *The Genetical Theory of Natural Selection* (Oxford Univ. Press, Oxford, 1930).
19. A. Pomiankowski, Y. Iwasa, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5106 (1998).
20. H. Kokko, R. Brooks, J. M. McNamara, A. I. Houston, *Proc. R. Soc. London Ser. B Biol. Sci.* **269**, 1331 (2002).
21. R. Boyd, P. J. Richerson, *Culture and the Evolutionary Process* (Univ. of Chicago Press, Chicago, 1985).
22. J. M. McNamara, A. I. Houston, in *Social Information Transmission and Human Biology*, J. C. K. Wells, S. Strickland, K. Laland, Eds. (CRC, Boca Raton, FL, 2006), pp. 59–88.
23. R. D. Lee, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9637 (2003).
24. R. Mace, *Behav. Ecol. Sociobiol.* **38**, 75 (1996).
25. D. B. Downey, *Am. Sociol. Rev.* **60**, 746 (1995).
26. P. Kristensen, T. Bjerkedal, *Science* **316**, 1717 (2007).
27. R. Mace, in *The Oxford Handbook of Evolutionary Psychology*, R. I. M. Dunbar, L. Barrett, Eds. (Oxford Univ. Press, Oxford, 2007), pp. 383–396.

10.1126/science.1153960

PERSPECTIVE

Health and Urban Living

Christopher Dye

The majority of people now live in urban areas and will do so for the foreseeable future. As a force in the demographic and health transition, urbanization is associated with falling birth and death rates and with the shift in burden of illness from acute childhood infections to chronic, noncommunicable diseases of adults. Urban inhabitants enjoy better health on average than their rural counterparts, but the benefits are usually greater for the rich than for the poor, thus magnifying the differences between them. Subject to better evidence, I suggest that the main obstacles to improving urban health are not technical or even financial, but rather are related to governance and the organization of civil society.

For the citizens of 18th century London and Paris, it was the worst of times. As far as public health was concerned, the best of times would be for future generations. By modern health standards, London in the 1700s was a slum: Between 10% and 30% of infants died before their first birthdays (*1*), and although the death rate of young children was lower in richer parts of the city, there was little variation in life expectancy across social classes. Edwin Chadwick's "sanitation revolution" gained momentum in the early 1800s and was given greater impetus by the Public Health Act of 1848, but even in 1858, the River Thames brought "the sewage of three millions of people…to seethe and ferment…in one vast open cloaca." [Winslow in (*2*)]. Conditions were no better in 19th century Paris: Relatively high food prices and poor sanitation left Parisian men more stunted than men elsewhere in France (*3*, *4*). In Europe today, about 70% of people live in urban areas. In the Europe of 1800, only 10 to 15% of people did so, partly because of the atrocious living conditions. Cholera, dysentery, measles, plague, smallpox, tuberculosis, typhus, and other infections, exacerbated by undernourishment, imposed an "urban penalty" such that deaths, mostly of children, exceeded births (Fig. 1). London, Paris, and other European cities could only grow by immigration from the countryside (*5*).

The 1848 act focused on sanitation—piping clean water to homes and safely disposing of human waste—but led on to a wider range of environmental improvements that had benefits for health, including ventilation of dwellings and streets, the preservation of green spaces, and the upgrading of road surfaces (*6*). By the start of the 20th century, urban health was typically improving faster than rural health in the industrialized world, and towns and cities grew faster than their hinterlands. As cities expanded, they started to provide a variety of indirect benefits to health: large markets with a steady and diverse food supply, economies of scale with low transportation costs, organized public services, and a critical mass of educated people that was needed

Fig. 1. An illustration of the beginning of a cholera epidemic in Paris, April 1832, by Jules Pelcoq, from an 1866 edition of *Histoire Populaire de la France*, published by Hachette. [Stefano Bianchetti/Corbis]

World Health Organization, 1211 Geneva 27, Switzerland. E-mail: dyec@who.int

to establish centers of enterprise, learning, and innovation.

Today, more than half the world's population—about 3.3 billion people—lives in urban areas (*7*), including roughly 50,000 settlements of at least 50,000 people (*8*, *9*). By 2015, and for the foreseeable future beyond, population growth will be mainly urban, mainly in the 500 or so cities that have 1 million to 10 million inhabitants, and mainly in poorer countries. Although three-quarters of the people who earn less than a dollar a day still live in rural areas, the proportion and number of poor people living in urban areas are rising (*10*). About one in three urban inhabitants—roughly one billion people—now live in slums, but the proportions are much higher than this average in sub-Saharan Africa and South Asia (*7*).

Some of these contemporary statistics, set against the historical backdrop of urbanization in Europe, give reason for thinking that life in built-up areas could be worse than in the surrounding countryside. The risks to health are obvious where urban water supplies are polluted, coastal sites are susceptible to flooding, crowding promotes the spread of infectious diseases, electricity supplies are intermittent, health services are inaccessible, life without family and social support is desolate, and roads and recreation areas are dangerous. Yet the sanitation revolution, and its aftermath, makes it clear that urban health has the potential to be far better than rural health. Although the nascent literature on urban living gives examples of both positive and negative effects, the general features of urban health are only just being described and explained. Here, I describe five characteristics of urban health that underpin the debate about how to foster healthy urban living in the future.

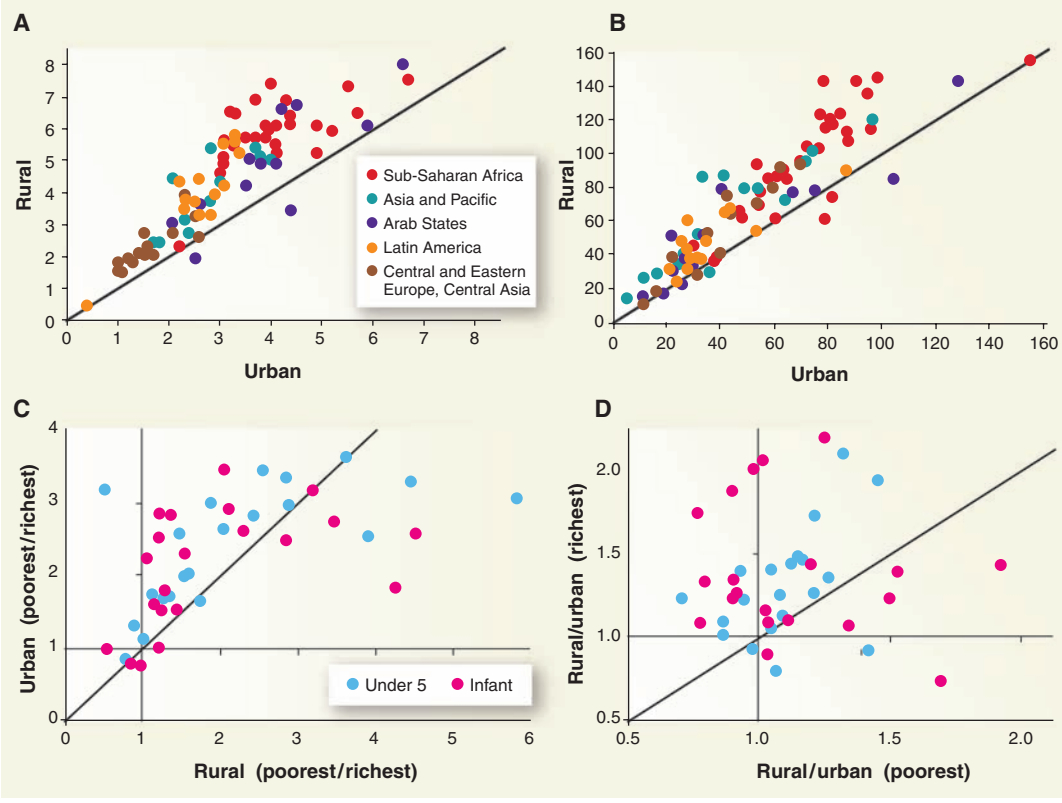


Fig. 2. (**A**) Fertility rates (births per woman 15 to 49 years) and (**B**) infant mortality rates (<1 year, per 1000 births) are typically higher in rural than urban areas (i.e., above the diagonals that mark the points where rates are the same in urban and rural areas). Regions in (B) are color-coded as in (A). (**C** and **D**) Child mortality ratios. Inequality in infant (red) and under 5 (blue) mortality tends to be greater in urban than rural areas, as judged from mortality ratios in (C) poorest 20%/richest 20% of families and (D) rural/urban areas (ratios mostly above diagonal). (D) Although urban living is especially beneficial for the rich, the poor generally benefit, too (ratios mostly >1). Poverty and wealth are determined from a household asset score in Demographic and Health Surveys (*20*). Data in (A) and (B) are from 94 and 90 countries, respectively (*13*), data in (C) are from 22 surveys in 18 countries, and data in (D) are from 22 surveys in 17 countries (*21*).

First, within countries, health is generally better on average in urban than in rural areas (*8*, *10*, *11*). By contrast with Europe up to the 19th century, births exceed deaths in most, if not all, urban agglomerations today. Consequently, many urban centers now show endogenous growth, in addition to growth by immigration from the countryside. Furthermore, although the number of slum-dwellers is growing in most parts of the world, the number of richer people is growing faster, mainly as a result of the wide range of economic opportunities that cities provide. Between 1990 and 2001, the slum population of Indonesia grew by 1.4% annually, but the whole urban population was rising at 4.4%, doubling in 16 years (*12*). In general, slum dwellers are a diminishing fraction of urban populations, and this is one reason that urban health is, on average, getting better.

The comparative health advantage of urban living is also revealed in lower fertility (Fig. 2A) and infant mortality rates (Fig. 2B), which have numerous interlinked determinants, including improved sanitation and nutrition, and easier access to contraception and health care (*13*, *14*). The strong correlations in Fig. 2, A and B, make a second important point: Although fertility and mortality rates tend to be lower in cities, the rates in urban and rural areas remain tied. In comparisons among countries, low urban mortality is associated with low rural mortality. Cities do not exist in isolation; they are part of the "national metabolism." Studies on the link between urban and rural poverty have suggested that the growing wealth of cities brings direct benefits to people living in rural areas (*10*).

Third, while urbanization appears to be a force for better health (*10*), the force does not operate in the same way everywhere. In comparisons among countries, 40% of the variation in child mortality is explained by the proportion of the population living in urban areas, but most of this (34%) is due to interregional differences (Fig. 3) (*15*, *16*). The mortality rate of children under 5 years in sub-Saharan Africa is about 10 times as high as it is in the established market economies, but in neither region is child mortality much affected by the level of urbanization. Clearly, health can and does tend to improve with urbanization, but the scale of the benefits is conditional on other factors, such as the effectiveness of public services and the opportunity for private enterprise.

Fourth, the health benefits of urbanization are not uniform (*17*–*19*). Urbanization, poverty reduction, and improvements in health are linked through economic growth (*10*), but economic growth is also associated with greater health inequalities within countries, as measured in

terms of the variation in child mortality, stunting, underweight, and life expectancy, and the association holds within the richest and the poorest nations alike (*19*).

Urban living is one factor associated with these growing disparities. Demographic and Health Surveys (DHS) (*20*, *21*), which provide unusually reliable data on urban health, reveal that the children of both rich and poor families gain from urban living, but the rich gain more. The ratio of child mortalities (infant and under 5 years) in the poorest 20% of families to the richest 20% is typically higher in urban than rural areas (Fig. 2C). To reinforce the point, rural/urban mortality ratios are mostly greater than 1 (upper right quadrant), but higher for the richest than for the poorest families (above diagonal, Fig. 2D) (*21*). Ratios less than 1 sometimes arise when the children of the urban poor (e.g., slum dwellers) suffer high mortality rates compared with the rural population (*22*, *23*). However, while urbanization magnifies the disparities in child survival in many countries, it does not do so everywhere; the exceptions revealed in DHS include Bolivia, the Dominican Republic, Egypt, Indonesia, Morocco, and Peru.

Fifth, we may assume that the health of adults, as for children, tends to be better in urban areas. However, no investigation has yet shown that the health benefits of urban living generally outweigh the health risks. City dwellers are comparatively wealthy and lead more sedentary lives with easier access to low-cost, low-fiber, high-energy, high-fat food. The proportion of adults (and children) who are overweight is rising in both rural and urban settings, but it is rising faster in cities (*24*), with implications for the incidence of diabetes, heart disease, certain cancers, and stroke. Nevertheless, where a higher proportion of people is overnourished, a lower proportion is undernourished (*24*), reducing stunting, wasting, and other conditions due to micro- and macronutrient deficiencies. In some countries, such as China, indoor air pollution is worse in certain rural than urban areas and has a bigger impact on chronic obstructive pulmonary disease (*25*). The poorer inhabitants of cities, though, are often exposed both to indoor and outdoor air pollution, and the effects of air pollutants on lung diseases in cities have not been systematically measured. Traffic accidents, mostly in the rapidly growing, congested cities of developing countries, now kill more than a million children and adults each year, and the number of casualties worldwide is bound to rise (*26*). However, richer countries have shown how measures to improve road safety can be rewarded by a steady fall in casualties. Communicable diseases pose both greater and lesser risks to adults in towns and cities, depending on the life cycles of the pathogens involved. The lower incidence of malaria in urban areas (*27*) may, depending on the setting, be offset by the greater incidence of HIV infection (*28*) or tuberculosis (*29*). Whether the transmission of Chagas disease and dengue is augmented or diminished in urban areas depends on details such as the quality of house construction (resting sites for triatomine insect vectors in cracked walls) and the distribution of standing water (habitat for larval mosquitoes).

In sum, urbanization is a force in the global demographic transition from high to low birth rates and short to long life spans, and in the health and nutritional transitions that are shifting the burden of illness from acute childhood infections to chronic and mostly noncommunicable diseases of adults. Children have a higher chance of surviving to adulthood in urban areas, but the potential benefits of urbanization have been unevenly exploited around the world. The urban environment favors many of its inhabitants, but especially the rich and not always the poor. Adults, too, probably enjoy better health in cities, but hard facts are hard to find.

Although rural-urban comparisons provide a convenient framework for analysis, the urban health goal is not simply to be better on average than rural areas. Nor is it satisfactory to trade one medical condition against another in summary statistics on life expectancy. The ambition must be to attain good health for all absolutely. In this context, the Millennium Development Goal of achieving "a significant improvement in the lives of at least 100 million slum dwellers" by 2020, a mere 10% of current numbers, is decidedly modest (*30*). In deciding, based on the available evidence, how to reach and exceed these goals, some provocative comparisons can be made between cities. For instance, why are infant mortality rates in Curitiba, Brazil (~10/1000 births), so much less than in Nairobi, Kenya (~40/1000 births on average and over 90/1000 births in slums) (*22*)? Among the many differences between these two cities, which are crucial?

Many of the prescriptions for better urban health are in fact self-evident and are often inexpensive: healthy housing, primary health care, communicable disease control through sanitation and vaccination, safe roads, and targeted assistance to women. They are also not specific to urban areas. The tough problem is that technical solutions need a framework in which they can be executed. Hence, the call for "healthy governance," regulated land ownership, probity in financial investment, social cohesion, the empowerment of civil society, and foresight in planning the physical environment (*7*, *11*, *30*). The right structure is hard to create because there are no recipes for social cohesion and good governance. Yet there is an imperative to succeed: If cities are the "defining artifacts of civilisation" (*31*), a nation may now be judged by the health of its urban majority.

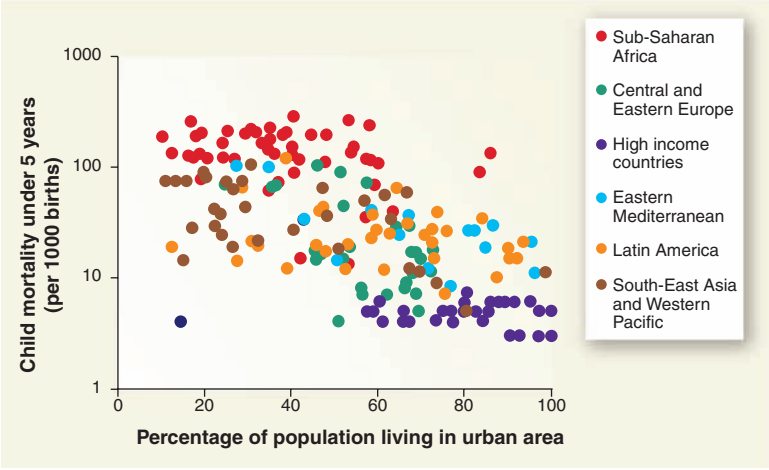


Fig. 3. Mortality rates of children under 5 years are lower in countries that are more urbanized, but mainly through differences between rather than within regions. Data are for 2005 (*15*, *16*).

References and Notes

1. G. Clark, *A Farewell to Alms: A Brief Economic History of the World* (Princeton University Press, Princeton, 2007).
2. R. A. Easterlin, *Eur. Rev. Econ. Hist.* **3**, 257 (2006).
3. G. Postel-Vinay, D. Sahn, "Explaining stunting in nineteenth century France" [Laboratoire d'Economie Appliquee, Institut National de la Recherche Agronomique (INRA), 2006].
4. D. S. Barnes, *The Great Stink of Paris and the Nineteenth-Century Struggle Against Filth and Germs* (Johns Hopkins Univ. Press, Baltimore, 2006).
5. E. A. Wrigley, *Poverty, Progress, and Population* (Cambridge Univ. Press, Cambridge, 2004).
6. C. Hamlin, S. Sheard, *BMJ* **317**, 587 (1998).
7. United Nations Population Fund, "State of world population 2007: unleashing the potential of urban growth" (United Nations Population Fund, New York, 2007).
8. S. Galea, D. Vlahov, *Annu. Rev. Public Health* **26**, 341 (2005).
9. D. Satterthwaite, "The scale of urban change worldwide 1950–2000 and its underpinnings" (International Institute for Environment and Development, Human Settlements Discussion Paper Series, 2005).
10. M. Ravallion, *Finance Dev.* **2007**, 15 (2007).
11. R. Godfrey, M. Julien, *Clin. Med. (Northfield Il)* **5**, 137 (2005).

12. United Nations Human Settlements Programme (UN-HABITAT), "Global urban observatory, urban indicators programme, phase III" (UN-HABITAT, New York, 2005).
13. United Nations Population Fund and the Population Reference Bureau, "Country profiles for population and reproductive health: policy developments and indicators" (UNFPA and the Population Reference Bureau, New York, 2005).
14. M. Garenne, in *Africa on the Move: African Migration in Comparative Perspective* M. Tienda, Ed. (Wits Univ. Press, Johannesburg, South Africa, 2006) pp. 252–279.
15. United Nations Population Division, "World urbanization prospects: the 2005 revision population database" (United Nations Population Division, New York, 2006).
16. World Bank, "World development indicators" (World Bank, Washington, DC, 2007).
17. C. Stephens, *Environ. Urban.* **8**, 9 (1996).
18. S. Yusuf, K. Nabeshima, W. Ha, *J. Urban Health* **84**, 35 (2007).
19. A. Wagstaff, "Inequalities in health in developing countries: swimming against the tide?" (World Bank, Washington, DC, 2002).
20. Measure Demographic and Health Surveys, (Measure DHS, 25 July 2007); www.measuredhs.com/.
21. D. Gwatkin, K. Johnson, A. Adam Wagstaff, S. Rutstein, R. Pande, "PovertyNet Library: socio-economic differences in health, nutrition, and population" (World Bank, Washington, DC, 2007); http://poverty2.forumone.com/library/view/15080.
22. African Population and Health Research Center (APHRC), "Population and health dynamics in Nairobi's informal settlements" (African Population and Health Research Center, 2002).
23. I. M. Timaeus, L. Lush, *Health Transit. Rev.* **5**, 163 (1995).
24. B. M. Popkin, *Am. J. Clin. Nutr.* **84**, 289 (2006).
25. D. M. Mannino, S. A. Buis, *Lancet* **370**, 765 (2007).
26. World Health Organization, "World report on road traffic injury prevention" (World Health Organization, Geneva, Switzerland, 2004).
27. S. I. Hay, C. A. Guerra, A. J. Tatem, P. M. Atkinson, R. W. Snow, *Nat. Rev. Microbiol.* **3**, 81 (2005).
28. UNAIDS, Joint United Nations Programme on HIV/AIDS, "Report on the global AIDS epidemic" (UNAIDS, New York, 2006).
29. V. K. Chadha, P. Kumar, P. S. Jagannatha, P. S. Vaidyanathan, K. P. Unnikrishnan, *Int. J. Tuberc. Lung Dis.* **9**, 116 (2005).
30. Knowledge Network on Urban Settings, World Health Organization Commission on Social Determinants of Health, "Our cities, our health, our future: acting on social determinants for health equity in urban settings" (World Health Organization Kobe Centre, Japan, 2007); www.who.or.jp/knusp/knus.html.
31. J. Reader, *Cities* (William Heinemann, London, 2004).
32. I thank D. Gwatkin, E. Rehfuess, B. Williams, A. Bierrenbach, and K. Lonnroth for helpful comments on the manuscript.

10.1126/science.1150198

PERSPECTIVE

The Size, Scale, and Shape of Cities

Michael Batty

Despite a century of effort, our understanding of how cities evolve is still woefully inadequate. Recent research, however, suggests that cities are complex systems that mainly grow from the bottom up, their size and shape following well-defined scaling laws that result from intense competition for space. An integrated theory of how cities evolve, linking urban economics and transportation behavior to developments in network science, allometric growth, and fractal geometry, is being slowly developed. This science provides new insights into the resource limits facing cities in terms of the meaning of density, compactness, and sprawl, and related questions of sustainability. It has the potential to enrich current approaches to city planning and replace traditional top-down strategies with realistic city plans that benefit all city dwellers.

Throughout the 19th century, social commentators universally damned the growth of cities, the chorus rising to a crescendo in the writings of William Morris, who spoke of "the hell of London and Manchester" and "the wretched suburbs that sprawl all round our fairest and most ancient cities" (*1*). These sentiments have dominated our approach to cities and their planning to this day: Cities are still seen as manifesting a disorder and chaos requiring control through the imposition of idealized geometric plans. There have been few dissenting voices, an exception being Jane Jacobs (*2*), who argued half a century ago that far from being homogeneous and soulless, cities are essential crucibles for innovation, tolerance, diversity, novelty, surprise, and most of all, for economic prosperity.

In the past 25 years, our understanding of cities has slowly begun to reflect Jacobs's message. Cities are no longer regarded as being disordered systems. Beneath the apparent chaos and diversity of physical form, there is strong order and a pattern that emerges from the myriad of decisions and processes required for a city to develop and expand physically (*3*). Cities are the example par excellence of complex systems: emergent, far from equilibrium, requiring enormous energies to maintain themselves, displaying patterns of inequality spawned through agglomeration and intense competition for space, and saturated flow systems that use capacity in what appear to be barely sustainable but paradoxically resilient networks.

Centre for Advanced Spatial Analysis, University College London, 1-19 Torrington Place, London WC1E 6BT, UK. E-mail: m.batty@ucl.ac.uk

The Size and Scale of Cities

Urban complexity has its basis in the regular ordering of size and shape across many spatial scales (*4*). Cities grow larger to facilitate a division of labor that generates scale economies (*5*), and it is a simple consequence of competition and limits on resources that there are far fewer large cities than small. However, the self-similarity observed across many spatial levels implies that the processes that drive agglomeration and clustering in small cities are similar to those in large cities; indeed in cities of any size.

A lot of the work on scaling has taken cities, firm sizes, and incomes as key exemplars. In the 1930s, Christaller first showed that market areas or hinterlands around cities scaled across a geometric hierarchy in terms of their population size (*6*). Gibrat (*7*) argued that such scaling could be approximated from log-normal distributions, which emerge when objects (cities and firms) grow randomly but proportionately, whereas Simon's simple birth and death models (*8*) have been widely applied to demonstrate the same logic. Recently Gabaix, Solomon, and others (*9*, *10*) have shown that such growth generates scaling in the steady state, which is consistent with various economic models that explain how systems grow through agglomeration. A consequence of all this is that many physical (geometric) and functional (economic) explanations are converging (*11*, *12*). The volume of work is now so extensive that a wide variety of size distributions are now known to show scaling (*13*). Examples for city populations over 1 million, for cities in the United States with over 100,000 people, and for the 200 tallest buildings in the world are shown in Fig. 1A.

There are still many puzzles associated with such scaling. Gibrat's law assumes that not only are growth rates random but so is their variance, yet there is now considerable evidence that such rates and their variances scale with size (*14*, *15*). Despite agglomeration effects that relate to size, there is a strong suspicion that the best places to locate new growth are in smaller rather than larger cities, reflecting the tradeoff between economies of scale and congestion, which both increase as cities get bigger. The implications are controversial. The age-old question of what the "optimal" size for a city is is as open as it has ever been.

Interactions, Networks, and Densities

Where the focus is on interactions between cities in terms of trade or migration, and within

cities in terms of commuting, shopping, and other social movements, scaling has recently been discovered with respect to such networks. In the past, the focus was almost entirely on modeling traffic flows rather than on the properties of such networks per se (*16*), although the distribution of traffic volumes originating from or destined for different locations in a city has long been known to be scaling (Fig. 1B). Density distributions are also essential outcomes from urban economic models where the focus is on the tradeoff between travel cost or distance and the cost of space, as in rent, house prices, and land values (*17*). These distributions generate an approximate scaling against distance from an established center shown for London in Fig. 1C. As yet, there are no integrated theories tying these ideas together in an economic framework consistent with physical scaling, although progress is being made (*18*). Nor are there any serious uses of such theory to determine ways in which realistic city plans might be devised, although many land-use–transportation models that incorporate such ideas are being used to evaluate the feasibility of new urban plans (*19*). After 40 years of effort, their use is hardly routine but this is still progress.

With the growth of network science (*20*), the focus has been on physical infrastructures, such as the topology and geometry of street and rail systems. These systems are characterized by scale-free activity at the nodes as measured by their number of connections, for example, but it is now clear that this type of scaling is also reflected in traffic volumes at nodes as we imply in Fig. 1B. Much of the work in network science to date has been on classifying network topologies into various shapes of graphs through their statistical properties. Where it is being applied, it is being used to inform the way in which people and vehicular traffic move at quite fine spatial scales, such as in pedestrian densities and dynamics in street networks, which show similar scaling to city size (*21*, *22*). Because network science is not rooted primarily in Euclidean space but deals as much with topologies, such as social networks, this suggests ways in which our longstanding physical approach to cities can be consistently linked to urban economic and social functions that only obliquely manifest themselves in geographical or physical terms. Interesting and useful insights about connectivity and inequality that reflect new ideas about how close or how segregated and congested people are in cities are being discovered (*23*). All this is essential to understanding how information flows both replace and complement material flows of resources that have underpinned the spatial organization of cities hitherto.

Urban Geometry and Morphology

City morphology is reflected in a hierarchy of different subcenters or clusters across many scales, from the entire city to neighborhoods, organized around key economic functions. These in turn reflect the resources needed to service them and the spatial range over which their demand is sustainable. Cities are thus classic examples of fractals in that their form reflects a statistical self-similarity or hierarchy of clusters (*24*). Large cities often develop as existing towns coalesce, with new edge cities being developed on their periphery as they change in scale. The way such fractal growth occurs has been likened to various physical growth processes ranging from percolation to diffusion-limited aggregation (*25*). These map onto the more established notions of density decay with respect to distance in cities from their established center. A typical picture for greater London is shown in Fig. 2A.

Presenting this structure in terms of the transportation network in Fig. 2B provides another perspective on fractal structure consistent with scale-free networks. Allometric methods can be used to link the size and shape of living objects to the networks they use to deliver resources to their parts (*26*). West and his colleagues have recently shown that as cities grow in size, physical networks tend to grow more slowly than city size; that is, the physical infrastructure used to move resources around does not increase as fast as the number of such resources, whereas key economic activities such as the number of innovations as measured through financial services, patents, and scientific products increase faster than city size in terms of population (*27*). Thus, big cities appear more attractive to the most productive industries, but it is easier to move resources around in small cities.

Models that simulate fractal structures can be calibrated to real situations and used for future predictions based on simple rules of land development (*28*). But their most effective use is to deconstruct the rules that have been used in the past to design idealized cities (Fig. 2). A typical city plan from Renaissance Italy (Fig. 2C) is a stylized symmetric construction whose fractal structure is highly contrived but could be formally generated by tight rules being placed on the size and shape of development. Ebenezer Howard's "city of tomorrow" (*29*) (Fig. 2D) presented the geometric logic according to which many 20th-century new towns were designed, again implying strict rules of morphological placement with respect to the components that make the town function at different scales. When implemented, most of these idealizations rarely provide the quality of life for their inhabitants that such order anticipates. They are simply too naïve with respect to the workings of the development process, the competition for the use of space that characterizes the contemporary city, and

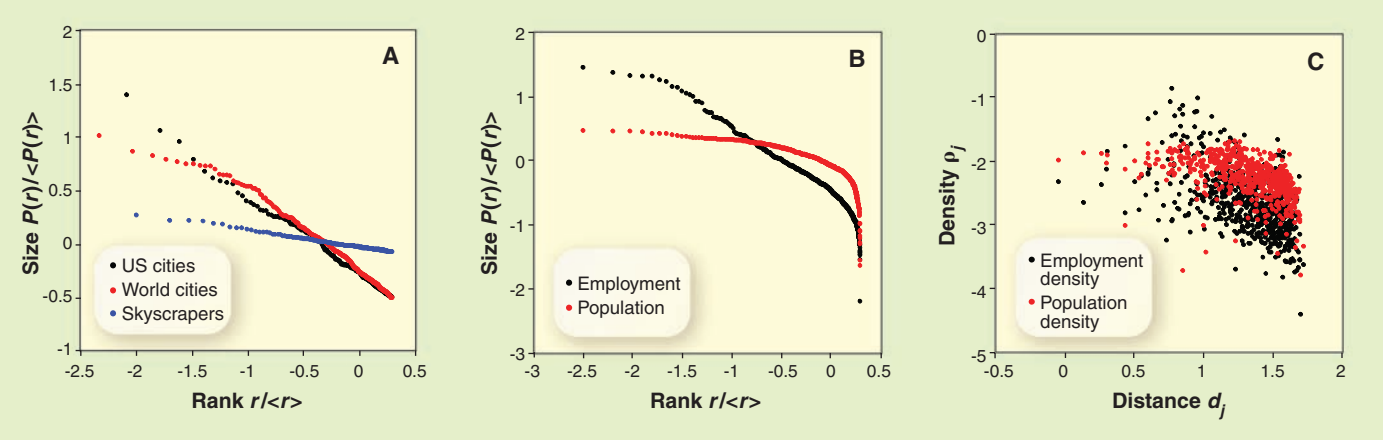


Fig. 1. Scaling in cities. (**A**) City and building size distributions. (**B**) Rank-size scaling in London. (**C**) Density scaling in London. In (A) and (B), vertical axes are populations in rank order from largest to smallest, $P(r)$, normalized by their mean values $<P(r)>$, and horizontal axes are ranks r normalized by their mean values $<r>$. In (C), the vertical axis is population density ρ_j at place j with the horizontal axis, d_j, being distance to j from the center of the metropolis.

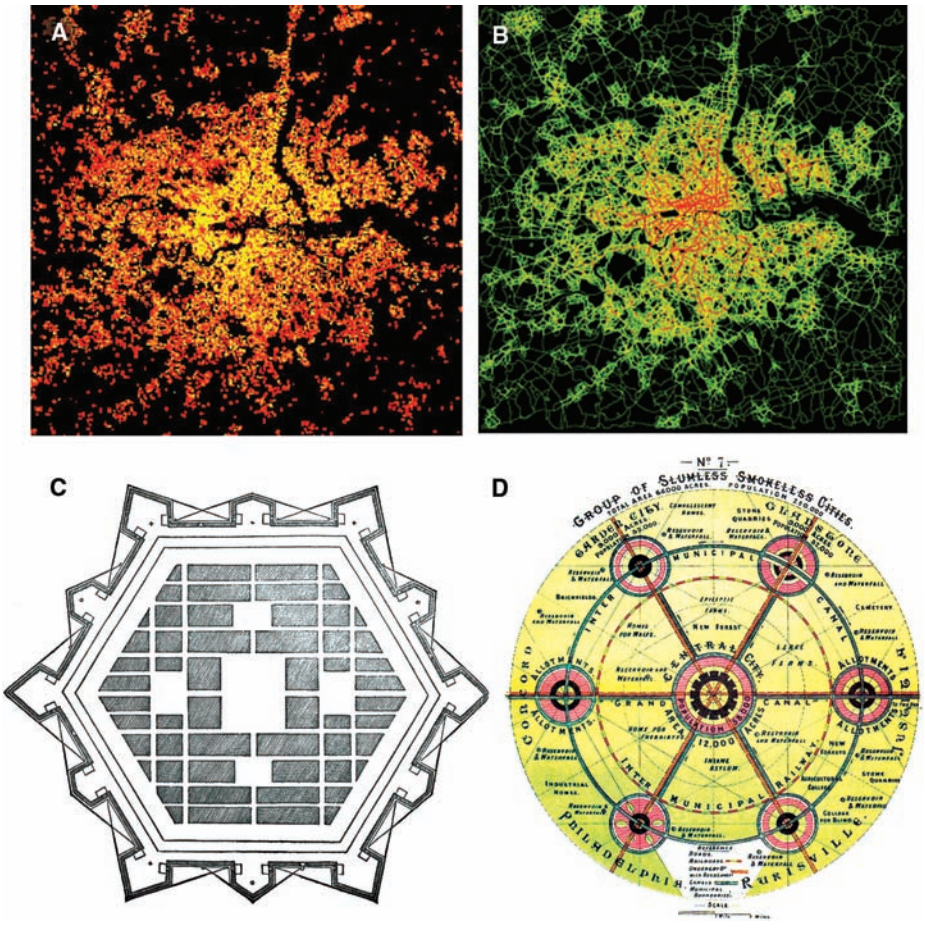


Fig. 2. Fractal cities. (**A**) Population morphology of London. (**B**) The road network in London colored by level of connectivity. (**C**) An idealized geometric city. (**D**) Howard's garden city of tomorrow (*29*).

the degree of diversity and heterogeneity that the most vibrant cities manifest.

A New Science for City Planning?

In the study of cities, there are many competing paradigms. This science has the potential not only to join some of these together but also to improve theories to the point where city planners can develop operational tools grounded in extensive empirical data. In terms of size and scale, we do not yet have a clear view of how big a city is in terms of the density of its activities, the volume of its built and natural space, and the way in which materials, information, and people interact to sustain such forms. We cannot have a clear view of what density means, what energies and costs are incurred by different urban geometries, and how feasible policies are for increasing compactness and managing sprawl until we have good answers to these questions.

The science advocated here has the potential to address these questions. As cities grow in size, they change in shape through allometry and this changes the energy balance used to sustain them. What we are currently learning is that different sizes and shapes of cities imply different geographical advantages, and this again casts doubt on the question of what the ideal size of city should be. Network science provides a way of linking size to the network forms that enable cities to function in different ways. How materials are processed, their resulting waste products and pollution, and their multiplier effects on other urban activities can be tracked using the network dynamics that is implicit in this science, whereas the speed at which change can be initiated through such networks provides essential insights into the potential effectiveness or otherwise of different urban policies. The impacts of climate change, the quest for better economic performance, and the seemingly intractable problems of ethnic segregation and deprivation due to failures in job and housing markets can all be informed by a science that links size to scale and shape through information, material, and social networks that constitute the essential functioning of cities.

We have only just started in earnest to build theories of how cities function as complex systems. We do know, however, that idealized geometric plans produced without any regard to urban functioning are not likely to resolve any of our current urban ills, and this new physics makes us much more aware of the limits of planning. It is likely to lead to a view that as we learn more about the functioning of such complex systems, we will interfere less but in more appropriate ways (*30*). Changes that we propose are then likely to be much more effective in resolving problems than the ways in which city planning has operated in the past. The challenge is to aggressively enrich this science and move it to the point where it can be successfully used to plan better cities. We are but at the beginning.

References and Notes

1. W. Morris, *Architecture, Industry and Wealth: Collected Papers* (Longmans, Green, and Co., London, 1902).
2. J. Jacobs, *The Death and Life of Great American Cities* (Random House, New York, 1961).
3. M. Batty, *Cities and Complexity: Understanding Cities Through Cellular Automata, Agent-Based Models, and Fractals* (MIT Press, Cambridge, MA, 2005).
4. V. Pareto, *Cours d'Economie Politique* (Droz, Geneva, Switzerland, 1896).
5. G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, MA, 1949).
6. W. Christaller, *Die Zentralen Orte in Suddeutschland* (Gustav Fischer, Jena, Germany, 1933).
7. R. Gibrat, *Les Inégalités Économiques* (Librarie du Recueil Sirey, Paris, 1931).
8. H. A. Simon, *Biometrika* **42**, 425 (1955).
9. X. Gabaix, *Q. J. Econ.* **114**, 739 (1999).
10. A. Blank, S. Solomon, *Physica A* **287**, 279 (2000).
11. G. Duranton, *Am. Econ. Rev.* **97**, 197 (2007).
12. J. Eeckhout, *Am. Econ. Rev.* **94**, 1429 (2004).
13. A. Clauset, C. Rohilla Shalizi, M. Newman, preprint available at http://arxiv.org/abs/0706.1062v1 (2007).
14. M. H. R. Stanley *et al.*, *Nature* **379**, 804 (1996).
15. M. Batty, *Nature* **444**, 592 (2006).
16. A, G. Wilson, *Entropy in Urban and Regional Modelling* (Pion Press, London, 1970).
17. C. Clarke, *J. R. Stat. Soc. Ser. A* **114**, 490 (1951).
18. M. A. Fujita, A. Venables, P. Krugman, *The Spatial Economy: Cities, Regions and International Trade* (MIT Press, Cambridge, MA, 1999).
19. M. Wegener, in *GIS, Spatial Analysis, and Modeling,* D. J. Maguire, M. Batty, M. F. Goodchild, Eds. (ESRI Press, Redlands, CA, 2005), pp. 203–220.
20. M. Newman, A. L. Barabasi, D. J. Watts, *The Structure and Dynamics of Networks* (Princeton Univ. Press, Princeton, NJ, 2005).
21. S. Scellato, A. Cardillo, V. Latora, S. Porta, *Eur. Phys. J. B* **50**, 221 (2006).
22. D. Helbing, L. Buzna, A. Johansson, T. Werner, *Transp. Sci.* **39**, 1 (2005).
23. G. Chowell, J. M. Hyman, S. Eubank, C. Castillo-Chavez, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **68**, 066102 (2003).
24. M. Batty, P. A. Longley, *Fractal Cities: A Geometry of Form and Function* (Academic Press, San Diego, CA, 1994).
25. H. A. Makse, S. Havlin, H. E. Stanley, *Nature* **377**, 608 (1995).
26. G. B. West, J. H. Brown, B. J. Enquist, *Science* **284**, 1677 (1999).
27. L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G. B. West, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7301 (2007).
28. I. Benenson, P. M. Torrens, *Geosimulation: Automata-Based Modeling of Urban Phenomena* (Wiley, London, 2004).
29. E. Howard, *To-Morrow: A Peaceful Path to Real Reform* (Routledge, London, 1898; new ed. 2003).
30. P. W. Anderson, *Science* **177**, 393 (1972).
31. The author thanks S. Marshall and D. Smith of University College London for help with Figs. 2C and 2B, respectively.

10.1126/science.1151419

REVIEW

Urbanization and the Wealth of Nations

David E. Bloom,* David Canning, Günther Fink

The proportion of a country's population living in urban areas is highly correlated with its level of income. Urban areas offer economies of scale and richer market structures, and there is strong evidence that workers in urban areas are individually more productive, and earn more, than rural workers. However, rapid urbanization is also associated with crowding, environmental degradation, and other impediments to productivity. Overall, we find no evidence that the level of urbanization affects the rate of economic growth. Our findings weaken the rationale for either encouraging or discouraging urbanization as part of a strategy for economic growth.

According to United Nations (UN) projections, more than half of the world's population will live in urban areas by the end of 2008 (*1*). The continued increase in the share of the population living in urban areas in recent decades has been welcomed by many economists, who view urbanization as a positive achievement on the path toward wealth and prosperity. According to this view, urbanization underpins and enhances economic growth and therefore increases the wealth of nations in the long run (*2*).

Urbanization is a complex phenomenon. Under favorable conditions, initially very small settlements can rapidly develop, first into small towns and then into cities as populations grow and new economic and political structures emerge. Successful sectors within the city will attract further investment, generate increased demand for labor, and trigger migration to the city as a further mechanism of urban growth. With a better quality of life, however, cities may become a major attractor for poor rural populations, leading to large urban unemployment, poverty, and, in many cases, also increased urban violence, congestion, and environmental degradation. The growth of urban areas has promoted concentrations of land, water, and air pollution (*3*) and is associated with the formation of large and rapidly growing slum populations in and around many major cities. According to the UN, more than 1 billion people, or about 14 percent of the total global population, lived in areas classified as slums in 2005 (*4*). These squalid settlements have come to be known for their hopelessness, their atmosphere of fear, and the social exclusion of their residents, and they are hardly a symbol of economic progress.

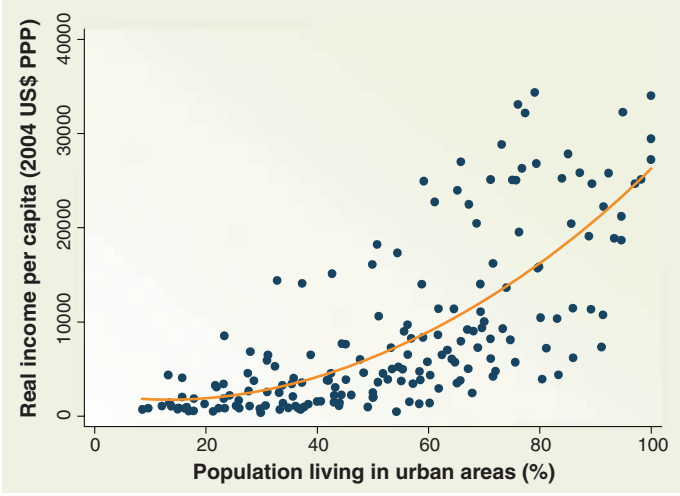


Fig. 1. Income per capita and urban population shares for 180 countries with available data in 2000. The income variable is gross domestic product (GDP) per capita at purchasing power parity (PPP) (*35*). The percentage of people living in urban areas comes from (*36*). These are the data we use throughout the paper. The main source of UN urban population share data is decennial censuses. The numbers rely on country-specific definitions, which are not always comparable across countries and not necessarily consistent across time (*32*, *37*). Nevertheless, the picture looks virtually identical when geography-based data such as those generated by the Global Rural Urban Mapping Project (*38*) or alternative measures such as the urban/rural ratio are used (*37*).

Governments have often undertaken active policies affecting the urbanization process. On the one hand, policy-makers can inhibit urbanization by providing support for rural populations in the form of agricultural subsidies or directly target rural populations by rural transfers or infrastructure projects. On the other hand, governments have historically often displayed large biases toward urban areas and urban populations in the developing world (*5*, *6*) at the expense of the rural population. Because of the political pressure exerted by urban dwellers, central governments have encouraged urbanization by keeping agricultural prices low, by direct investment in urban industries, and by a more generous provision of public services such as health and education (*7*).

We show here that there is no empirical evidence of a causal effect of the level of urban population share on the pace of economic growth. Although the agglomeration of diffuse populations into urban areas will generally increase output per capita, very large or rapidly growing urban areas can have offsetting negative effects through crowding and environmental degradation, and by overwhelming city administrations' capacities. Policies regarding urban development should weigh carefully the positive and negative spillovers of urbanization, without a presumption that urbanization is a policy for promoting economic growth.

The Effects of Urbanization

The economics literature is replete with references to urbanization as a natural concomitant of modernization and industrialization (*2*, *8*). Cities as locations of concentrated economic activities offer large and diversified labor pools and are closer to customers and suppliers, (*9*). Cities also offer increased opportunities for division of labor and make intra-industry specialization more likely (*10*). Firms not only may profit from horizontal and vertical spillovers but also are able to respond to market demand changes more effectively (*11*–*14*). Relatively cheaper transport combines with the proximity to customers and suppliers to reduce the costs of trade (*15*, *16*). Moreover, by aggregating educated and creative people in one place, cities incubate new ideas and technologies and may lead to efficient growth by potentiating the full social returns to increased human capital (*17*, *18*). Although these models highlight the advantages of urbanization, the implied optimal level of urban population share is unclear. As urban areas become more densely populated, changes in urban population share may have no effect on productivity growth.

Figure 1 shows the positive cross-country relationship between the level of income and the urban population share in 2000. The correlation between real income per capita and the fraction of the population in urban areas is 0.8. The theoretical arguments for urbanization as a source of productivity gains and the strong correlation between income and urban population share have been the foundation of a literature promoting urban growth as the path toward wealth and development (*19*) and have led to the perception of a

Harvard School of Public Health, Harvard University, 665 Huntington Avenue, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: dbloom@hsph.harvard.edu

strong, positive effect of urbanization on economic growth (*2*).

The Effect of Urban Population Share on Income Levels

Although the cross-country scatter plot of income and urban population share is striking, it provides little insight into the causal relation between urban population share and economic development. At issue is whether urbanization promotes economic well-being. That is, would raising the level of urban population share in a country promote economic growth and a higher level of income? The relationship in Fig. 1 could be due simply to an effect of the income level on the level of urban population share, with no feedback from urban population share to the level of income. Alternatively, it could be due to a common factor that affects both urban population share and income, without either of them affecting the other.

Two previous studies have failed to find a link between the share of the population urbanized and economic growth (*20*, *21*). We investigate the question in more detail using a number of new approaches and argue that the result is robust.

Our first approach is to investigate the effects of urban population share on income levels across countries for 1960 to 2000. The data set is configured into four 10-year intervals to limit the influence of short-term economic cycles and to avoid reliance on interpolated urban population share data. We carry out a Granger causality test between income per capita and urban population share (*22*). Urban population share is said to "Granger-cause" income if, controlling for past income levels, past values of urban population share are predictive of future values of income. This is a low-threshold test insofar as it does not require direct evidence of a structural, causal link from urban population share to income; it merely requires that urbanization has some predictive power for future income levels.

The results of the Granger causality test are summarized in Table 1 below. We take real income per capita (*23*) as the dependent variable and explain it with 10-year lagged income as well as the 10-year lagged level of urban population share.

Table 2. Long-term effects of urban population share. Robust standard errors in parentheses.

	Dependent variable: Average annual growth 1970 to 2000 full sample		Dependent variable: Average annual growth 1970 to 2000 low-income countries	
	(1)	(2)	(3)	(4)
East Asia dummy variable†	1.609**	1.606**	5.075***	4.725*
	(0.68)	(0.67)	(1.29)	(2.23)
Primary schooling 1970	0.00363	0.00298	0.0438	0.0480
	(0.012)	(0.012)	(0.034)	(0.043)
Investment price 1970	–0.00483***	–0.00486***	–0.00405***	–0.00414**
	(0.0013)	(0.0013)	(0.0011)	(0.0015)
Log (GDP) 1970	–1.669***	–1.560***	–3.424**	–2.529
	(0.30)	(0.55)	(1.31)	(3.50)
Fraction of area tropical	–0.787*	–0.766*	–0.473	–0.316
	(0.42)	(0.44)	(1.68)	(1.90)
Coastal density	0.000657*	0.000485	0.0120*	0.0123
	(0.00038)	(0.00048)	(0.0060)	(0.0081)
Fraction Confucian‡	2.072*	2.249**		
	(1.08)	(1.12)		
Life expectancy 1970	0.181***	0.184***	0.0780	0.0666
	(0.059)	(0.061)	(0.084)	(0.10)
Africa dummy variable†	–0.0647	–0.0888	–0.0800	–0.183
	(0.61)	(0.68)	(1.30)	(1.63)
Urban population share 1970	–0.00950	–0.00664	–0.0379	0.312
	(0.011)	(0.059)	(0.038)	(1.56)
Urban population share 1970 squared		0.000204		0.00141
		(0.00041)		(0.0056)
Urban population share x Log(GDP)1970		–0.00257		–0.0569
		(0.0094)		(0.22)
Constant§	5.284**	4.608	20.84**	15.33
	(2.36)	(3.54)	(8.82)	(21.6)
P value for F test of joint significance of urban share terms		0.857		0.794
Sample	All countries	All countries	Low-income	Low-income
Observations (no.)	86	86	24	24
R^2	0.65	0.65	0.73	0.74

†The Africa and East Asia dummy variables are indicators that equal 1 if the country is in respective region and 0 otherwise. The inclusion of these indicators is based on the hypothesis that all countries in a geographic area share some region-specific characteristics that affect economic growth. ‡Following Barro (*27*), Fraction Confucian is the proportion of the population indicating Confucianism as primary religion. §The constant represents the intercept of the fitted regression line. *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$

We also add fixed effects that control for country-specific factors that are constant over time, and time dummies to allow for a changing relationship over time. As shown in Table 1, once we control for the lagged income level, the lagged urban population share is not a statistically significant predictor of future income levels. Knowledge of a country's level of urban population share does not appear to help us predict its future income level.

The results of the Granger causality tests have to be interpreted with caution. Despite the fact that our specification incorporates time and country fixed effects (*24*), we may be excluding a number of other variables that affect economic growth and thus may obtain biased estimates. In addition, 10 years may be too short a time horizon for an economy to capture the benefits of urbanization (*25*).

To address this issue, we investigate whether the initial level of urban population share in 1970 affects the rate of economic growth over the period 1970 to 2000. In this cross-sectional approach, we have to control for other factors that may influence growth. Rather than specify our own ad hoc model of growth, we use a specification based on the work of Sala-i-Martin, Doppelhofer, and Miller (*26*) (hereafter SDM), who examined 67 potential explanatory variables, not including urban population share, as possible explanations for economic growth, and identified those variables with the highest posterior probability (via Bayesian updating) given

Table 1. Granger causality test. Country and time fixed effects are included in all specifications. Estimation results are based on the bias-corrected least-square dummy variable estimator developed by Bruno (*39*). Bootstrapped standard errors in parentheses.

	Dependent variable
	Log GDP per capita
Urban population share	–0.0473
(10 years previously)	(0.062)
Log GDP per capita	0.712*
(10 years previously)	(0.063)
Observations	561
Number of countries	163

*$P < 0.01$

the evidence. In column 1 of Table 2, we show the results for a model with their nine best predictors, plus the initial fraction of the population living in urban areas. We estimate a negative, but statistically insignificant, effect of urban population share. In column 2, we allow for a nonlinear relationship between income and urban population share (using a squared term) and for the optimal level of urbanization to depend on the level of income (using an interaction term between income and urban population share). The results in column 2 are similar to those in column 1: None of the terms that include urban population share have a statistically significant effect on future economic growth, either individually or jointly.

We wish to assess the robustness of these growth regression results. There are two factors that might change the results. First, low-income countries may be different from other countries; and, second, changing the set of covariates used in the growth regression may change the estimated effect of urban population share. In columns 3 and 4 of Table 2, we repeat the growth regression exercise for low-income countries as defined by the World Bank. This gives somewhat different estimates and has less precision because the sample size is smaller, but again we find no effect of the level of urban population share on economic growth.

In Table 3 (top), we report the coefficient of urban population share for a number of regressions with different sets of covariates. We start with just urban population share and initial income, a specification similar to that used in Table 1, but now without country fixed effects. We then examine the effect of adding sequentially the 5, 9, and 16 covariates that provide the best prediction of economic growth as found by SDM (*26*). In none of these regressions is urban population share statistically significant. In Table 3 (bottom), we repeat the exercise for low- and middle-income countries and find the same lack of significance.

Overall, our results imply that countries with a higher degree of initial urban population share do not experience any faster or slower economic growth than countries with a low degree of initial urban population share. Moreover, this result appears quite robust with respect to the specification of the growth model. Taken together, none of our empirical tests provide support for the view that urban population share has a causal effect on the level of income. Although urbanization is part of the process of economic development, it does not appear to have an independent influence on economic growth (*27*).

Different Types of Urbanization

Understanding the process of rural-to-urban migration is central to understanding the effects of urbanization. There are several distinct channels through which urbanization can occur. The first involves the movement of people from rural to urban areas. Empirically, migration is estimated to contribute on average between 40 and 50% of total urban population growth (*28*, *29*). Second, the rate of natural population increase may be different in urban and rural areas as a result of differences in birth- and death rates. Birthrates are usually lower in urban than rural settings, but mortality rates are often lower as well. The third important channel is the reclassification of rural settlements as urban as a result of rural population growth and increasing population density; a person can become "urbanized" while standing still.

Although migration can be triggered by increased demand for labor in urban areas, blooming economies are not the only cause of urban migration. Political instability within countries has led to major refugee flows toward cities like Kinshasa (Democratic Republic of the Congo) and Karachi (Pakistan). Natural disasters, such as extended droughts and floods, have destroyed the economic basis of rural life in several regions in the developing world and induced major population flows toward the cities in countries like Angola, Ethiopia, and Mauritania (*30*). Even in the absence of natural disasters, economic conditions in the countryside—sometimes brought about by policies that favor city dwellers—can push rural residents to cities that, in the long run, offer them little in the way of economic sustenance. In such cases, cities may grow rapidly in population without being economic "success stories."

One possible explanation for the absence of an empirical link running from urban population share to income per capita lies in the different types of urbanization observed across countries or continents. Although the share of the population living in urban areas has risen from slightly below 20% to a level around 36% in both Asia and Africa between 1960 and 2000 (Fig. 2), per capita income has increased 340% in Asia compared with only 50% in Africa. If the initial level of urbanization were the key factor in economic growth, we would have expected both regions to grow at about the same rate because they had about the same urban population share. If we thought that urban population share simply follows the level of income, Asia should have urbanized much more quickly than Africa given its superior growth performance; but, again, this was not observed. Rather, it appears that urbanization in Asia has been driven mostly by industrialization and a plethora of job opportunities in urban areas (*31*), whereas urbanization in Africa seems to be more the result of population pressure, civil conflict, and changing political regimes (*32*), as well as ethnic tensions and a momentum effect (*33*).

The fact that rapid urbanization has gone hand in hand with economic growth in Asia, while proceeding just as rapidly but without growth in Africa, is perhaps surprising. However, it is a common pattern that, over long periods of time, development indicators can move in quite different directions (*25*). Development is a multifaceted process, and economic growth, or the absence thereof, is not a strong indicator of progress in other dimensions of development.

Although the share of the total population living in urban areas seems to have little effect on economic growth, it may be that other measures of urban composition do matter. Primacy, the portion of the urban population living in the largest city, or urban concentration, the proportion of urban dwellers living in large cities, may affect economic growth, although the size of these effects appears to be highly nonlinear (*20*, *21*). Primacy and urban concentration are measures of the distribution of the urbanized population between small and large cities, and in

Table 3. Robustness checks table. Robust standard errors in parentheses. No values were statistically significant.

	Full sample			
	Dependent variable: Average annual growth 1970 to 2000			
Model	Lagged GDP only	SDM 5	SDM 9	SDM 16
Urban population share 1970	0.00917	0.00621	−0.0130	−0.0105
	(0.016)	(0.014)	(0.011)	(0.013)
Observations (no.)	78	78	78	78
R-squared	0.04	0.55	0.68	0.74
	Low- and middle-income countries only			
	Dependent variable: Average annual growth 1970 to 2000			
Model	Lagged GDP only	SDM 5	SDM 9	SDM 16
Urban population share 1970	0.00179	−0.0144	−0.0231	−0.0266
	(0.021)	(0.021)	(0.019)	(0.020)
Observations (no.)	55	55	55	55
R-squared	0.00	0.52	0.66	0.73

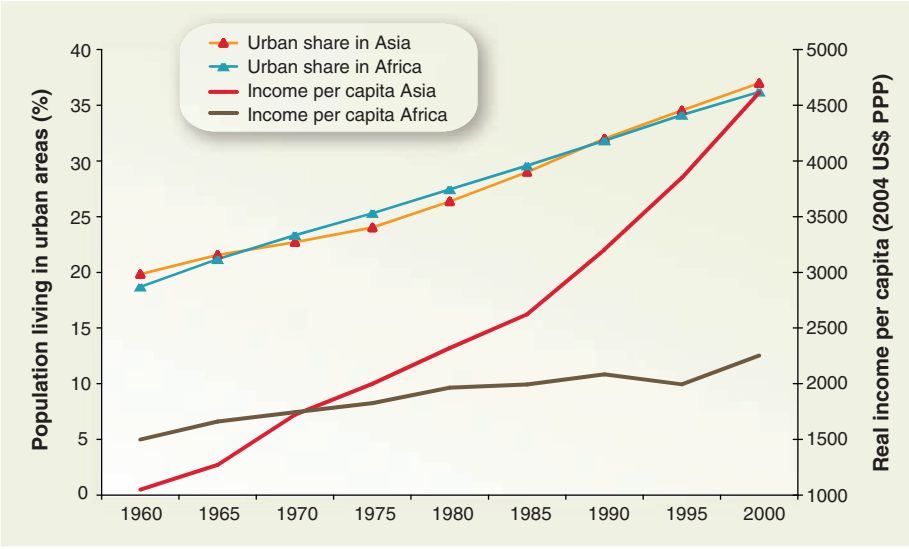


Fig. 2. Average per capita income and average urban shares for African and Asian countries over the period 1960 to 2000.

practice they have a very low correlation with the level of urban population share of the overall population that we investigate here.

The Bottom Line

Urban development is important; with a majority of the world population living in urban areas, well-functioning cities will be one of the key factors in the ongoing battle against poverty (*34*). However, it appears that urbanization is more an indicator than an instrument of economic development. Our findings imply that policies specifically aimed at accelerating, or retarding, urbanization are unlikely to speed up economic development. Even though regional urban areas and success stories like the Silicon Valley or Bangalore are key drivers of economic growth, the notion that a larger fraction of a country's population living in urban areas improves economic performance does not seem to have empirical support. Policy-makers who hope to increase the long-term economic growth of their countries by supporting, or inhibiting, urbanization are likely to miss their target.

References and Notes

1. United Nations, *World Population Prospects: The 2004 Revision. CD-ROM Edition - Extended Dataset* (United Nations, New York, 2005).
2. C. M. Becker, in *International Handbook of Development Economics*, A. Dutt, J. Ros, Eds. (Edward Elgar Publishing, Northampton, MA, 2008).
3. United Nations Human Settlements Programme (UN-HABITAT)/DFID, *UN-HABITAT/DFID Nairobi* (2002).
4. UN-HABITAT, *State of the World's Cities 2006/7* (Earthscan Publications Ltd., London, 2007).
5. M. Lipton, *Why Poor People Stay Poor: Urban Bias in World Development* (Temple Smith, London, 1977).
6. M. Lipton, *J. Dev. Stud.* **29**, 229 (1993).
7. M. Lipton, *J. Dev. Stud.* **20**, 139 (1984).
8. Y. W. Bradshaw, E. Fraser, *Am. Sociol. Rev.* **54**, 986 (1989).
9. A. Ciccone, R. Hall, *Am. Econ. Rev.* **86**, 54 (1996).
10. R. Becker, V. Henderson, in *Economics of Cities: Theoretical Perspectives*, J.-M. Huriot, J.-F. Thisse, Eds. (Cambridge University Press, Cambridge, 2000).
11. H. Abdel-Rahman, M. Fujita, *J. Reg. Sci.* **30**, 165 (1990).
12. A. K. Dixit, J. E. Stiglitz, *Am. Econ. Rev.* **67**, 297 (1977).
13. M. Fujita, H. Ogawa, *Paper Prepared for Prince Bertil Symposium on the Dynamic Firm, Stockholm* (1982).
14. J. E. Rauch, *J. Urban Econ.* **34**, 380 (1993).
15. A. Caplin, J. Leahy, *Econ. J.* **108**, 62 (1998).
16. P. Krugman, *J. Polit. Econ.* **99**, 483 (1991).
17. D. Black, J. V. Henderson, *J. Polit. Econ.* **107**, 252 (1999).
18. M. R. Montgomery, R. Stren, B. Cohen, H. E. Reed, Eds., *Cities Transformed: Demographic Change and Its Implications in the Developing World*, vol. 273 (National Academies Press, Washington, DC, 2003).
19. J. G. Williamson, *Econ. Dev. Cult. Change* **13**, 1 (1965).
20. L. Bertinelli, E. Strobl, *CREDIT Research Paper* **03/14**, Centre for Research in Economic Development and International Trade, University of Nottingham, UK, (2003); http://papers.ssrn.com/sol3/papers.cfm?abstract_id=464202.
21. J. V. Henderson, *J. Econ. Growth* 8, 47 (2003).
22. C. W. J. Granger, *Econometrica* **37**, 424 (1969).
23. Because the relationship between urbanization and income is nonlinear, we use the natural logarithm of income per capita in our empirical specification, which does result in a linear relationship. Note that the change over time in the natural logarithm of income is the rate of economic growth.
24. The inclusion of country fixed effects implies that the empirical analysis abstracts from all factors that are constant within a given country over the entire sample period. Similarly, the inclusion of time fixed effects implies that the empirical analysis abstracts from all factors that affect all countries equally in a given year.
25. W. Easterly, *J. Econ. Growth* **4**, 239 (1999).
26. X. Sala-i-Martin, G. Doppelhofer, R. I. Miller, *Am. Econ. Rev.* **94**, 813 (2004).
27. In addition to urban population share, we also considered primacy: the fraction of the urban population living in the country's most populated city. The inclusion of the primacy variable did not change any of our empirical results.
28. N. Keyfitz, *Geogr. Anal.* **12**, 142 (1980).
29. S. H. Preston, *Popul. Dev. Rev.* **5**, 195 (1979).
30. United Nations, Office for the Coordination of Humanitarian Affairs, Reliefweb (2007); www.reliefweb.int.
31. G. Hugo, in *Africa on the Move: African Migration and Urbanisation in Comparative Perspective*, M. Tienda, S. Findley, S. Tollman, E. P. Whyte, Eds. (Wits University Press, Johannesburg, 2006).
32. D. Satterthwaite, *IIED Human Settlements Discussion Paper Urban Change* **4** (2007).
33. M. Fay, C. Opal, *Policy Research Working Paper No. 2412* (World Bank, Washington, DC, 2000); http://papers.ssrn.com/sol3/papers.cfm?abstract_id=632483.
34. UNFPA, *State of the World Population 2007,* UNFPA 45 (United Nations Population Fund, New York, 2007); www.unfpa.org/swp/2007/english/introduction.html.
35. A. Heston, R. Summers, B. Aten, Penn World Table Version 6.2, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania (2006); http://pwt.econ.upenn.edu/php_site/pwt_index.php.
36. United Nations, *World Urbanization Prospects: The 2005 Revision Population Database* (United Nations, New York, 2006).
37. D. E. Bloom, D. Canning, G. Fink, T. Khanna, P. Salyer, paper prepared for the World Institute for Development Economics Research of the United Nations University, Project Workshop: Beyond the Tipping Point: Development in an Urban World, hosted by the London School of Economics and Political Science, 18 to 20 October 2007; available as PGDA working paper no. 29 at www.hsph.harvard.edu/pgda/working.htm.
38. Socioeconomic Data and Applications Center, Center for International Earth Science Information Network (CIESIN), Columbia University (2007), vol. 2007; http://sedac.ciesin.columbia.edu/gpw.
39. G. S. F. Bruno, *Econ. Lett.* **87**, 361 (2005).

40. The authors are grateful to M. Montgomery, L. Rosenberg, and two anonymous referees for their comments and suggestions.

10.1126/science.1153057

Experienced Saxophonists Learn to Tune Their Vocal Tracts

Jer Ming Chen, John Smith, Joe Wolfe*

How much do the acoustics of the vocal tract influence performance on single-reed instruments (clarinets and saxophones)? To summarize a 25-year debate, we note that scientists' opinions have ranged from "negligible" (*1*) to "vocal tract resonant frequencies must match the frequency of the required notes" (*2*). Musicians' opinions are also diverse. The longevity of the debate is due to the difficulty of making nonperturbing, precise acoustical measurements inside the mouth during playing (*3*), that is, in a variable, humid environment with very loud sounds.

The notes played on wind instruments depend on the acoustic impedance, Z: the ratio of sound pressure to the oscillatory component of air flow. Reed instruments usually play at frequencies very near a high, sharp peak in the impedance spectrum of the bore of the instrument. In a standard simple model (*4*), the impedance of the vocal tract, Z_t, is in series with that of the bore, Z_b. For most playing conditions, Z_b is much larger than Z_t (e.g., Fig. 1A), so acousticians usually neglect the role of the tract.

In saxophones and clarinets, a single reed vibrates to modulate the flow into the instrument and is in turn driven to vibrate by standing waves in the bore of the instrument. We studied the tenor saxophone, whose mouthpiece is large enough to support devices that we constructed for measuring Z_t with minimal perturbation to the player. Five professional and three amateur saxophonists sustained notes for several seconds while we measured Z_t by using an acoustic signal comprising 224 sine waves (*5*–*7*).

In Fig. 1, (A) and (B) show typical experimental data. The dark line is Z_t. Upon this are superimposed narrow spikes, which are harmonics of the note being played. The pale line shows Z_b for that note.

Over the lower-frequency range of the instrument, the peaks in Z_b are much greater than those in Z_t. Also in this range, the peaks in Z_t varied greatly among players and showed no consistent relation to the note played (Fig. 1, A and C). In the high (altissimo) range, however, the professional players consistently tuned a strong peak in Z_t near to or slightly above the fundamental of the note played (Fig. 1C).

The amateur players, who did not tune a strong peak in Z_t, were unable to play in the altissimo range. A configuration of keys pressed (a fingering) produces a Z_b with several peaks, whose magnitudes decrease at high frequency. In the normal range, the peak of Z_b at the appropriate frequency is large and dominates the series combination ($Z_b + Z_t$). In the high range, the relevant peak in Z_b is weak (Fig. 1, A and B, are representative). Here, experienced players use a peak in Z_t, which may be several times greater than that in Z_b, to choose the playing frequency. For instance, experienced players can hold a single fingering and play a different note at each of the first dozen peaks in Z_b by tuning Z_t.

We conclude that the vocal tract resonances have only modest effects on the sounding pitch over much of the instrument's range. However, to play notes in the altissimo range, players learn to tune a resonance of the tract near to the note to be played. Although demonstrated only for saxophone, similar effects are likely to be important in other single- and double-reed instruments, whose players also report the importance of the tract for special effects, including high-register playing.

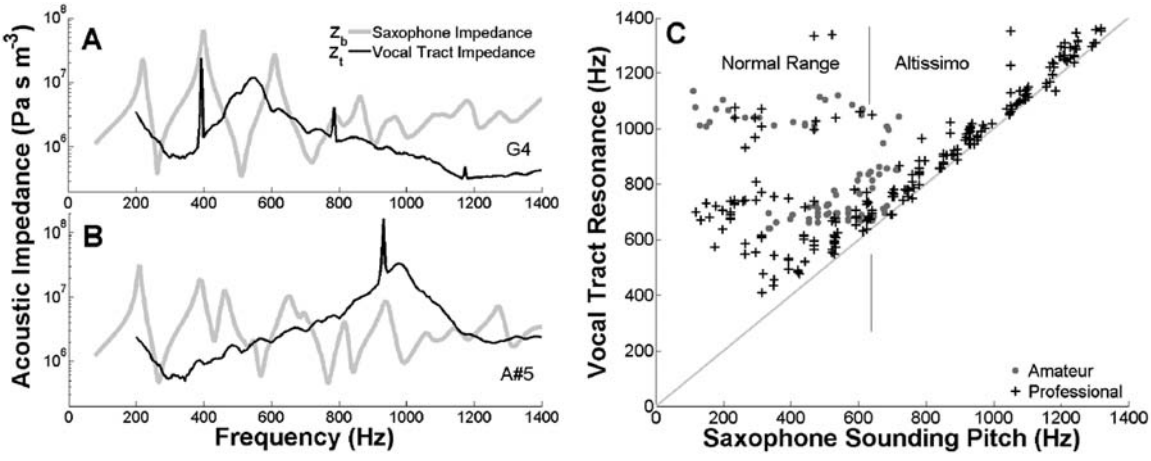


Fig. 1. Representative examples of acoustic impedances of the tract (Z_t) for a professional player (**A**) for the note G4 (near 400 Hz, in the normal range) and (**B**) for the note A#5 (near 940 Hz, altissimo range) (sound files in supporting material). The narrow peaks are harmonics of the notes played, the broad peaks are tract resonances. Z_b for the saxophone for each note is shown with a pale line. (**C**) The frequencies of the relevant resonance in the vocal tract are plotted against the frequency of the note played. The diagonal line shows the equation: tract resonance frequency = pitch frequency.

References and Notes

1. J. Backus, *J. Acoust. Soc. Am.* **78**, 17 (1985).
2. P. G. Clinch, G. J. Troup, L. Harris, *Acustica* **50**, 280 (1982).
3. C. Fritz, J. Wolfe, *J. Acoust. Soc. Am.* **118**, 3306 (2005).
4. A. Benade, in *Vocal Fold Physiology, Biomechanics, Acoustics, and Phonatory Control*, I. R. Titze, R. C. Scherer, Eds. (Denver Center for the Performing Arts, Denver, CO, 1985), p. 425.
5. J. Wolfe, J. Smith, J. Tann, N. H. Fletcher, *J. Sound Vib.* **243**, 127 (2001).
6. P. Dickens, J. Smith, J. Wolfe, *J. Acoust. Soc. Am.* **121**, 1471 (2007).
7. Materials and methods are available on *Science* Online.

8. We thank the Australian Research Council for support, the Yamaha Corporation of Japan for the saxophone, and our participants.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5864/776/DC1
Materials and Methods
Fig. S1
References
Audio S1 to S4

5 October 2007; accepted 26 November 2007
10.1126/science.1151411

School of Physics, University of New South Wales, Sydney, NSW 2052, Australia.

*To whom correspondence should be addressed. E-mail: J.Wolfe@unsw.edu.au

Innate Immune Homeostasis by the Homeobox Gene *Caudal* and Commensal-Gut Mutualism in *Drosophila*

Ji-Hwan Ryu,[1]* Sung-Hee Kim,[1]* Hyo-Young Lee,[1,2] Jin Young Bai,[1] Young-Do Nam,[3] Jin-Woo Bae,[3] Dong Gun Lee,[4] Seung Chul Shin,[1,5] Eun-Mi Ha,[1] Won-Jae Lee[1]†

Although commensalism with gut microbiota exists in all metazoans, the host factors that maintain this homeostatic relationship remain largely unknown. We show that the intestinal homeobox gene *Caudal* regulates the commensal-gut mutualism by repressing nuclear factor kappa B–dependent antimicrobial peptide genes. Inhibition of *Caudal* expression in flies via RNA interference led to overexpression of antimicrobial peptides, which in turn altered the commensal population within the intestine. In particular, the dominance of one gut microbe, *Gluconobacter* sp. strain EW707, eventually led to gut cell apoptosis and host mortality. However, restoration of a healthy microbiota community and normal host survival in the *Caudal-RNAi* flies was achieved by reintroduction of the *Caudal* gene. These results reveal that a specific genetic deficiency within a host can profoundly influence the gut commensal microbial community and host physiology.

The mucosal epithelia of all metazoans, such as those found in the gastrointestinal tract, are in intimate contact with a large number of commensal microbiota (*1*). As a consequence, commensal bacteria are known to influence many aspects of the host gut physiology, including innate immunity, development, and homeostasis (*2*–*7*). However, the lack of a genetically amenable animal model has limited in-depth analyses of gut-microbe interactions in vivo.

Recent studies have shown that the *Drosophila* gut activates host antimicrobial defense through the production of microbicidal reactive oxygen species (ROS) and antimicrobial peptides (AMPs) (*8*–*11*). During most gut-pathogen interactions, intestinal redox homeostasis, mediated via infection-induced ROS generation by the dual oxidase enzyme and subsequent ROS elimination by immune-regulated catalase, is critical for host survival (*8*, *9*). The direct contact between gut epithelia and ingested pathogens also activates the immune deficiency (IMD) pathway and subsequent nuclear localization of the p105-like NF-κB, Relish, which in turn leads to de novo synthesis of diverse immune effectors: AMPs and immunosuppressive enzymes such as the peptidoglycan recognition proteins PGRP-SC and PGRP-LB (*10*, *12*–*16*). Although pathogen-initiated gut immunity is fairly well documented in *Drosophila* (*8*–*10*, *14*, *16*, *17*), the molecular interaction between commensals and the *Drosophila* gut is poorly understood.

Role of gut commensal microbiota in the IMD-Relish pathway. To investigate the molecular mechanism for commensal-gut interactions, we first asked whether the commensal microbiota could elicit host gut immune responses. We produced germ-free wild-type (GF^{WT}) and conventionally reared wild-type (CR^{WT}) animals (fig. S1) to enable examination of the gut IMD-Relish pathway potential in the absence or presence of commensals. A large amount of the nuclear-translocated active form of Relish was detected in the intestine cells in the presence of commensals (CR^{WT} flies) and was even more pronounced after gut infection with the *Drosophila* pathogen *Erwinia carotovora carotovora-15* (*Ecc15*) (Fig. 1A).

The well-characterized constitutive nuclear localization of Relish in the intestines of CR^{WT} flies was almost completely abolished in the absence of commensals in GF^{WT} or in antibiotics-treated CR^{WT} flies. Similarly, it was not seen in flies carrying a mutation in the IMD-Relish pathway (CR^{Dredd} flies) (Fig. 1A). The expression levels of Relish-dependent immunosuppressive enzymes such as PGRP-SC and PGRP-LB (*12*, *13*) were significantly higher in the guts of CR^{WT} flies than in the guts of GF^{WT} or CR^{Dredd} flies (Fig. 1B), which confirms that the IMD-Relish pathway was activated by commensals under normal gut conditions. However, the Relish-dependent immune effector molecules, such as AMP genes including *Cecropin* (*Cec*) and *Diptericin* (*Dpt*), were largely silenced in the CR^{WT} gut despite chronic Relish activation. Similarly, no significant difference in AMP expression levels was observed between the GF^{WT} and CR^{WT} midguts (Fig. 1B). Taken together, these results indicate that although commensal organisms of the gut can chronically induce a high level of local IMD–NF-κB pathway activation, only a subset of target genes (notably excluding AMP genes) are activated.

Role of Caudal in gut AMP repression. We next investigated the potential mechanisms of repression of the gut AMP genes to determine how the selective silencing toward commensals might occur. Currently, the cis-regulatory elements controlling epithelial AMP gene expression are poorly understood. It is known, however, that the κB elements in the promoter regions responsible for nuclear factor kappa B (NF-κB)–Relish binding are essential for inducing epithelial AMP expression (*14*–*16*), whereas the CDREs, responsive elements for the homeobox transcription factor Caudal (Cad) responsible for Cad binding, are found to be critical for constitutive AMP expression in certain types of epithelia, such as salivary glands and the ejaculatory duct (*18*). The homeobox transcription factor Cad was originally identified on the basis of its regulatory role in the anteroposterior body axis formation of the *Drosophila* embryo (*19*, *20*). Because *Cad* expression in postembryonic life is known to be mostly restricted to the intestine (*19*) (fig. S2), we analyzed the contribution of CDREs to gut AMP gene expression in vivo. To accomplish this, we used green fluorescent protein (GFP) reporter–expressing transgenic flies carrying a *Cec* promoter in which the CDREs were mutated ($Cec^{CDRE\text{-}mut}$-*GFP*) and compared GFP expression with that of transgenic flies carrying the wild-type promoters fused to GFP (*Cec-GFP*) (Fig. 1C). We were unable to detect any *Cec* reporter activity in the midgut of *Cec-GFP* flies under normal conditions (Fig. 1C). Interestingly, however, $Cec^{CDRE\text{-}mut}$-*GFP* flies were found to exhibit high constitutive expression of *Cec* reporter activity in the posterior midgut and the proventriculus in the absence of oral infection (Fig. 1C).

To test whether Cad was involved in the negative regulation of *Cec* expression through CDREs, we generated transgenic flies that carried the *UAS-Cad-RNAi* construct to mimic the loss of function (fig. S3) via RNA interference (RNAi). A spontaneous activation of all tested AMPs was observed in the gut of $CR^{Cad\text{-}RNAi}$ flies under conventional conditions without microbial infection, which was not the case in the control flies (Fig. 1D). Furthermore, similar AMP derepression was also observed in *Cad-RNAi* flies with different GAL4 drivers (fig. S4). The role of Cad as a repressor was further confirmed by a strong expression of *Dpt* reporter activity observed in the gut of $CR^{Cad\text{-}RNAi}$ flies carrying the *Dpt-LacZ* reporter (fig. S5). Cad-dependent AMP repression is highly tissue-specific because no AMP derepression was observed in the fat body of $CR^{Cad\text{-}RNAi}$ flies (Fig. 1D). Introduction of *Cad-RNAi* had no effect on the expression of PGRP-SC and PGRP-LB (Fig. 1D), which indicates that any repressive role of Cad is restricted to a distinct

[1]Division of Molecular Life Science, Ewha Woman's University and National Creative Research Initiative Center for Symbiosystem, Seoul 120-750, South Korea. [2]Laboratoire de BBMI, Institut Pasteur, Paris 75724, France. [3]Biological Resource Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-806, South Korea. [4]School of Life Science and Biotechnology, Kyungpook National University, Daegu 702-701, South Korea. [5]Brain Korea 21 Program, Yonsei University College of Medicine, CPO Box 8044, Seoul 120-752, South Korea.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: lwj@ewha.ac.kr

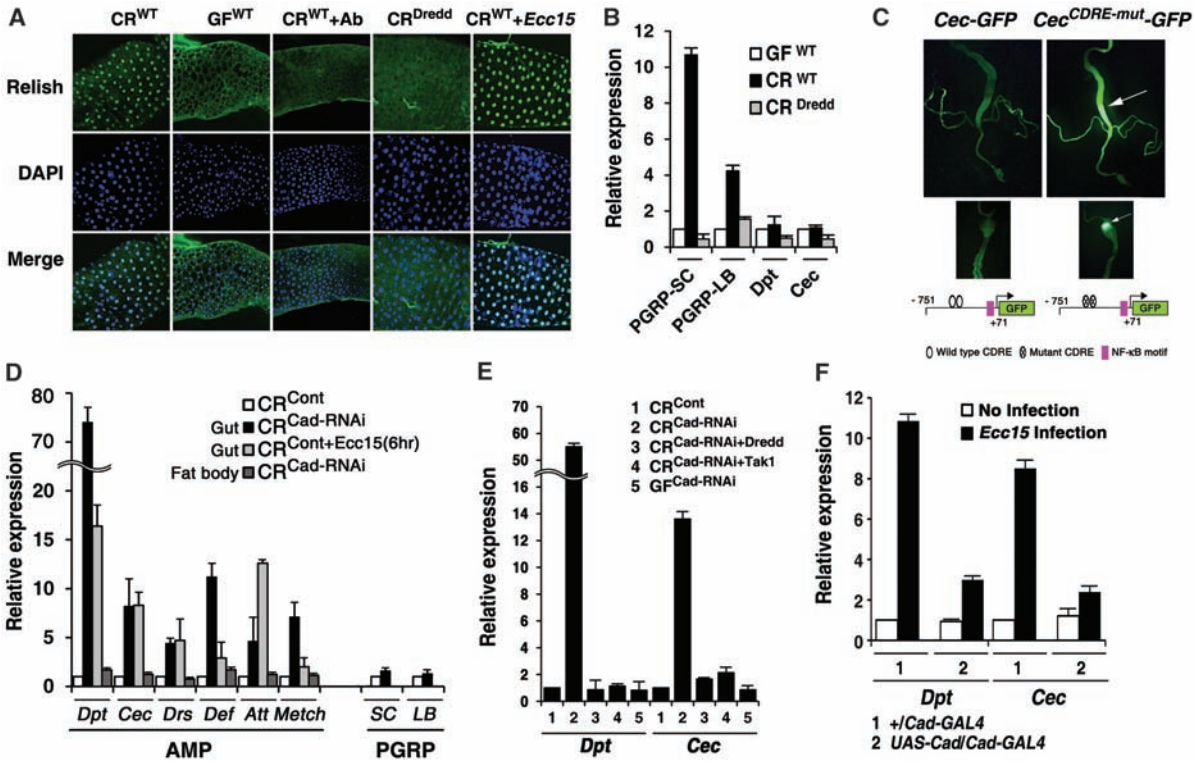


Fig. 1. Caudal acts as a gut-specific repressor for NF-κB–dependent AMP genes. (**A**) Nuclear-translocated active form of Relish (antibody to Relish, green) in posterior midgut (5- or 10-day-old flies). CR^{WT}, conventionally reared wild-type flies; GF^{WT}, germ-free wild-type flies; CR^{WT} + Ab, antibiotics-treated CR^{WT} flies as described in fig. S1; CR^{Dredd}, conventionally reared *Dredd* mutant flies; CR^{WT} + *Ecc15*, natural gut infection with *Ecc15*. Nuclear staining was performed with 4′,6′-diamidino-2-phenylindole (DAPI). (**B**) Quantitative real-time PCR analysis of PGRP-SC, PGRP-LB, Diptericin (Dpt), and Cecropin (Cec) using dissected posterior midguts (without malpighian tubules) of 5-day-old flies. The target gene expression level in the tissues of GF^{WT} flies was taken arbitrarily as 1. (**C**) CDREs are required for the repression of *Cec* expression (indicated by arrow) in the posterior midgut (upper panel) and the proventriculus (lower panel). A schematic diagram of the *Cec* promoter is also shown. (**D**) Quantitative real-time PCR analysis using posterior midguts or fat body. Genotypes of flies: CR^{Cont} (*c729-GAL4/+*) and $CR^{Cad\text{-}RNAi}$ (*c729-GAL4/+; UAS-Cad-RNAi/+*). The target gene expression level in the gut or fat body of CR^{Cont} flies was taken arbitrarily as 1. Dpt, Diptericin; Cec, Cecropin; Drs, Drosomycin; Def, Defensin; Att, Attacin; Metch, Metchnikowin. (**E**) Quantitative real-time PCR analysis. Genotypes of flies (5-day-old GF or CR): Cont (*c729-GAL4/+*); Cad-RNAi (*c729-GAL4/+; UAS-Cad-RNAi/+*); Cad-RNAi + Dredd ($Dredd^{B118}$; *c729-GAL4/+; UAS-Cad-RNAi/+*); Cad-RNAi + TAK1 (*TAK1; c729-GAL4/+; UAS-Cad-RNAi/+*). The target gene expression level in the CR^{Cont} gut was taken arbitrarily as 1. (**F**) Overexpression of *Cad* can abolish infection-induced AMP expression. The target gene expression level in uninfected control gut was taken arbitrarily as 1. Natural gut infection for 6 hours with *Ecc15* in (A), (D), and (F) was performed as described in (*21*); relative expression levels in (B), (D), (E), and (F) are expressed as means ± SD ($P < 0.05$) of three different experiments.

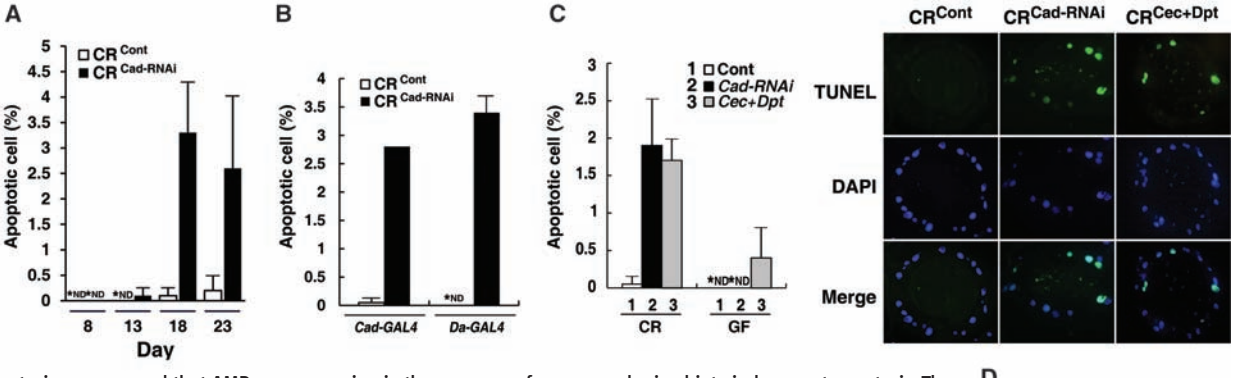


Fig. 2. Apoptosis assays reveal that AMP overexpression in the presence of commensal microbiota induces gut apoptosis. The posterior midguts of adult flies at different ages (**A**) and those of 18-day-old adult flies (**B** to **D**) were used. An apoptosis index was determined by dividing the number of apoptotic cells by the total number of cells and multiplying by 100. Values represent means ± SD ($P < 0.05$) of five independent experiments. The genotypes of *Cad-RNAi* flies with the *c729-GAL4* driver in (A) and (C) were described in Fig. 1E. (A) Time course analyses of gut epithelial cell apoptosis in *Cad-RNAi* flies. (B) Apoptosis in *Cad-RNAi* flies with different GAL4 drivers: the intestine-specific *Cad-GAL4* (*Cad-GAL4/+; UAS-Cad-RNAi/+*) and the strong ubiquitous *Da-GAL4* (*UAS-Cad-RNAi/Da-GAL4*). Flies carrying only GAL4 driver were used as controls. (C) Apoptosis in CR and GF flies. The genotype of Cec + Dpt flies was *c729-GAL4/UAS-Dpt; UAS-Cec/+*. Left panel, apoptosis index; right panel, representative images of terminal deoxynucleotidyl transferase–mediated deoxyuridine triphosphate nick end labeling (TUNEL) staining. TUNEL-positive cells are labeled in green. (D) Apoptosis in flies overexpressing a single AMP. Fly genotypes: Cont (*Da-GAL4/+*); *Cec* (*UAS-Cec/+; Da-GAL4/+*); *Dpt* (*UAS-Dpt/+; Da-GAL4/+*). *ND, not detected.

subset of NF-κB–dependent genes such as AMPs. When we examined *Cad-RNAi*–induced AMP derepression in the IMD-Relish pathway mutant genetic backgrounds ($CR^{Cad\text{-}RNAi\ +\ Dredd}$ or $CR^{Cad\text{-}RNAi\ +\ TAK1}$ flies) or in the absence of commensals ($GF^{Cad\text{-}RNAi}$ flies), the high level of AMP derepression was completely abolished (Fig. 1E).

These data show that Cad acts as a gut-specific transcriptional repressor exerting its antagonistic role in commensal-induced NF-κB–dependent AMP induction. Furthermore, the overexpression of *Cad* in the gut could abolish the infection-induced AMP expression (Fig. 1F). Thus, it is likely that a dynamic equilibrium between Cad and Relish is one of the major mechanistic aspects determining the selective deployment of gut IMD-AMP pathway potential (*21*) (Fig. 1 and fig. S6). Because AMPs have a microbicidal effect against a broad spectrum of microorganisms, Cad-mediated AMP repression is likely required for establishing an optimal environment for the commensal microbiota.

Role of *Cad* in gut homeostasis. Several recent lines of evidence suggest that deregulation of the intestinal NF-κB pathway may be relevant to the etiology and pathology of many important diseases, including inflammatory bowel diseases (IBDs) (*22*–*29*). Given that IBDs typically involve apoptosis of intestinal cells, we examined whether cell death might be occurring in the gut of *Cad-RNAi* flies. Time-course analyses with adult flies showed that gut epithelial cell apoptosis was detected in *Cad-RNAi* flies at day 18, but not in control flies of the same age (Fig. 2A). Cell death was also observed in the intestines of $CR^{Cad\text{-}RNAi}$ flies with different GAL4 drivers: *c729-GAL4*, *Cad-GAL4*, and *Da-GAL4* (Fig. 2, A and B). Interestingly, gut epithelial cell apoptosis was abolished in the $GF^{Cad\text{-}RNAi}$ gut (Fig. 2C), which suggests the involvement of commensal organisms in *Cad-RNAi*–mediated gut pathology. To determine whether the high apoptosis level seen in the *Cad-RNAi* gut was due to secondary effects of AMP hyperactivation, we overexpressed two AMP genes (*Cec* and *Dpt*) to mimic the *Cad-RNAi* genotype. AMP overexpression in CR flies ($CR^{Cec+Dpt}$) was sufficient to induce a high level of gut epithelial cell apoptosis (Fig. 2C). This was not the case in the absence of commensal microbiota ($GF^{Cec+Dpt}$) (Fig. 2C). When we overexpressed only a single AMP (*Cec* or *Dpt*) in CR flies (CR^{Cec} or CR^{Dpt}), similar gut epithelial cell apoptosis could be also observed (Fig. 2D). Overall, these in vivo results show that gut AMP overexpression in the presence but not in the absence of gut commensal microbes is sufficient to cause gut pathology.

Role of *Cad* in the gut commensal community structure. Because constitutive production of AMPs in the guts of *Cad-RNAi* flies likely affects the ecosystem of normal commensals, we next determined the dominant commensal species in the midgut of control (CR^{Cont}) and *Cad-RNAi* ($CR^{Cad\text{-}RNAi}$) flies. In wild-type flies, five commensal species dominate: *Lactobacillus plantarum* (*LP*), *Lactobacillus brevis* (*LB*), *Acetobacter pomorum* (*AP*), and two novel strains: *Gluconobacter* sp. strain EW707 (*G707*), and a bacterium in the family *Acetobacteraceae*, strain EW911 (*A911*) (*21*) (figs. S7 to S9 and table S1). All of these commensal bacteria, but not other bacteria such as *Ecc15* and *Escherichia coli*, could persist in the gut, demonstrating their competences as commensal bacteria (Fig. 3A). Real-time polymerase chain reaction (PCR)–based quantitative analyses of each commensal (*21*) (figs. S9 to S11) clearly revealed that the commensal community structure of the *Cad-RNAi* flies differed from that of the control flies (Fig. 3B). Three bacteria—*AP*, *LP*, and *LB*—were commonly dominant in the gut of both control and *Cad-RNAi* flies (Fig. 3B). However, *A911* [a dominant commensal member in controls, ~1.4×10^5 colony-forming units (CFUs) per gut] was markedly diminished and maintained at a very low level in the *Cad-RNAi* gut (~900 CFUs per gut), whereas *G707* (a minor commensal member in the control gut, ~800 CFUs per gut) emerged as a dominant commensal in the *Cad-RNAi* gut (~1.7×10^4 CFUs per gut) (Fig. 3B). Time-course analyses with *Cad-RNAi* flies showed that loss of *A911* was visible from as

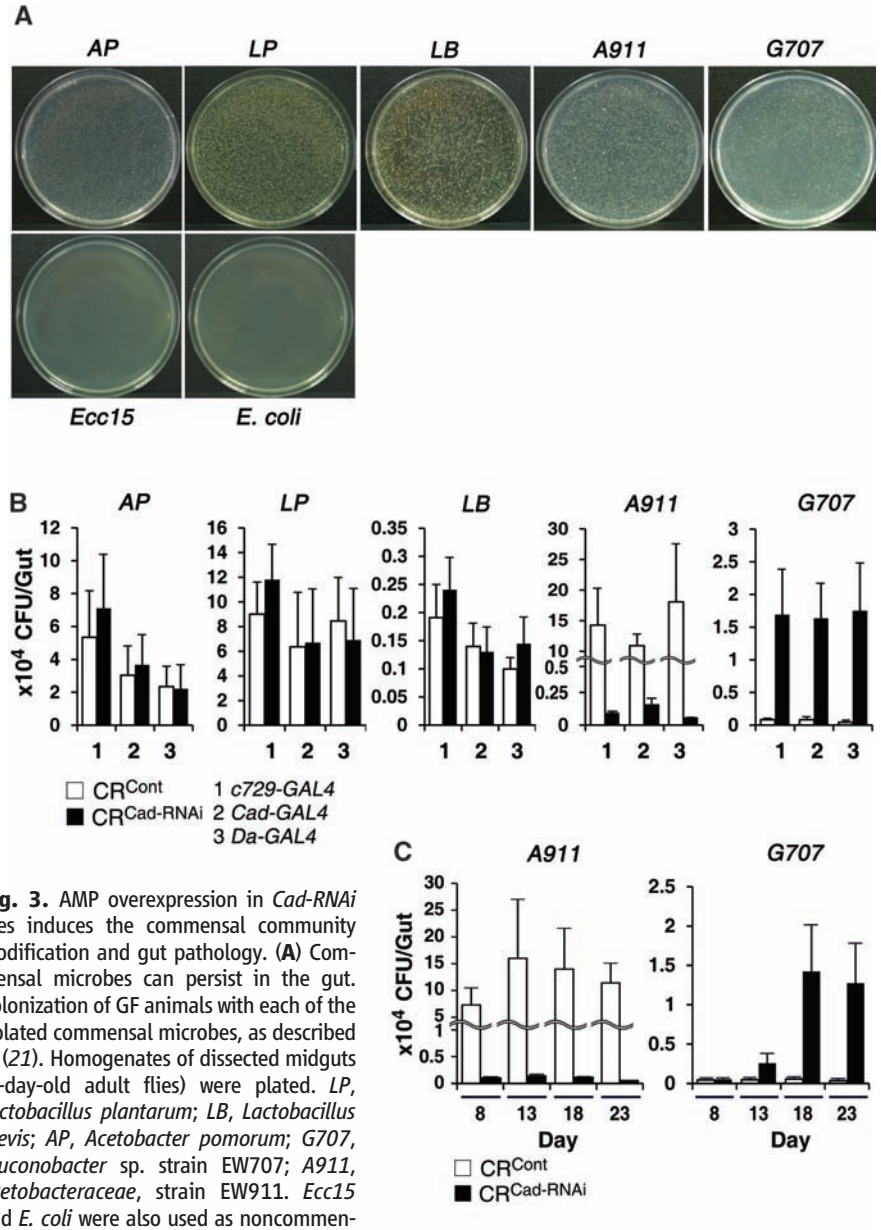


Fig. 3. AMP overexpression in *Cad-RNAi* flies induces the commensal community modification and gut pathology. (**A**) Commensal microbes can persist in the gut. Colonization of GF animals with each of the isolated commensal microbes, as described in (*21*). Homogenates of dissected midguts (8-day-old adult flies) were plated. *LP*, *Lactobacillus plantarum*; *LB*, *Lactobacillus brevis*; *AP*, *Acetobacter pomorum*; *G707*, *Gluconobacter* sp. strain EW707; *A911*, *Acetobacteraceae*, strain EW911. *Ecc15* and *E. coli* were also used as noncommensal microbes. (**B** and **C**) Real-time PCR-based analysis to quantify the number of each commensal microbe in the posterior midguts. Values represent means ± SD ($P < 0.05$) of three independent experiments. The genotypes of *Cad-RNAi* flies with the *c729-GAL4* driver were described in Fig. 1E. *Cad-RNAi* flies with *Cad-GAL4* or *Da-GAL4* were described in Fig. 2B. The posterior midguts of 18-day-old adult flies (B) and those of adult flies at different ages (C) were used.

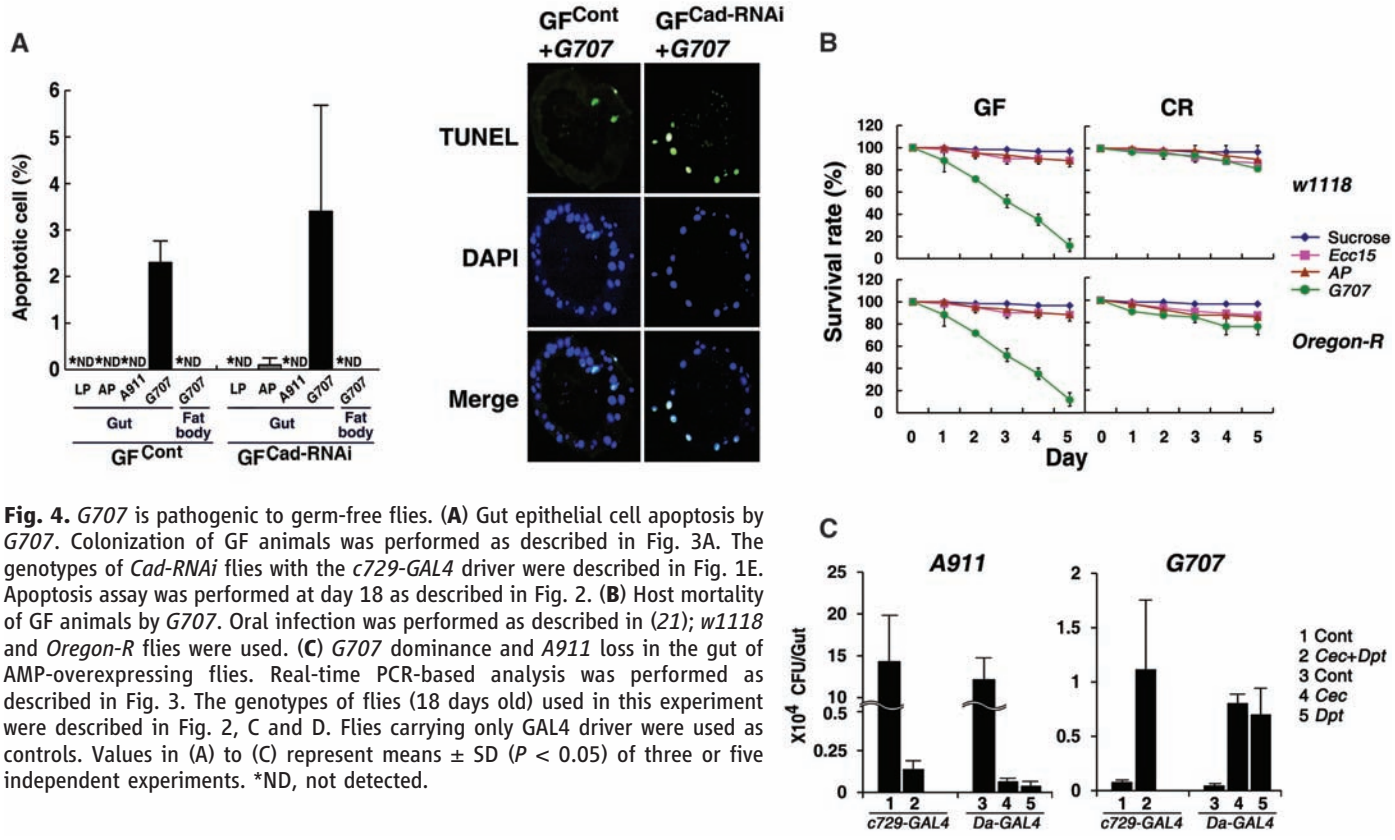


Fig. 4. *G707* is pathogenic to germ-free flies. (**A**) Gut epithelial cell apoptosis by *G707*. Colonization of GF animals was performed as described in Fig. 3A. The genotypes of *Cad-RNAi* flies with the *c729-GAL4* driver were described in Fig. 1E. Apoptosis assay was performed at day 18 as described in Fig. 2. (**B**) Host mortality of GF animals by *G707*. Oral infection was performed as described in (*21*); *w1118* and *Oregon-R* flies were used. (**C**) *G707* dominance and *A911* loss in the gut of AMP-overexpressing flies. Real-time PCR-based analysis was performed as described in Fig. 3. The genotypes of flies (18 days old) used in this experiment were described in Fig. 2, C and D. Flies carrying only GAL4 driver were used as controls. Values in (A) to (C) represent means ± SD ($P < 0.05$) of three or five independent experiments. *ND, not detected.

Table 1. In vitro antibacterial assay using synthetic Cec A1, performed as described in (*21*). The minimal Cec A1 concentration that prevented the growth of a given test organism was determined and was defined as the minimum inhibitory concentration (MIC).

MIC (μg/ml)					
Gram-positive bacteria		Gram-negative bacteria			
LP	*LB*	*AP*	*G707*	*A911*	*E. coli**
10 to 20	10 to 20	2.5	2.5	0.625	1.25

**E. coli* ATCC 25922.

early as day 8, whereas dominance of *G707* started at day 13 and reached the maximum level at day 18 (Fig. 3C). Given the high apoptosis levels in the gut epithelial cells of GF^{Cont} flies fed on the homogenates of the $CR^{Cad\text{-}RNAi}$ gut, but not on homogenates of the CR^{Cont} gut (fig. S12), we reasoned that the dominant commensals in the *Cad-RNAi* gut, such as *G707*, may be involved in the gut pathology.

To validate this hypothesis, we introduced each of the isolated commensal bacteria into GF^{Cont} or $GF^{Cad\text{-}RNAi}$ embryos and maintained these bacteria until the adult stage to generate monoassociated flies. High gut epithelial cell apoptosis was observed in all of the GF flies (GF^{Cont} or $GF^{Cad\text{-}RNAi}$) when the single organism was *G707* (Fig. 4A), but not when the single organism was one of the other commensal organisms (*LP*, *AP*, or *A911*) (Fig. 4A). Additionally, *G707* did not induce the apoptosis in the fat body (Fig. 4A), and *G707* was pathogenic to the host when GF flies were subjected to gut infection with *G707* (Fig. 4B). These results show that *G707* is indeed pathogenic in a gut-specific manner when allowed to become the dominant gut microbe.

To further confirm that reorganization of the gut microbiota composition seen in the *Cad-RNAi* flies was due to the constitutive overexpression of microbicidal AMPs, we tested whether the artificial shift of gut microbiota composition could occur as a result of overexpression of AMPs. We found that overexpression of *Cec* and *Dpt* ($CR^{Cec+Dpt}$) was sufficient to induce the modification of commensal community (i.e., *G707* dominance and a loss of *A911*), as in the case of the *Cad-RNAi* guts (Fig. 4C). Similar results could also be observed when we overexpressed only a single AMP (CR^{Cec} or CR^{Dpt}) (Fig. 4C). To confirm that the loss of *A911* was due to the high sensitivity of this strain to AMP, we performed an in vitro antibacterial test with synthetic Cec A1 and determined the minimum inhibitory concentration (*21*). The result showed that *A911* was highly susceptible to low concentrations of synthetic Cec A1, whereas all other commensals as well as *G707* exhibited relatively high resistance to Cec A1 (Table 1). Thus, we conclude that gut AMP overexpression in *Cad-RNAi* flies acts as a distinct selection pressure on different commensal microbes, resulting in modification of the commensal community structure and pathogenic conditions in the gut.

Role of the normal commensal community in gut homeostasis. Because *G707* is normally present at a very low level in the wild-type commensal community, we investigated whether the normal wild-type commensal community structure could antagonize the dominance of *G707*. To accomplish this, we introduced *G707* by oral feeding into the gut of either GF^{WT} or CR^{WT} flies and then examined its persistence. The results indicated that *G707* bacteria could persist in the absence of other commensal organisms in GF^{WT} flies but disappeared rapidly and were maintained at a low level in the presence of the normal commensal community in CR^{WT} flies (Fig. 5A). Consistent with these results, we observed a high-level gut epithelial cell apoptosis in the *G707*-challenged GF^{WT} flies, but not in the *G707*-challenged CR^{WT} flies (Fig. 5A), which demonstrates the role of normal commensal community structure in gut homeostasis by maintaining *G707* at a low level.

Given the numerical inferiority of *A911* in *G707*-dominant gut environments, we investigated whether the presence of *A911* was suf-

ficient to suppress *G707* dominance in the gut. To accomplish this, we generated monoassociated flies by introducing isolated *A911* into GF^{WT} embryos and maintaining this presence until the adult stage. When *G707* was introduced into these monoassociated flies through oral feeding, the *G707* bacteria rapidly decreased to a low level in the presence of *A911*, whereas a high initial *G707* level was able to persist in the absence of *A911* (in the case of GF^{WT} flies) (Fig. 5A). After *G707* introduction, a high apoptosis index was observed in the gut of the GF^{WT} flies, but not in the gut of *A911*-monoassociated flies (Fig. 5A). This result shows that the presence of *A911* is sufficient to prevent *G707* dominance, and that the loss of *A911* in the AMP-overexpressing genotype flies (such as *Cad-RNAi* flies or *Cec*- or *Dpt*-overexpressing flies) is likely to be responsible for *G707* dominance.

Taken together, these results reveal that Cad acts as a critical host factor that maintains the immune homeostasis responsible for preservation of the normal commensal community structure. Failure of the balanced regulation of AMP, as in the case of *Cad-RNAi* flies, can act as a novel selection pressure leading to modification of the gut commensal structure. Because *G707* is a highly pathogenic organism (Fig. 4, A and B), the *G707*-dominating host genotype acts as an initial cause of gut apoptosis, whereas the dominance of *G707* acts as a direct cause of gut apoptosis.

Role of *Cad* in host physiology. The apoptosis of gut cells seen in *Cad-RNAi* flies with different GAL4 drivers was accompanied by elevated mortality (Fig. 5E and figs. S13 and S14). The high mortality of $CR^{Cad\text{-}RNAi}$ flies was significantly ameliorated in the absence of commensal ($GF^{Cad\text{-}RNAi}$ flies) (Fig. 5E) or in the presence of commensals under the IMD pathway mutant genetic backgrounds ($CR^{Cad\text{-}RNAi + Dredd}$) (fig. S14). The normal survival rate could be also observed in $GF^{Cad\text{-}RNAi}$ flies stably associated with major commensal microbiota (*AP*, *LP*, *LB*, and *A911*) excluding *G707*, which implies the involvement of *G707* in *Cad-RNAi*–mediated host mortality (Fig. 5E). Furthermore, the c-Jun N-terminal kinase pathway and interleukin-1β converting enzyme were shown to be involved in the microbiota-induced gut apoptosis and mortality process of the *Cad-RNAi* flies (fig. S15).

To demonstrate that the initial cause of death in the *Cad-RNAi* flies was the reduced expression of *Cad*, we attempted an in vivo rescue experiment. Restoration of the basal AMP level (Fig. 5B), healthy microbiota community structure (Fig. 5C), reduced apoptosis (Fig. 5D), and normal host survival levels (Fig. 5E) could be achieved in the *Cad-RNAi* flies by genetic reintroduction of *Cad* ($CR^{Cad\text{-}RNAi + Cad}$ flies). Taken together, our results show that the intestinal homeobox *Cad* gene is responsible for the delicate immune homeostasis in the microbe-contacting gut tissue, which is essential for preservation of the normal microbiota community, gut homeostasis, and host survival.

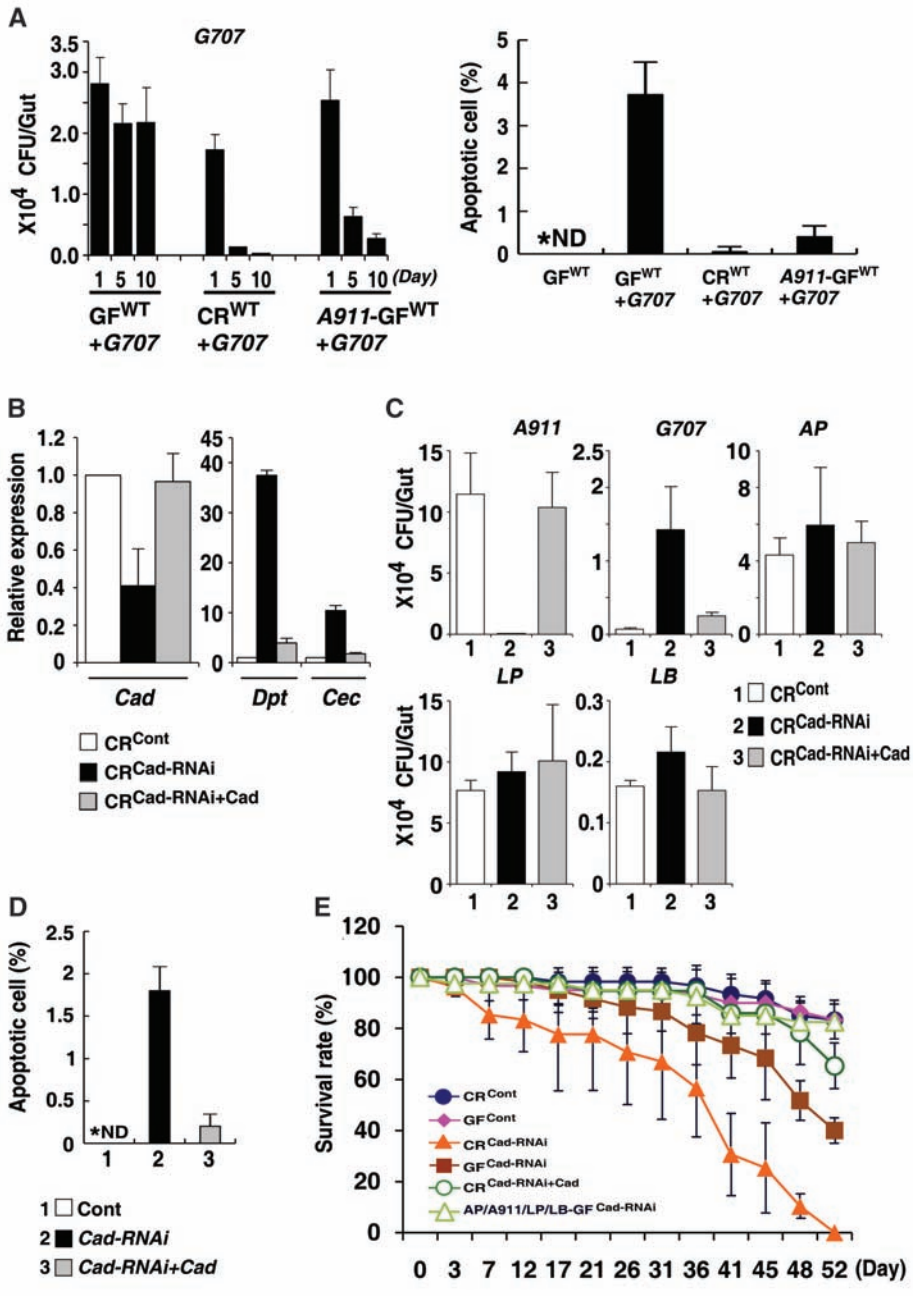


Fig. 5. Cad is indispensable for immune homeostasis in preserving the indigenous commensal community and host health. (**A**) Normal commensal community structure is important to suppress *G707* dominance in the gut. Germ-free wild-type flies (GF^{WT}), wild-type flies carrying normal commensal microbiota (CR^{WT}), and wild-type flies monoassociated with *A911* (*A911*-GF^{WT}) were used. *G707* was introduced to these flies (3 days old) through oral feeding for 48 hours. Left panel: Real-time PCR-based analysis to quantify *G707* in the gut was performed at 1, 5, and 10 days after *G707* feeding. Right panel: Apoptosis assay was performed as described in Fig. 2. Values represent means ± SD ($P < 0.05$) of five independent experiments. (**B** to **E**) Restoration of basal AMP levels (B), normal commensal community structure (C), reduced apoptosis (D), and normal survival levels (E) could be achieved in the *Cad-RNAi* flies by genetic reintroduction of *Cad*. The posterior midguts of 18-day-old adult flies were used for analyses of commensal community structure and apoptosis. Apoptosis assay was performed as described in Fig. 2. Genotypes of CR flies: Cont (*c729-GAL4/+*); Cad-RNAi (*c729-GAL4/+; UAS-Cad-RNAi/+*); Cad-RNAi + Cad (*c729-GAL4/UAS-Cad; UAS-Cad-RNAi/+*). GF^{Cont} and $GF^{Cad\text{-}RNAi}$ flies were also used. AP/A911/LP/LB-$GF^{Cad\text{-}RNAi}$ flies carrying four major commensals (*AP*, *A911*, *LP*, and *LB*, excluding *G707*) were generated by colonizing GF embryos with commensals as described in (*21*). Values represent means ± SD ($P < 0.05$) of three independent experiments. *ND, not detected.

Discussion. The gut epithelia of virtually all organisms have evolved to form a mutually beneficial strategic alliance with microorganisms (*6*). However, the role of the host factor in gut homeostasis has been largely overlooked, and little information regarding the molecular principles of gut homeostasis established by the interrelationship between the host immunity and commensal microbiota is available. In mammalian gut epithelia, deregulation of the NF-κB and AMP signaling pathways was found to be implicated in the pathogenesis of chronic IBDs (*22*–*29*). However, the enormous diversity of the resident microbiota community of the mammalian gut (e.g., 500 to 1000 different species in the human gut) (*30*) and the genetic complexity of the host immune system make it difficult to clearly establish the molecular links that would clarify the relations among immune genotype, commensal microbiota structure, and disease phenotype at the organism level.

By using a genetically amenable model organism harboring an extremely simple gut commensal structure, we have shown that the commensal microbiota community structure links the defective immune genotype to the gut disease phenotype. Surprisingly, *Drosophila* gut epithelia have evolved an immune strategy by recruiting a developmental master control gene, *Cad*, to maintain appropriate AMP levels for preservation of the normal flora community structure. Defective regulation of the AMP level, as seen in the case of the *Cad-RNAi* genotype, promotes gut pathology by exerting a selection pressure that favors the dominance of a pathogenic commensal, *G707*, rather than by acting as a direct cause of the disease phenotype. The emergence of a disease-causing commensal organism under an immune-defective genotype indicates the involvement of a microorganism as an origin of chronic inflammation. Further elucidation of the link between the immune genotype–based commensal community structure and host physiology may provide important insights into the causative role of pathogenic commensal microbes in a variety of chronic inflammatory diseases of the commensal-contacting epithelia.

References and Notes

1. L. V. Hooper, J. I. Gordon, *Science* **292**, 1115 (2001).
2. T. T. Macdonald, G. Monteleone, *Science* **307**, 1920 (2005).
3. F. Bäckhed, R. E. Ley, J. L. Sonnenburg, D. A. Peterson, J. I. Gordon, *Science* **307**, 1915 (2005).
4. T. A. Koropatnick *et al*., *Science* **306**, 1186 (2004).
5. P. J. Turnbaugh *et al*., *Nature* **444**, 1027 (2006).
6. C. Dale, N. A. Moran, *Cell* **126**, 453 (2006).
7. S. Rakoff-Nahoum, J. Paglino, F. Eslami-Varzaneh, S. Edberg, R. Medzhitov, *Cell* **118**, 229 (2004).
8. E.-M. Ha, C.-T. Oh, Y. S. Bae, W.-J. Lee, *Science* **310**, 847 (2005).
9. E. M. Ha *et al*., *Dev. Cell* **8**, 125 (2005).
10. J. H. Ryu *et al*., *EMBO J.* **25**, 3693 (2006).
11. B. Lemaitre, J. Hoffmann, *Annu. Rev. Immunol.* **25**, 697 (2007).
12. A. Zaidman-Remy *et al*., *Immunity* **24**, 463 (2006).
13. V. Bischoff *et al*., *PLoS Pathog.* **2**, e14 (2006).
14. D. Ferrandon *et al*., *EMBO J.* **17**, 1217 (1998).
15. T. Onfelt Tingvall, E. Roos, Y. Engstrom, *EMBO Rep.* **2**, 239 (2001).
16. P. Tzou *et al*., *Immunity* **13**, 737 (2000).
17. N. Vodovar *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11414 (2005).
18. J. H. Ryu *et al*., *Mol. Cell. Biol.* **24**, 172 (2004).
19. M. Mlodzik, W. J. Gehring, *Cell* **48**, 465 (1987).
20. E. Moreno, G. Morata, *Nature* **400**, 873 (1999).
21. See supporting material on *Science* Online.
22. M. Karin, T. Lawrence, V. Nizet, *Cell* **124**, 823 (2006).
23. C. Zaph *et al*., *Nature* **446**, 552 (2007).
24. K. S. Kobayashi *et al*., *Science* **307**, 731 (2005).
25. S. Maeda *et al*., *Science* **307**, 734 (2005).
26. A. Nenci *et al*., *Nature* **446**, 557 (2007).
27. R. J. Xavier, D. K. Podolsky, *Nature* **448**, 427 (2007).
28. H. Xiao *et al*., *Immunity* **26**, 461 (2007).
29. N. H. Salzman, M. A. Underwood, C. L. Bevins, *Semin. Immunol.* **19**, 70 (2007).
30. P. B. Eckburg *et al*., *Science* **308**, 1635 (2005); published online 14 April 2005 (10.1126/science.1110591).

31. Supported by the National Creative Research Initiative Program of the Ministry of Science and Technology, Korea.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1149357/DC1
Materials and Methods
SOM Text
Figs. S1 to S15
Table S1
References

17 August 2007; accepted 21 December 2007
Published online 24 January 2008;
10.1126/science.1149357
Include this information when citing this paper.

REPORTS

Quantum Phase Extraction in Isospectral Electronic Nanostructures

Christopher R. Moon,[1] Laila S. Mattos,[1] Brian K. Foster,[2] Gabriel Zeltzer,[3] Wonhee Ko,[3] Hari C. Manoharan[1]*

Quantum phase is not directly observable and is usually determined by interferometric methods. We present a method to map complete electron wave functions, including internal quantum phase information, from measured single-state probability densities. We harness the mathematical discovery of drum-like manifolds bearing different shapes but identical resonances, and construct quantum isospectral nanostructures with matching electronic structure but divergent physical structure. Quantum measurement (scanning tunneling microscopy) of these "quantum drums"—degenerate two-dimensional electron states on the copper(111) surface confined by individually positioned carbon monoxide molecules—reveals that isospectrality provides an extra topological degree of freedom enabling robust quantum state transplantation and phase extraction.

The local structure of wave functions can now be experimentally determined in many materials. However, just as bonding in molecules or conductivity in solids depends not only on the magnitude of the orbital wave functions but also on the phase, the electronic properties and dynamics of nanostructures critically depend on determining both the magnitude and the phase of wave functions. In a classical system, measuring the phase of a standing wave is trivial, but the internal phase of a quantum wave function ψ is not an observable; only the probability density $|\psi|^2$ can be determined from direct measurement. To determine phase, some form of additional information is required.

Recently, quantum phase has been inferred in several experiments, including spectral interferometry of Rydberg wave packets (*1*) and tomography of molecular orbitals (*2*). Such experiments are performed on atoms or molecules in the gas phase and require ultrafast measurements of the response of a quantum state to an impinging wave or excitation. In solid-state systems, phase-sensitive measurements have been performed in quantum dots and rings (*3*) and with computational post-processing of diffraction patterns (*4*). All of these experiments retrieve the underlying phase differences through interferometry once a reference beam and geometry are defined.

Here, we present a noninterferometric method to map the internal quantum phase of a solitary, time-independent state by harnessing the topological property of isospectrality as the additional degree of freedom. We create a particular pair of geometric shapes with matching spectral properties. We then show that the complete phase information of wave functions in both structures can be experimentally deter-

[1]Department of Physics, Stanford University, Stanford, CA 94305, USA. [2]Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA. [3]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA.

*To whom correspondence should be addressed. E-mail: manoharan@stanford.edu

mined. Because this technique is based on a fundamental mathematical symmetry, it is general and should apply to several types of nanostructures and materials, but we demonstrate it for wave functions of confined regions on metal surfaces. To develop this method, we have drawn from recent answers to a famous, long-standing question in mathematics.

A general method for obtaining an excitation spectrum from a given geometry is to apply the Laplacian operator to the oscillation variable. Much scrutiny has been given to the relation, familiar to any musician, between the allowed frequencies of a wave—the eigenvalues of the Laplacian—and the shape of the boundary that encloses it. Given all of the eigenfrequencies of a resonator, its area can be determined (*5*), as well as its perimeter and the number of embedded holes (*6*). But does a resonator's eigenspectrum uniquely specify its geometry? In other words, as Kac asked 40 years ago (*7*), "Can one hear the shape of a drum?"

When Kac's seminal paper was presented in 1966, Milnor had already proven the existence of two noncongruent 16-dimensional flat tori in which the Laplacian's eigenvalue spectrum is identical (*8*). In later decades, mathematicians struggled to find such isospectral geometries in lower-dimensional systems. Sunada discovered a group-theoretic method for proving isospectrality (*9*) in the 1980s, and Buser proved a class of isospectral manifolds by "pasting" two-dimensional (2D) flat tiles together in higher dimensions (*10*). However, the dimensionality of proven isospectral domains was only slowly whittled down to three dimensions (*10*, *11*). [Ironically, the one-dimensional analog of Kac's question is trivial—one can always hear the length ("shape") of a string by identifying its fundamental frequency.]

Did Kac's 2D drums mimic their closest cousins in 3D, where isospectral pairs had been found, or match the 1D case where the spectrum uniquely defines the geometry? It was not until 1992 that Gordon *et al*. mathematically discovered the first 2D isospectral domains (*11*), finally answering Kac's enigma: No, one cannot hear the shape of a drum.

This result in ostensibly obscure mathematics crossed over to general audiences because of its connection with simple ideas of sound and shape (*12*, *13*). However, there were still many questions about whether this theoretical spectral equality could be observed in real, imperfect systems. Sridhar and Kudrolli (*14*) performed the first experimental test of classical isospectrality by measuring resonances in microwave transmission through thin metal cavities. Even and Pieranski (*15*) verified isospectrality in actual vibrating drums made from liquid crystal films.

Because the time-independent Schrödinger equation is also a wave equation defined by the Laplacian and boundary conditions, systems of quantum particles can also theoretically be isospectral. Such systems—an electron resonator, for example—can have interesting properties impossible to study in classical drums. Indeed, it has recently been argued that quantum effects can allow one to distinguish between structures that otherwise would be isospectral (*16*).

To explore quantum isospectrality and its sensitivity to our nanofabrication abilities, it is necessary to find a system whose frequency response is highly sensitive to geometry, but whose particle-particle interactions remain relatively constant. Here, we detail studies that establish the limits of quantum isospectrality in a system where the 2D electron density is sufficiently high that Coulomb interactions are minimized and further screened by bulk electrons (*17*), and quasi-particle lifetimes are sufficiently long to preserve quantum coherence. We then apply these isospectral geometries to quantum phase measurement.

We designed quantum mechanical isospectral electron resonators by following symmetry and tiling rules from the mathematical literature (*11*, *18*, *19*). For our "vibrating" medium, we used the 2D Fermi sea of electrons that inhabits the Cu(111) surface state (*20*). These electrons were confined to theoretically isospectral shapes with walls made from CO molecules positioned with the tip of a scanning tunneling microscope (STM) (*21*, *22*).

We chose CO because it is easily and accurately manipulated, yet stable over a wide range of sample voltage V and tunnel current I (*23*). We can stably pack CO molecules closer than adsorbed atoms, enabling us to form nearly continuous walls (Fig. 1, A to C). For these experiments, we operated a home-built STM at 4.2 K in ultrahigh vacuum (*13*). CO bonds directly above Cu atoms, and it images as a topographical depression of ~50 pm at low biases. After ~0.012 monolayers of CO were dosed, the molecules were positioned to form the complex boundaries required for isospectral geometries (*13*). We then performed successively stricter tests of isospectrality: spectral matching, amplitude matching, and the transplantation of wave functions. The results were then used to map quantum phase.

The most basic known planar isospectral domains (*11*) are nine-sided polygons previously termed "Bilby" and "Hawk" (*24*). Each is a polyform composed of seven identical triangles and created via a specific series of reflections,

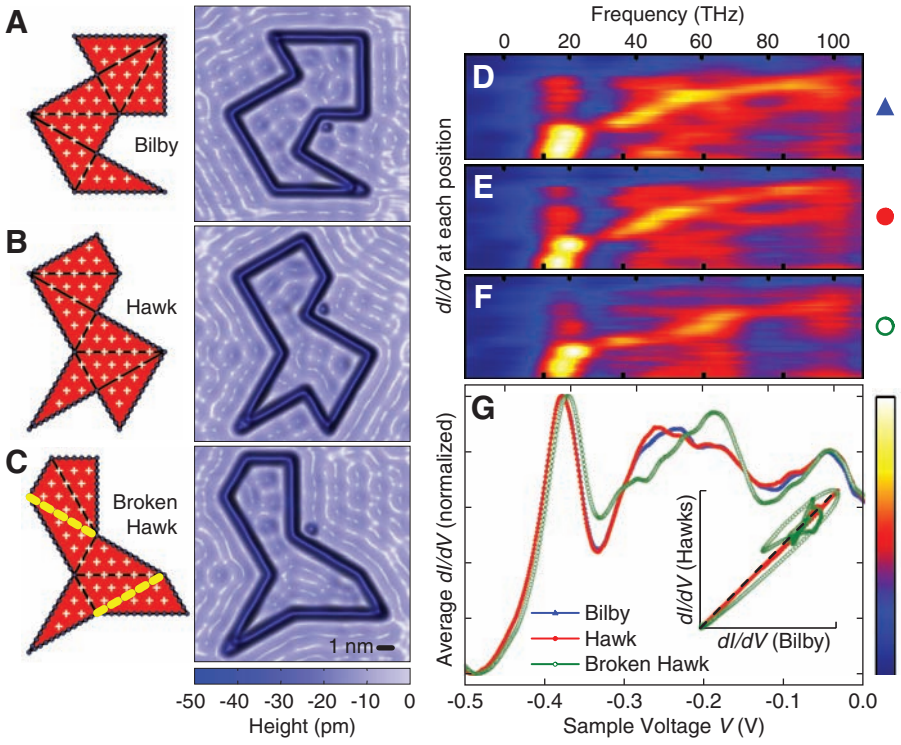


Fig. 1. Design and realization of quantum isospectral resonators assembled from CO molecules on the Cu(111) surface. (**A** to **C**) Schematics and STM topographs of the Bilby (A), Hawk (B), and Broken Hawk (C) domains. The seven identical 30°-60°-90° triangles composing each shape are shown in red. Blue dots indicate the positions of wall molecules. White crosses mark locations where *dI/dV* spectroscopy was performed. STM topographs are 15 nm by 15 nm (V = 10 mV, I = 1 nA). A single CO molecule used for registration between spectra (*13*) accompanies each nanostructure. (**D** to **F**) Spectral fingerprints (*dI/dV* spectra) acquired throughout Bilby (D), Hawk (E), and Broken Hawk (F). (**G**) The normalized averages of the Bilby and Hawk spectra match closely, consistent with isospectrality, whereas the average Broken Hawk spectrum clearly differs. Inset: Spectral correlation plot (dashed line denotes perfect match) quantifying Bilby-Hawk isospectrality and its departure in Broken Hawk.

such that every triangle is the mirror image of its neighbors. Because there is proven flexibility in choice of interior angles for the base triangle (*18*) and the Cu(111) surface is a close-packed triangular lattice, we chose tilings based on the 30°-60°-90° triangle. The molecular designs superimposed over the ideal Bilby and Hawk shapes and the STM topographs of the experimentally realized quantum resonators are shown in Fig. 1, A and B. Each structure was assembled from 90 individual CO molecules, bounds an area of ~57 nm^2, and holds ~30 electrons (*13*).

We also carried out control experiments where we violated the construction principles (*13*) of the isospectral structures. One such structure, "Broken Hawk," was created by flipping two sections of Hawk (along yellow dashed lines, Fig. 1C). The area, perimeter, and molecule count of Broken Hawk are identical to those of Hawk and Bilby.

We performed *dI/dV* spectroscopy throughout each of these structures to measure their eigenenergies (*13*). To observe all eigenmodes, we acquired dense sets of spectra along an interior triangular lattice (white crosses in Fig. 1, A to C; 46 spectra for each structure). To graphically compare the overall spectral content of the three structures, we condensed the total 138 spectra into "spectral fingerprints" (Fig. 1, D to F). Here, for each structure, the *dI/dV*-versus-*V* point spectra are sorted by energy of largest peak and plotted with *dI/dV* mapped to the color scale shown. Although there are gross features in common among the structures (as expected because all have the same area and perimeter), the detailed features reveal that Bilby and Hawk are most closely matched, whereas Broken Hawk visibly differs. To further collapse the data set, we plotted the average *dI/dV* (Fig. 1G) for each structure. Again, we find excellent agreement between the Bilby and Hawk shapes and a substantial deviation in Broken Hawk. For example, the first peak (which arises from the first two eigenmodes of each structure; see below) has been shifted by ~7 mV in Broken Hawk, and the subsequent structure shows further departures. These data provide initial evidence of quantum isospectrality in suitably assembled geometries.

The natural resonance frequencies of 2D electrons in these nanostructures are on the order of $v_F/l \sim 100$ THz, where the Fermi velocity v_F is $\sim 10^8$ cm/s and the characteristic structure size is $l \sim 10$ nm. For reference, we have added this frequency scale to Fig. 1, D to G. As another simple yet compelling demonstration of quantum isospectrality, we have converted the measured average spectra in Fig. 1G to audio frequencies so that one can "listen" to each of the structures. The results (movie S1) portray the THz ringing of electrons as one might hear them. Indeed, Bilby and Hawk—as quantum drums—"sound the same," whereas Broken Hawk can be audibly distinguished.

Thus, we have been able to observe evidence for isospectrality in a quantum system even with nonideal boundaries. In an effort to quantify the limits of isospectrality in these quantum structures, we devised a complementary analysis of the ensemble data set.

In an infinite hard-wall potential, and if the confined electrons had very long lifetimes, our spectra would show sharp spikes at the eigenenergies. However, several factors widen these peaks into broader resonances. Electrons can tunnel across the confinement potential imposed by the molecular walls (*25*) or scatter from the molecules into bulk states (*26*). Intrinsic lifetime broadening of the surface states also occurs, the result of electron-electron and electron-phonon interactions that dominate for electrons at low energies (*17*).

To overcome these linewidth effects, we applied a self-consistent fitting procedure that simultaneously matches a series of Lorentzians to all of the spectra taken in a structure (*13*). The fitted eigenenergies for Bilby and Hawk (Fig. 2A) are remarkably similar, confirming their isospectrality in our quantum system. By plotting the amplitudes of each Lorentzian across the isospectral domains (Fig. 2C, Bilby; Fig. 2D, Hawk), the spatial structure of the underlying eigenmodes is revealed, presenting a purely experimental confirmation that the fitted energies correspond to actual quantum resonances.

All possible two-way correlations among the three data sets of Bilby, Hawk, and Broken Hawk (Fig. 2A) show that Bilby and Hawk are numerically closest by a factor of ~9 in root-mean-square (RMS) energy difference. To quantify variation from perfect isospectrality, we plotted the deviation ΔV from the average Bilby/Hawk energy at each mode (Fig. 2B). Here, it is evident that statistically Bilby and Hawk are isospectral (within measurement error) and Broken Hawk is distinct. As a conservative bound on quantum isospectrality in this system, we can say that Bilby and Hawk are isospectral to better than ±2 mV (Fig. 2B, dashed lines), and the alterations built into Broken Hawk take it well outside this window. These tests demonstrate quantum isospectrality's extreme sensitivity to correct geometry and gentle topological perturbations. Results for other geometric perturbations are shown in fig. S4.

Next, we studied isospectral quantum resonators with greater complexity (*18*), consisting of 21 triangles each (Fig. 3, A and B). We refer to these structures as "Aye-aye" and "Beluga" (or *A* and *B* in shorthand). In addition to being isospectral, this pair of structures theoretically possesses points that are homophonic; that is, not only will the same frequencies result if the drums are "struck" at these points, but the relative amplitudes of those frequencies will be

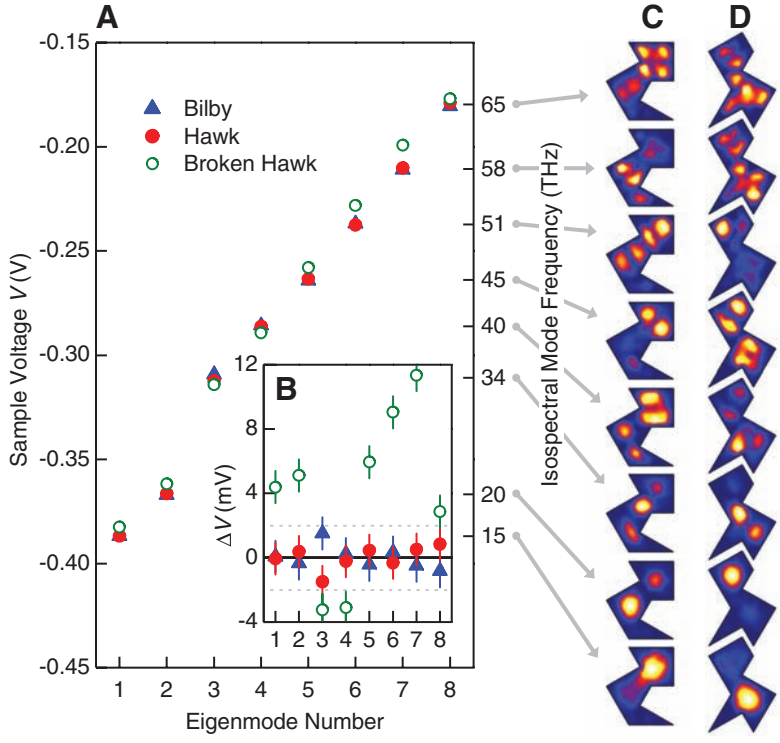


Fig. 2. Quantum isospectrality. (**A**) The first eight electron energy levels in Bilby, Hawk, and Broken Hawk resonators, found by simultaneously fitting Lorentzians (see text) to all spectra in Fig. 1, D to F. (**B**) The difference between each of the energies in (A) and the mean of the Bilby-Hawk energy [error bars discussed in (*13*)]. (**C** and **D**) Spatial maps of the amplitudes of the fitted Lorentzians reveal the eigenmodes of the isospectral structures Bilby (C) and Hawk (D).

identical (*27*). The homophonic points occur at the meeting of six triangles in the interiors of the two shapes.

Our experimental constructs were assembled from 105 CO molecules and hold ~14 electrons. We extracted the eigenenergies of these structures after dense spectroscopy (29 locations each) and performed the same Lorentzian-fitting procedure described above, and found them to be isospectral within experimental error (±3 mV). Here we focus on the special aspect of these geometries: the response at the homophonic points (white crosses in Fig. 3, A and B). The spectra (Fig. 3C) show that the electronic structure at these locations is nearly identical, despite their entirely different environments.

This agreement should be contrasted with the difference observed in any other pair of isolated *dI/dV* traces acquired from isospectral resonators. For example, we measured the RMS difference between each pair of *dI/dV* traces in *A* and *B*. Of the 406 distinct combinations, the minimum difference (the maximum correlation) occurs between the homophonic points in *A* and *B*; the next closest pair has twice this difference. Remarkably, if the geometry of *A* is subtly broken, the homophonic point spectrum changes far more than when the geometry is radically changed to the *B* shape (fig. S5). Our results are highlighted in movie S2, which plays the audio conversion (*13*) of three representative pairs of point spectra, including the homophonic pair. Here, the similarity between the homophonic points is audibly contrasted with the easily discernable differences between other points. In analogy to two differently shaped drums struck exactly at homophonic points, this quantum version demonstrates how to materialize identical local electronic structure at two remote locations surrounded by different global environments.

Having verified the equivalence of the electron energies in specific pairs of quantum nanostructures, we now turn to their eigenfunctions. The mathematical underpinnings of the proof of isospectrality rely on Berard's generalization (*28*, *29*) of Sunada's theorem (*9*), which established a relationship between the normal modes of these shapes. Specifically, if an eigenfunction of one member of an isospectral pair is known, there exists a nonisometric transformation (*30*) yielding the corresponding eigenfunction of its isospectral complement. The transformation works by cutting the wave function data for one domain into its constituent triangles, which are then combined by appropriate superpositions and transplanted onto the second domain (*13*). For example, in the 21-triangle homophonic structures, the triangular sections of the wave functions are labeled (Fig. 3, A and B) and are represented as column vectors, so that $\mathbf{A} = (A_1, A_2, \ldots, A_{21})$ and $\mathbf{B} = (B_1, B_2, \ldots, B_{21})$. Then, the transplantation operation is an isomorphism that can be represented by a 21 × 21 matrix **T** (*13*, *18*), and $\mathbf{B} = \mathbf{TA}$. Every row of **T** has five nonzero elements, meaning that every triangle of a *B* wave function is the superposition of five sections of the *A* wave function. Every nonzero element of **T** is ±1, where the negative sign is accompanied by flipping the triangle about one of its sides to match the symmetry of its destination triangle. We note that **T** is independent of energy and hence applies to any wave function. Incidentally, there also exists a second (linearly independent) transplantation matrix, **T′**, which has 16 nonzero elements per row; therefore, any norm-preserving linear combination of **T** and **T′** will also yield valid transplantations.

To measure the wave functions of the homophonic structures, we acquired high-resolution open-loop *dI/dV* maps at each mode energy (*13*). We start with the experimental measurements of the ground-state mode in *A* and *B* (Fig. 4A). When the *A* mode is then transplanted, the result is in excellent agreement with the data for the *B* mode. Concordantly, by inverting the transplantation matrix **T**, the *B* data can be transformed into a match for the *A* data. We have also tested the **T′** matrix described above with similar results, and we emphasize that the structures of the matrices afford no fitting parameters because they are exactly determined by geometry and symmetry. As the original mathematical proof of isospectrality made use of transplantation, this experimental observation is arguably the strongest confirmation of quantum isospectrality in our nanostructures. Perhaps the most tantalizing component of transplantation, however, appears when addressing the excited states. As is, their transplantation fails utterly (fig. S6) because we require a final crucial ingredient: the internal, quantum mechanical phase of the wave functions.

We recognized that our quantum invocation of Sunada's theorem is inherently phase-sensitive because it must act on ψ rather than $|\psi|^2$. Hence, our basic idea was to determine the phase of

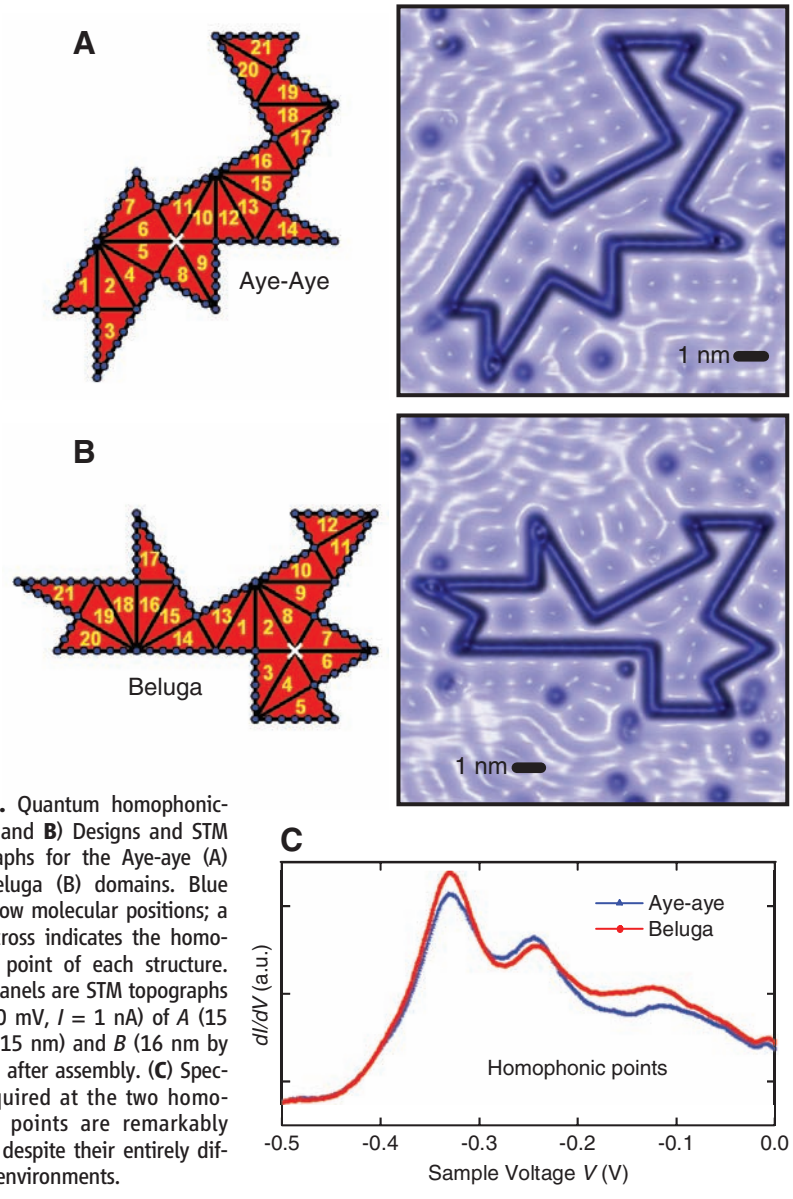


Fig. 3. Quantum homophonicity. (**A** and **B**) Designs and STM topographs for the Aye-aye (A) and Beluga (B) domains. Blue dots show molecular positions; a white cross indicates the homophonic point of each structure. Right panels are STM topographs (V = 10 mV, I = 1 nA) of *A* (15 nm by 15 nm) and *B* (16 nm by 16 nm) after assembly. (**C**) Spectra acquired at the two homophonic points are remarkably similar despite their entirely different environments.

measured wave function data inside one isospectral domain by optimizing its transplantation to the data for its partner domain. That is, we wished to minimize the error $\delta = \sum_{\mathbf{r}} ||\mathbf{T}|A(\mathbf{r})\rangle|^2 - ||B(\mathbf{r})\rangle|^2|$, summing over every data point in real space $\mathbf{r}$. Here, $||A\rangle|^2$ and $||B\rangle|^2$ are the measured probability densities determined from our *dI/dV* maps (*13*). In the homophonic structures (Fig. 3, A and B), we measured these probability densities for the first few eigenmodes at ~10^5 points each. Separating A into its amplitude (known) and phase (unknown), $|A\rangle = ||A(\mathbf{r})\rangle| \exp[i\phi(\mathbf{r})]$, we designed and operated a "quantum transplantation machine" (QTM) to extract the optimal phase $\phi(\mathbf{r})$ by minimizing δ (*13*). The QTM (fig. S3) has two inputs and two outputs, and consists of simple operations surrounding the transplantation operation $\mathbf{T}$ (*13*).

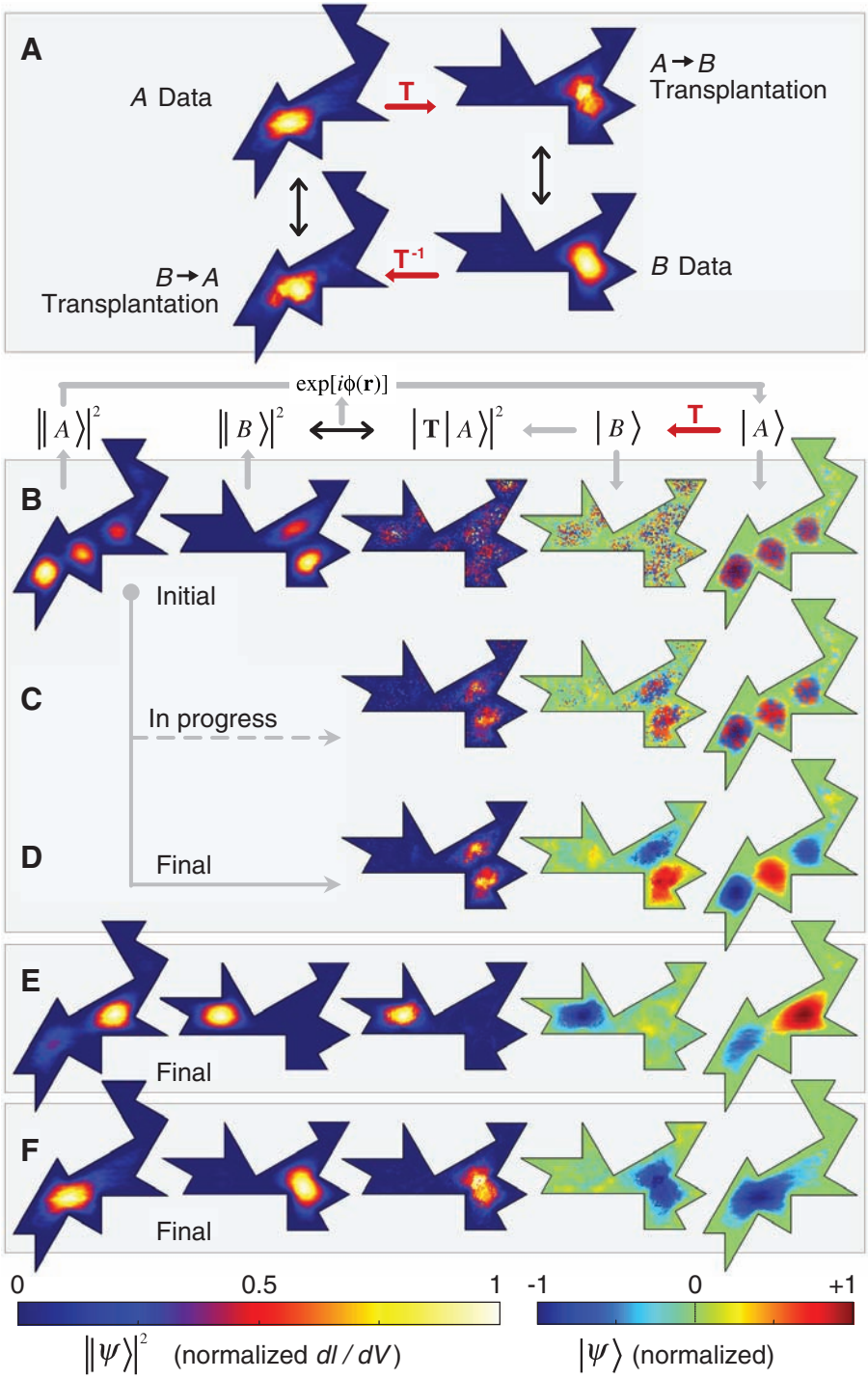


Fig. 4. (**A**) Quantum transplantation. Open-loop *dI/dV* map ($V = -0.329$ V) for the ground state of Aye-aye (upper left) is transformed by the transplantation operator $\mathbf{T}$ into a map that is an excellent match to the measured ground-state eigenfunction for Beluga (lower right, same parameters). Similarly, $\mathbf{T}^{-1}$ maps the *B* data onto the *A* shape, in excellent agreement with the measured *A* data. (**B** to **F**) Quantum phase extraction. (B) Initially, a random phase of 0 or π is assigned to every data point, and the transplantation $|\mathbf{T}|A\rangle|^2$ is a poor match for $||B\rangle|^2$. (C) An intermediate state, with phase extraction still in progress. (D) The final extracted wave functions for the third eigenmode of the two shapes. (E) Results for the second mode. (F) The results for the first mode match those in (A) because all parts of the wave function are in phase. Color scale bars are normalized to maximum values.

In this parameterization of the wave function, the magnitude is continuously varying, so the internal phase (modulo an overall arbitrary phase factor) can be expressed as a set $\phi(\mathbf{r}) \in \{0,\pi\}$ for all $\mathbf{r}$. That is, the quantum phase difference between any two points in a single wave function is a well-defined quantity and can only be 0 or π. This phase, or equivalently the sign information of the wave function $\exp[i\phi(\mathbf{r})] = \pm 1$, is normally irretrievably obscured by quantum measurement of a single wave function. The importance of $\phi(\mathbf{r})$ stems from its crucial role in superpositions and dynamics, as this phase evolves in time according to the time-dependent Schrödinger equation.

Determination of these internal signs of a wave function is a type of inverse problem in the class of binary optimization. To solve it, we began by assigning a random phase at every point; at this stage, $|A\rangle$ and $\mathbf{T}|A\rangle$ were speckled and irregular (Fig. 4B, initial QTM state for mode 3). Next, the phase was adjusted iteratively to minimize δ using a greedy algorithm (*13*, *31*). Figure 4C shows an intermediate state of the QTM while phase extraction is still in progress. Upon convergence, the final output of the QTM (Fig. 4D) is not only $|A\rangle$, but also $\mathbf{T}|A\rangle = |B\rangle$; that is, the full phase information of the wave functions of both isospectral structures has been determined. Similarly, the mode 2 and mode 1 electron wave functions extracted for both the *A* and *B* resonators are displayed in Fig. 4, E and F, respectively.

Movie S3 presents the full evolution of the QTM for each of these modes as the phase is extracted. The convergence of the QTM is quite robust, thanks to the acute phase sensitivity of transplantation. Time-reversal symmetry forces our wave functions to be real; hence, the only additional degree of freedom allowed in the phase extraction is an overall π phase shift, or sign change. In fact, we observe this vividly in the statistics of the extraction process: The QTM, for random initial phase seed, converges with 50% probability to each of two possible outcomes linked by a global $\exp(i\pi)$ phase factor. An example can be seen by comparing the QTM results for mode 2 in movie S3 with Fig. 4E.

Our work shows that phase extraction is surprisingly robust even in the presence of noise and imperfect boundary conditions—indicating accessibility in other systems—and we note that other realizations can be even simpler (fig. S1)

when not constrained by a lattice as in our experiment. The application of fundamental geometry to quantum systems is a growing field (*32*, *33*). Quantum phase extraction using topological symmetries has general applicability to the fields of quantum dots, nanoscale devices, and molecular electronics. The necessary criterion is geometric control on the scale of the relevant wavelength, already attainable in technologies such as nanolithography, self-assembly, and molecular design. Quantum dots, suitably patterned in semiconductors or metals, should mimic the physics described here, and calculations exist as a guideline for building isospectral systems (*34*). Measurements of nanomechanical resonators, now approaching the quantum limit (*35*), will similarly be phase-obscured and can benefit from these methods.

In chemistry, a complementary quest exists for isospectral and near-isospectral molecules, which can theoretically exist for a variety of different potentials but have not been experimentally observed (*36*, *37*). Just as this work allows the STM to be used as a scanning phase meter beyond charge sensitivity, we envision that other phase-sensitive experimental probes can be obtained by similar geometric tuning of quantum materials.

References and Notes

1. T. C. Weinacht, J. Ahn, P. H. Bucksbaum, *Phys. Rev. Lett.* **80**, 5508 (1998).
2. J. Itatani *et al*., *Nature* **432**, 867 (2004).
3. A. Yacoby, M. Heiblum, D. Mahalu, H. Shtrikman, *Phys. Rev. Lett.* **74**, 4047 (1995).
4. S. Marchesini, *Rev. Sci. Instrum.* **78**, 011301 (2007).
5. H. Weyl, *Nachr. Ges. Wiss. Goettingen* **1911**, 110 (1911).
6. H. P. McKean, I. M. Singer, *J. Differ. Geom.* **1**, 43 (1967).
7. M. Kac, *Am. Math. Mon.* **73**, 1 (1966).
8. J. Milnor, *Proc. Natl. Acad. Sci. U.S.A.* **51**, 542 (1964).
9. T. Sunada, *Ann. Math.* **121**, 169 (1985).
10. P. Buser, *Ann. Inst. Fourier* **36**, 167 (1986).
11. C. Gordon, D. Webb, S. Wolpert, *Inventiones Math.* **110**, 1 (1992).
12. B. Cipra, *Science* **255**, 1642 (1992).
13. See supporting material on *Science* Online.
14. S. Sridhar, A. Kudrolli, *Phys. Rev. Lett.* **72**, 2175 (1994).
15. C. Even, P. Pieranski, *Europhys. Lett.* **47**, 531 (1999).
16. E. Fradkin, J. E. Moore, *Phys. Rev. Lett.* **97**, 050404 (2006).
17. J. Kliewer *et al*., *Science* **288**, 1399 (2000).
18. P. Buser, J. Conway, P. Doyle, K. D. Semmler, *Int. Math. Res. Not.* **1994**, 391 (1994).
19. S. J. Chapman, *Am. Math. Mon.* **102**, 124 (1995).
20. M. F. Crommie, C. P. Lutz, D. M. Eigler, *Nature* **363**, 524 (1993).
21. J. A. Stroscio, D. M. Eigler, *Science* **254**, 1319 (1991).
22. M. F. Crommie, C. P. Lutz, D. M. Eigler, *Science* **262**, 218 (1993).
23. A. J. Heinrich, C. P. Lutz, J. A. Gupta, D. M. Eigler, *Science* **298**, 1381 (2002); published online 24 October 2002 (10.1126/science.1076768).
24. T. A. Driscoll, H. P. W. Gottlieb, *Phys. Rev. E* **68**, 016702 (2003).
25. A. A. Aligia, A. M. Lobos, *J. Phys. Condens. Matter* **17**, S1095 (2005).
26. E. J. Heller, M. F. Crommie, C. P. Lutz, D. M. Eigler, *Nature* **369**, 464 (1994).
27. Homophonicity is a stricter form of isospectrality. In this work, we study quantum isospectral nanostructures with Dirichlet-like boundary conditions. It has been proven that the Bilby/Hawk and Aye-aye/Beluga domains are isospectral even with Neumann boundary conditions, but homophonicity exists only within the latter pair and only for the Dirichlet case.
28. P. Bérard, *Math. Ann.* **292**, 547 (1992).
29. P. Bérard, *J. London Math. Soc.* **S2-48**, 565 (1993).
30. Tha pairs are not related by trivial transformations such as rotations, reflections, and translations.
31. T. H. Cormen, *Introduction to Algorithms* (MIT Press, Cambridge, MA, 2001).
32. L.-M. Duan, J. I. Cirac, P. Zoller, *Science* **292**, 1695 (2001).
33. M. A. Nielsen, M. R. Dowling, M. Gu, A. C. Doherty, *Science* **311**, 1133 (2006).
34. P. A. Knipp, T. L. Reinecke, *Phys. Rev. B* **54**, 1880 (1996).
35. M. D. LaHaye, O. Buu, B. Camarota, K. C. Schwab, *Science* **304**, 74 (2004).
36. W. C. Herndon, M. L. Ellzey, *Tetrahedron* **31**, 99 (1975).
37. E. Heilbronner, T. B. Jones, *J. Am. Chem. Soc.* **100**, 6506 (1978).

38. Supported by NSF grants CAREER DMR-0135122 and IMR DMR-0216913, U.S. Department of Energy grant DE-AC02-76SF00515, Office of Naval Research grant YIP/PECASE N00014-02-1-0351, Research Corporation grant RI0883, Stanford-IBM Center for Probing the Nanoscale grant NSF PHY-0425897, National Defense Science and Engineering Graduate program fellowships (C.R.M. and B.K.F.), and an Alfred P. Sloan Foundation fellowship (H.C.M.). We thank R. E. Schwartz, S.-H. Song, A. C. Manoharan, J. T. Moon, D. P. Arovas, M. Zworski, and M. R. Beasley for discussions, and R. G. Harris for expert technical assistance.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5864/782/DC1
SOM Text
Materials and Methods
Figs. S1 to S6
Movies S1 to S3
References

9 October 2007; accepted 18 December 2007
10.1126/science.1151490

Observation of the Spin Hall Effect of Light via Weak Measurements

Onur Hosten* and Paul Kwiat

We have detected a spin-dependent displacement perpendicular to the refractive index gradient for photons passing through an air-glass interface. The effect is the photonic version of the spin Hall effect in electronic systems, indicating the universality of the effect for particles of different nature. Treating the effect as a weak measurement of the spin projection of the photons, we used a preselection and postselection technique on the spin state to enhance the original displacement by nearly four orders of magnitude, attaining sensitivity to displacements of ~1 angstrom. The spin Hall effect can be used for manipulating photonic angular momentum states, and the measurement technique holds promise for precision metrology.

Hall effects, in general, are transport phenomena, in which an applied field on the particles results in a motion perpendicular to the field. Unlike the traditional Hall effect and its quantum versions, in which the effect depends on the electrical charge, the spin Hall effect is driven by the spin state of the particles. It was recently suggested (*1*, *2*) and observed (*3*) that, even in the absence of any scattering impurities, when an electric field is applied to a semiconductor, a dissipationless spin-dependent current perpendicular to the field can be generated. A photonic version of the effect—the spin Hall effect of light (SHEL)—was recently proposed (*4*) in which the spin-1 photons play the role of the spin-1/2 charges, and a refractive index gradient plays the role of the electric potential gradient.

We use an air-glass interface to demonstrate the SHEL, in which the transmitted beam of light splits by a fraction of the wavelength, upon refraction at the interface, into its two spin components (Fig. 1A): the component parallel ($s = +1$, right-circularly polarized) and antiparallel ($s = -1$, left-circularly polarized) to the central wave vector. This effect is different from (i) the previously measured (*5*) longitudinal Goos-Hänchen (*6*) and transverse Imbert-Fedorov (*7*, *8*) shifts in total internal reflection, which are described in terms of evanescent wave penetration, and (ii) the recently reported "optical spin Hall effect," which deals with optically generated spin currents of exciton-polaritons in a semiconductor microcavity (*9*). The splitting in the SHEL, implied by angular momentum conservation, takes place as a result of an effective spin-orbit interaction. The same interaction also leads to other effects such as the optical Magnus effect (*10*, *11*), the fine-splitting of the energy levels of an optical resonator (*12*) [in which the interaction resembles the spin-orbit (Russell-Saunders) coupling of electrons in atoms], and the deviation of photons from the simple geodesic paths of general relativity (*13*).

The exact amount of the transverse displacements due to the SHEL at an air-glass interface has been the subject of a recent debate (*4*, *14*–*16*). Our theory and experimental results support the predictions of Bliokh and Bliokh (*15*, *16*); although the calculations of other researchers (*4*, *11*, *14*) are not incorrect, they contain rather unfavorable initial conditions [see supporting online material (SOM)]. One can obtain close estimates of the magnitude of the displacements using solely the conservation of the z component of the total (spin plus orbital) angular mo-

Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

*To whom correspondence should be addressed. E-mail: hosten@uiuc.edu

mentum resulting from the rotational symmetry. However, only in certain circumstances will the result be exact (*16*). [For a general review of light's angular momentum, see (*17*).]

We can describe the SHEL as a consequence of a geometric phase (Berry's phase) (*18*), which corresponds to the spin-orbit interaction. It is already known that photons, when guided by an optical fiber with torsion, acquire a geometric phase whose sign is determined by the spin state (*19*, *20*). When a photonic wave packet changes direction because of a spatial variation in the refractive index, the plane-wave components with different wave vectors experience different geometric phases, affecting the spatial profile and resulting in the SHEL.

For a paraxial beam, the transverse beam state (for the relevant direction y and its associated wave vector k_y) at the air side of the interface (Fig. 1A), including the spin state $|s\rangle$, can be written as $|\Psi_a\rangle = \int dy\Psi(y)|y\rangle|s\rangle = \int dk_y\Phi(k_y)|k_y\rangle|s\rangle$, with $\Phi(k_y)$ being the Fourier transform of $\Psi(y)$. At the glass side of the interface, under the action of the geometric phases, the state becomes $|\Psi_g\rangle = \int dk_y\Phi(k_y)\exp(-ik_y\hat{\sigma}_3\delta)|k_y|s\rangle = \int dy\Psi(y - s\delta)|y\rangle|s\rangle$, with $\hat{\sigma}_3|s\rangle = s|s\rangle$, indicating $+\delta$ and $-\delta$ shifts for the wave packets of the parallel and antiparallel spin states. Here, the term $\exp(-ik_y\hat{\sigma}_3\delta)$ represents a coupling between the spin and the transverse momentum of the photons.

The origin of this "spin-orbit" interaction term lies in the transverse nature of the photon polarization: The polarizations associated with the plane-wave components experience different rotations in order to satisfy the transversality after refraction. This is depicted pictorially in Fig. 1B with incoming horizontal polarization ($|H\rangle$) (along x_I). In the spin basis, this state corresponds to $|H\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)$. In the lowest-order approximation, the change in the state after refraction is $|k_y\rangle|H\rangle \rightarrow |k_y\rangle(|H\rangle + k_y\delta|V\rangle) = |k_y\rangle|\varphi\rangle$, with $\varphi = k_y\delta << 1$ and $|V\rangle$ being vertical polarization. In the spin basis, $|\varphi\rangle = \frac{1}{\sqrt{2}}(\exp(-ik_y\delta)|+\rangle + \exp(ik_y\delta)|-\rangle)$, indicating the coupling $\exp(-ik_y\hat{\sigma}_3\delta)$.

As a result of the polarization-dependent Fresnel reflections at the interface, the opposite displacements of the two spin components actually depend on the input polarization state (see SOM for details). For $|H\rangle$ and $|V\rangle$ input polarizations, the displacements δ^H and δ^V are given by (Fig. 1C)

$$\delta^H_{|\pm\rangle} = \pm\frac{\lambda}{2\pi}\frac{\cos(\theta_T) - (t_s/t_p)\cos(\theta_I)}{\sin(\theta_I)},$$

$$\delta^V_{|\pm\rangle} = \pm\frac{\lambda}{2\pi}\frac{\cos(\theta_T) - (t_p/t_s)\cos(\theta_I)}{\sin(\theta_I)} \quad (1)$$

Here, θ_I and θ_T are, respectively, the central incident and transmitted angles related by Snell's law; t_s and t_p are the Fresnel transmission coefficients at θ_I; and λ is the wavelength of the light in the incident medium. In a continuously varying refractive index, the input polarization dependence disappears, and the motion can be formulated in terms of a particle moving in a vector potential in momentum space (*4*, *14*, *21*, *22*), along the same lines as with electronic systems (*23*, *24*).

For optical wavelengths, precise characterization of the displacements requires measurement sensitivities at the angstrom level. To achieve this sensitivity, we use a signal enhancement technique known from quantum weak measurements (*25*). In a quantum measurement, a property (observable $\hat{\mathbf{A}}$) of a system is first coupled to a separate degree of freedom (the "meter"), and then the information about the state of the observable is read out from the meter. At the single-photon level, the SHEL is actually equivalent to a quantum measurement of the spin projection along the central propagation direction (observable $\hat{\sigma}_3$, with eigenstates $|+\rangle$ and $|-\rangle$), with the transverse spatial distribution serving as the meter [similar to a Stern-Gerlach spin-projection measurement (*26*)]. However, the displacements generated by the SHEL here are much smaller than the width of the transverse distribution, resulting in a weak measurement: The meter states associated with different spin eigenstates overlap to a large extent. Therefore, the meter carries very little information about the state of the observable, leaving the initial state almost undisturbed. Although our experiment is at a classical level with a large number of photons in a quantum-mechanical coherent state, the results remain the same, with each photon behaving independently. Furthermore, in the paraxial regime, the dynamics of the transverse distribution are given by the Schrödinger equation with time replaced by path length, making the analysis identical to nonrelativistic quantum mechanics with an impulsive measurement interaction Hamiltonian $H_I = k_y\hat{\mathbf{A}}\delta$.

With the weak measurement taking place in between, the signal enhancement technique uses an appropriate preselection and postselection of the state of the observable to achieve an enhanced displacement in the meter distribution (*25*, *27*) (Fig. 2A). Given the preselected and postselected states $|\psi_1\rangle$ and $|\psi_2\rangle$, for sufficiently weak measurement strengths, the final position of the meter is proportional to the real part of the so-called "weak value" of the measured observable $\hat{\mathbf{A}}$

$$A_w = \frac{\langle\psi_2|\hat{\mathbf{A}}|\psi_1\rangle}{\langle\psi_2|\psi_1\rangle} \quad (2)$$

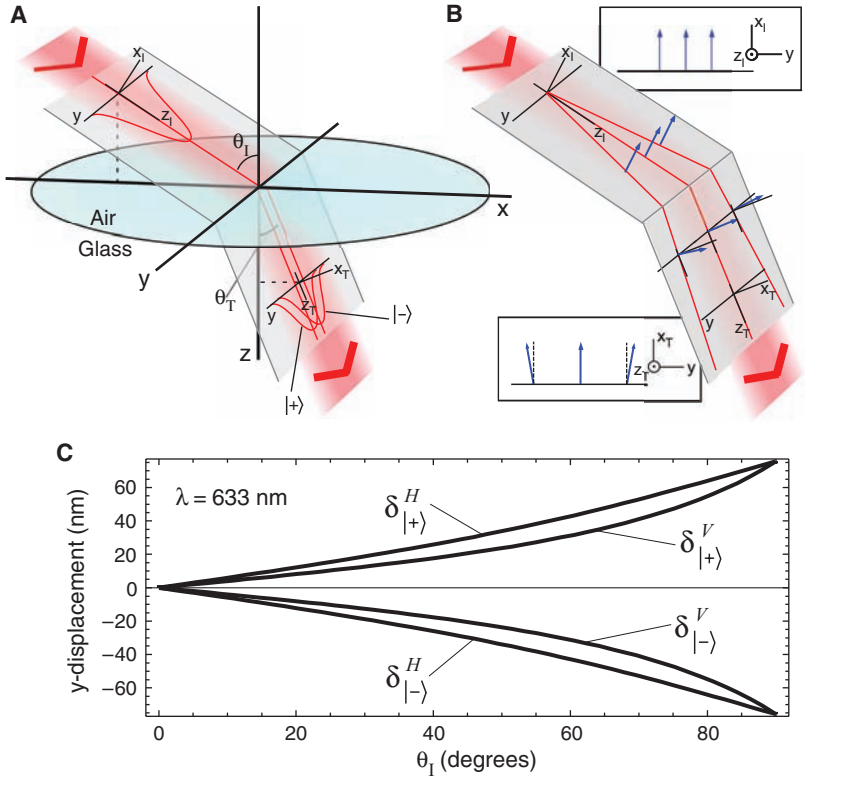


Fig. 1. The SHEL at an air-glass interface. (**A**) $|+\rangle$ and $|-\rangle$ spin components of a wave packet incident at angle θ_I experience opposite transverse displacements (not deflections) upon refraction at an angle θ_T. (**B**) Different plane-wave components acquire different polarization rotations upon refraction to satisfy transversality. The input polarization is in the x_I direction (equivalent to horizontal according to Fig. 3) for all constituent plane waves. Arrows indicate the polarization vectors associated with each plane wave before and after refraction. The insets clarify the orientation of the vectors. (**C**) Theoretical displacements of the spin components (Eq. 1) for horizontally and vertically polarized incident photons with wavelength λ = 633 nm.

which increases as the postselected state approaches being orthogonal to the preselected one. This effect was previously demonstrated (*28*) with real weak values, where a birefringent displacement of photons was enhanced by a factor of 20. The imaginary part of a weak value corresponds to a displacement in the momentum-space distribution of the meter, which with free evolution leads to the possibility of even larger enhancements. Furthermore, the order does not matter: The free evolution can take place first, followed by the weak measurement, but the final displacement will be identical (Fig. 2B). We describe the final displacement of the meter as the "modified weak value," $A_w^{mod} = F|A_w|$, where the factor F depends on the initial state of the meter and the amount of its free evolution before detection (see SOM).

In our setup (Fig. 3), the SHEL takes place at the front surface of a variable angle prism (VAP) for various incidence angles θ_I, with the back surface adjusted to be at normal incidence to avoid secondary Hall shifts (because there is obviously no Hall shift at normal incidence). The VAP is constructed by attaching two BK7 round wedge prisms together with the surface tension of a thin layer of index-matching fluid and is mounted loosely to avoid any stress-induced birefringence. The prisms can rotate with respect to each other, and the entire assembly can rotate around three orthogonal axes, allowing the desired surface orientations. The enhancement effect is achieved by preselecting and postselecting the polarization states of the incoming photons in states $|\psi_1\rangle$ and $|\psi_2\rangle$, by the calcite polarizers P1 and P2, respectively

$$|\psi_1\rangle = |H\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle),$$

$$|\psi_2\rangle = |V \pm \Delta\rangle = -i\exp(\mp i\Delta)|+\rangle + i\exp(\pm i\Delta)|-\rangle \quad (3)$$

With $\Delta << 1$ being a small angle, the weak value of the spin component is given by $(\hat{\sigma}_3)_w = \mp i\cot\Delta \approx \mp i/\Delta$. Lenses L1 and L2, respectively, focus and collimate the transverse spatial distribution of the incoming photons. The factor F in the modified weak value is determined by the transverse spatial state of the photons after lens L2 (see SOM)

$$F = \frac{4\pi\langle y_{L2}^2\rangle}{z_{eff}\lambda} \quad (4)$$

Here, $\langle y_{L2}^2\rangle$ is the variance of the y-direction transverse distribution after lens L2, and z_{eff} = 125 ± 5 mm is the effective focal length of L2. The design of the VAP assures that the optical path length inside of the VAP remains the same for all incidence angles, so that F is always the same. We achieve a displacement at the position sensor by an amount that is $(\hat{\sigma}_3)_w^{mod} = \mp F\cot\Delta \approx \mp F/\Delta$ times (~10^4 for our system) larger than the displacement caused by the SHEL

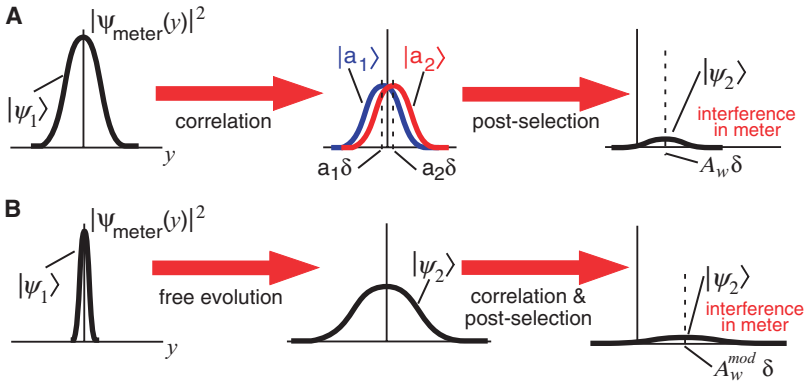


Fig. 2. A weak quantum measurement with preselection and postselection. (**A**) System starts in state $|\psi_1\rangle$. The measurement interaction weakly correlates the meter with the eigenstates of the measured observable $\hat{\mathbf{A}}$. A postselection on the system in state $|\psi_2\rangle$ gives rise to an interference in the meter, shifting it to its final position proportional to A_w (Eq. 2). (**B**) When A_w is imaginary, modifying the state of the meter by means of a free evolution (either before or after the measurement of $\hat{\mathbf{A}}$; the "before" condition is shown here) makes the final meter position proportional to A_w^{mod}, which can be much larger than A_w.

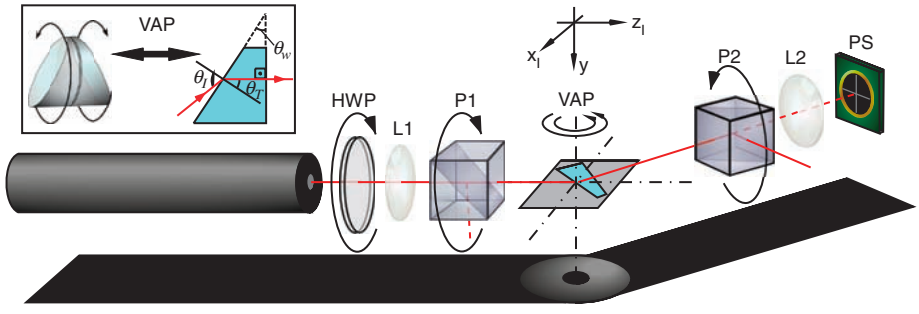


Fig. 3. Experimental setup for characterizing the SHEL. VAP (material: BK7; refractive index n = 1.515 at 633 nm); P1 and P2, Glan Laser polarizers; L1 and L2, lenses with effective focal lengths 25 and 125 mm, respectively; PS, position sensor (a split photodiode); HWP, half-wave plate (for adjusting the intensity after P1). The light source is a 10-mW linearly polarized He-Ne laser at 633 nm.

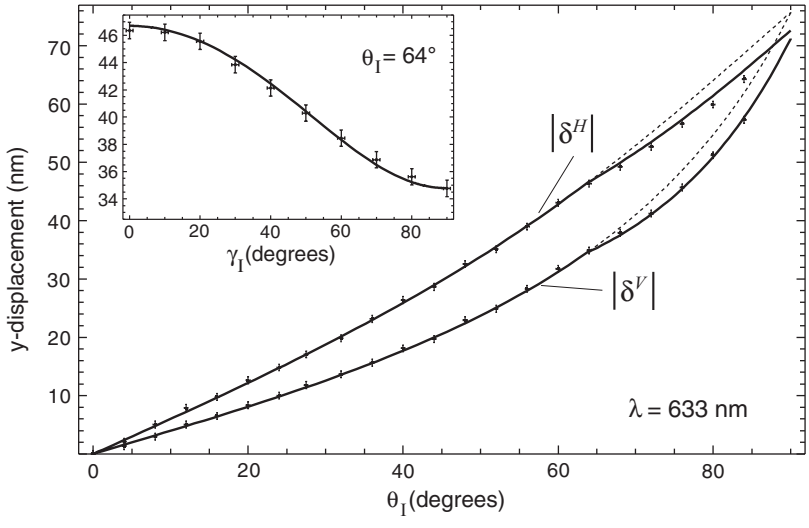


Fig. 4. Experimental results for the magnitude of the opposite shifts of the two spin components as a function of incidence angle. The data sets $|\delta^H|$ and $|\delta^V|$ are half the measured separation between the two spin components for horizontal and vertical incident polarization, respectively. Error bars represent two SDs. Dashed lines indicate the theory (Eq. 1), whereas solid lines indicate the modified theory (see text). The inset shows the results as a function of the incident polarization angle γ_I for a fixed incidence angle of 64°.

alone, and we measure this by reading out the difference between the displacements for the two postselections $|V\pm\Delta\rangle$.

In the results (Fig. 4), the dashed lines, which overlap with the solid lines up to 64°, represent the theory with the photons exiting the prism at normal incidence (Eq. 1). The solid lines also take into account the modifications (see SOM) arising from the fact that our VAP cannot satisfy the normal exiting condition beyond 64°. In order to stay in the linear response regime of both the enhancement technique and the position sensor, for incidence angles larger than 56°, the enhancement is adjusted (through the polarizer angle Δ) to be lower by a factor of 1.8: $|(\hat{\sigma}_3)_w| = 57.3 \pm 0.7$ for $\theta_I \leq 56°$, $|(\hat{\sigma}_3)_w| = 31.8 \pm 0.2$ for $\theta_I > 56°$. With the value of F determined from a single-parameter fit to be 156 ± 2, all the data points lie on the theoretical curves (Fig. 4) with a SD of 3.5 Å from the theory (excluding the last two data points of curve $|\delta^H|$, where there is a small, and as yet unexplained, discrepancy). By measuring $\langle y_{L2}^2 \rangle$, we can also experimentally determine the value of F, albeit less accurately; we find 157 ± 6, indicating the total enhancement factor to be $(\hat{\sigma}_3)_w^{mod} = (8.97 \pm 0.36) \times 10^3$ for $\theta_I \leq 56°$ and $(\hat{\sigma}_3)_w^{mod} = (4.99 \pm 0.20) \times 10^3$ for $\theta_I > 56°$. Figure 4 also shows the dependence of the displacements on the incident linear polarization (see SOM for the theory). For arbitrary incident polarization (other than $|H\rangle$ or $|V\rangle$), a kick in the y momentum is also expected; however, our measurements are not sensitive to this effect (see SOM).

The measurability of very small displacements is ultimately limited by the quantum noise of the light, because enough photons need to be collected to resolve the position of the transverse distribution (*29*). If the experiment is already quantum noise–limited, then (for a fixed number of photons entering the experimental setup) the enhancement due to preselection and postselection does not bring any gain in the measurement sensitivity: For any value of F, although the displacement is multiplied by a factor of $|(\hat{\sigma}_3)_w| \approx 1/\Delta$, only a fraction of the photons ($|\langle\psi_1|\psi_2\rangle|^2 = \sin^2 \Delta \approx \Delta^2$) makes it through the postselection, which cancels the advantage of the enhancement. Nevertheless, most experiments are limited by technical issues [e.g., in our setup, by the laser pointing stability, the intensity saturation of the position sensor, or the unwanted (~10-μm) displacements caused by rotating the polarizers]. In the experiment, we suppress all these technical issues by a factor of $(\hat{\sigma}_3)_w^{mod} \approx 10^4$ with respect to the signal and achieve a dc sensitivity/stability to displacements of ~1 Å, without the need for vibration or air-fluctuation isolation. The upper limit to the enhancement comes from the achievable extinction ratio of the polarizers and from the electronic noise in the position sensor.

The overall sensitivity of our measurements can be increased by orders of magnitude, by incorporating standard signal modulation and lock-in detection techniques, and thus holds great promise for precision metrology. In fact, in principle, Hall displacements themselves can be made arbitrarily large, separating beams associated with different spin states (*4*, *21*) [and also with different orbital angular momentum states (*22*)]. In addition, the SHEL itself may become an advantageous metrological tool (e.g., for characterizing refractive index variations measured at subwavelength distances from the region of interest, because any Hall displacement accompanying a beam deflection is larger than the displacement caused by the deflection itself at such propagation distances).

References

1. J. Sinova *et al.*, *Phys. Rev. Lett.* **92**, 126603 (2004).
2. S. Murakami, N. Nagaosa, S. C. Zhang, *Science* **301**, 1348 (2003).
3. J. Wunderlich, B. Kaestner, J. Sinova, T. Jungwirth, *Phys. Rev. Lett.* **94**, 047204 (2005).
4. M. Onoda, S. Murakami, N. Nagaosa, *Phys. Rev. Lett.* **93**, 083901 (2004).
5. F. Pillon, H. Gilles, S. Girard, *Appl. Opt.* **43**, 1863 (2004).
6. F. Goos, H. Hänchen, *Ann. Phys. (Leipzig)* **1**, 333 (1947).
7. F. I. Fedorov, *Dokl. Akad. Nauk SSSR* **105**, 465 (1955).
8. C. Imbert, *Phys. Rev. D* **5**, 787 (1972).
9. C. Leyder *et al.*, *Nat. Phys.* **3**, 628 (2007).
10. A. V. Dooghin, N. D. Kundikova, V. S. Liberman, B. Ya. Zel'dovich, *Phys. Rev. A* **45**, 8204 (1992).
11. V. S. Liberman, B. Ya. Zel'dovich, *Phys. Rev. A* **46**, 5199 (1992).
12. K. Yu. Bliokh, D. Yu. Frolov, *Opt. Commun.* **250**, 321 (2005).
13. P. Gosselin, A. Berard, H. Mohrbach, *Phys. Rev. D* **75**, 084035 (2007).
14. M. Onoda, S. Murakami, N. Nagaosa, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **74**, 066610 (2006).
15. K. Yu. Bliokh, Y. P. Bliokh, *Phys. Rev. Lett.* **96**, 073903 (2006).
16. K. Yu. Bliokh, Y. P. Bliokh, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **75**, 066609 (2007).
17. M. Padgett, J. Courtial, L. Allen, *Phys. Today* **57**, 35 (2004).
18. M. V. Berry, *Proc. R. Soc. London Ser. A* **392**, 45 (1984).
19. R. Y. Chiao, Y. S. Wu, *Phys. Rev. Lett.* **57**, 933 (1986).
20. A. Tomita, R. Y. Chiao, *Phys. Rev. Lett.* **57**, 937 (1986).
21. K. Yu. Bliokh, Y. P. Bliokh, *Phys. Lett. A* **333**, 181 (2004).
22. K. Yu. Bliokh, *Phys. Rev. Lett.* **97**, 043901 (2006).
23. M.-C. Chang, Q. Niu, *Phys. Rev. B* **53**, 7010 (1996).
24. G. Sundaram, Q. Niu, *Phys. Rev. B* **59**, 14915 (1999).
25. Y. Aharonov, L. Vaidman, *Phys. Rev. A* **41**, 11 (1990).
26. J. A. Wheeler, W. H. Zurek, Eds. *Quantum Theory and Measurement* (Princeton Univ. Press, Princeton, NJ, 1983).
27. J. Tollaksen, *J. Phys. Conf. Ser.* **70**, 012015 (2007).
28. N. W. M. Ritchie, J. G. Story, R. G. Hulet, *Phys. Rev. Lett.* **66**, 1107 (1991).
29. N. Treps *et al.*, *Science* **301**, 940 (2003).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1152697/DC1
SOM Text
Fig. S1
References

7 November 2007; accepted 27 December 2007
Published online 10 January 2008;
10.1126/science.1152697
Include this information when citing this paper.

Bond-Selective Control of a Heterogeneously Catalyzed Reaction

Daniel R. Killelea,* Victoria L. Campbell, Nicholas S. Shuman,† Arthur L. Utz‡

Energy redistribution, including the many phonon-assisted and electronically assisted energy-exchange processes at a gas-metal interface, can hamper vibrationally mediated selectivity in chemical reactions. We establish that these limitations do not prevent bond-selective control of a heterogeneously catalyzed reaction. State-resolved gas-surface scattering measurements show that the ν_1 C-H stretch vibration in trideuteromethane (CHD_3) selectively activates C-H bond cleavage on a Ni(111) surface. Isotope-resolved detection reveals a CD_3:CHD_2 product ratio > 30:1, which contrasts with the 1:3 ratio for an isoenergetic ensemble of CHD_3 whose vibrations are statistically populated. Recent studies of vibrational energy redistribution in the gas and condensed phases suggest that other gas-surface reactions with similar vibrational energy flow dynamics might also be candidates for such bond-selective control.

Chemical bonds participate in a richly choreographed dance of stretches and bends during reaction, and the selective excitation of key reagent vibrations before reaction has long been viewed as a potentially powerful strategy for achieving bond-selective chemistry (*1*). Experiments on partially deuterated water (HOD) validated this approach for unimolecular (*2*) and bimolecular (*3*, *4*) reactions in the gas phase, and more recent studies have extended it to polyatomic molecules in the gas phase (*5*–*8*) and unimolecular reactions of adsorbates on surfaces (*9*–*12*). Vibrational excitation in inelastic electron tunneling spectroscopy (IETS) has promoted rotation of adsorbed acet-

Department of Chemistry and W. M. Keck Foundation Laboratory for Materials Chemistry, Tufts University, Medford, MA 02155, USA.

*Present address: James Franck Institute, University of Chicago, Chicago, IL 60637, USA.
†Present address: Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
‡To whom correspondence should be addressed. E-mail: arthur.utz@tufts.edu

ylene on Cu(100) (*9*), promoted diffusion of CO on Pd(110) (*10*), and uncovered a vibrational mode-specific preference for diffusion versus desorption of ammonia on Cu(100) (*11*). Resonant infrared (IR) excitation of the Si-H bond enhanced H desorption from an H/D-covered Si(111) surface via a nonthermal process, although the mechanism for enhancement is not yet fully understood (*12*). It remains unclear, though, whether these strategies are transferable to bimolecular reactions between a laser-excited gas-phase reagent and a solid surface. Such reactions are central to heterogeneous catalysis and vapor deposition. We address that question with experiments that clearly establish the ability of gas-phase vibrational excitation to enable bond-selective C-H cleavage in the heterogeneously catalyzed dissociative chemisorption of CHD_3 on a Ni surface.

Vibrationally mediated bond-selective chemistry is intuitively appealing, but several factors complicate its practical implementation (*13*–*15*). The vibrational eigenstates produced by optical or inelastic electron tunneling excitation are most often concerted excitations of two or more bonds, but the reaction coordinate for dissociation or atom abstraction is localized in a single bond. Coherently exciting two or more eigenstates can prepare a superposition state whose excitation is localized, but intramolecular vibrational energy redistribution (IVR)—a process that disperses the superposition state's energy into other parts of the molecule—competes with the reaction and can limit or prevent bond-selective control. Synchronizing excitation of the rapidly dephasing superposition state with the reactive collision is also difficult. In favorable cases, a vibrational eigenstate closely resembles a localized bond stretch vibration. Excitation of that state is long-lived in the gas phase, but the reactive collision perturbs the molecule, induces IVR, and can still thwart bond-selective control.

We explore whether bond-selective control is possible in a heterogeneously catalyzed reaction. Controlling reactivity in this way depends on the vibrational structure of the reagent, the patterns and time scales of IVR, the structure of the transition state, and the role of the metal surface. CHD_3 is particularly well suited as a probe of these effects because it has chemically distinguishable bonds, multiple IVR pathways, and an eigenstate (the ν_1 normal mode) with a very localized C-H stretch (*16*). Because narrow bandwidth laser excitation of ν_1 prepares a single vibrational eigenstate, IVR does not take place in the isolated molecule. There is minimal collisional quenching in the molecular beam, and the state's IR radiative lifetime is long, so excitation remains localized in the C-H bond until the gas-surface collision. The IR- and Raman-accessible C-H stretching vibrations of other methane isotopologues are collective excitations of two or more C-H bonds. Collision-induced IVR may localize excitation (*17*, *18*), but there is no way to identify which one of the C-H bonds broke.

The IVR pathways and rates for partially deuterated methane molecules permit mode- and bond-selective atom abstraction chemistry in the gas phase. C-H stretch excitation in CHD_3 (*19*) and in CH_3D (*7*, *20*) leads to H-atom abstraction by Cl, whereas C-D excitation leads to D-atom abstraction. A simple spectator model provides a qualitative explanation; excitation localized in the reactive bond promotes transition-state access and bond cleavage, whereas excitation in the bends and stretches of nonreactive (spectator) bonds correlates with product excitation. Despite their near degeneracy, the symmetric and antisymmetric C-H stretches in CH_3D differ in reactivity. This difference arises from a collision-induced IVR process that preferentially localizes symmetric C-H stretch excitation into the C-H bond nearest the Cl atom (*18*). Reactivity patterns from these studies underscore the subtle but important role of IVR. Collision-induced IVR is fast enough to localize C-H stretch excitation and cause mode-specific reactivity, but IVR between C-H and C-D stretches is relatively slow and enables bond-selective reactivity.

Methane's dissociative chemisorption on Ni is well studied because of its role as the rate-limiting step in the industrial production of H_2. C-H stretch excitation promotes methane activation, and the many phonon-mediated and electronically mediated energy-exchange channels available at the gas-metal interface do not prevent mode-selective chemistry. Methane dissociates on Ni(111) via single C-H bond cleavage to form surface-bound methyl and H fragments (*21*). Vibrational state–resolved gas-surface scattering studies of CH_4 on Ni(100) show that the symmetric (ν_1) (*22*) and antisymmetric (ν_3) (*23*, *24*) C-H stretching states in CH_4 enhance reactivity, in accord with calculations that predict substantial C-H bond elongation at the transition state (*25*–*27*). Reactivity is mode-specific, with a relative vibrational efficacy of $\nu_1 > \nu_3 > \nu_4$ (triply degenerate bend) (*22*, *28*, *29*). CH_2D_2 dissociation on Ni(100) is also mode-specific. The |0,2> local mode state (two quanta in a single C-H bond) is more reactive than the |1,1> state (one quantum in each C-H bond). This result is consistent with the spectator model, but experiments did not distinguish the methyl isotopologues, which prevented an assessment of bond selectivity (*30*).

In our experiments, we prepared a supersonic molecular beam of CHD_3 molecules, exposed the beam to a clean Ni(111) surface, and quantified the surface-bound products of C-H (chemisorbed H and CD_3) and C-D (chemisorbed D and CHD_2) bond cleavage. Our earlier work

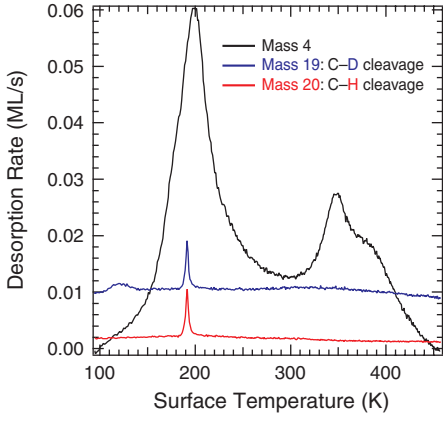


Fig. 1. Quantifying CHD_3 reaction products with TPD.

Table 1. Experimental parameters and measured reaction probabilities for CHD_3 ensembles with thermally populated vibrational states. Translational energies for the 2% CHD_3/He beams are from measured time-of-flight data, and vibrational energies are calculated for thermal ensembles at T_{nozzle}. Surface temperature was 90 K for all measurements. Error is based on the SD of three to five replicate measurements.

T_{nozzle} (K)	E_{trans} (kJ/mol)	E_{vib} (kJ/mol)	$S_0 \pm 2\sigma$
550	52.7	4.5	$(8.8 \pm 3.0) \times 10^{-6}$
600	57.4	5.9	$(2.6 \pm 1.5) \times 10^{-5}$
700	67.0	9.2	$(1.8 \pm 0.7) \times 10^{-4}$
830	79.8	14.4	$(8.3 \pm 4.5) \times 10^{-4}$
900	86.2	17.6	$(2.0 \pm 1.1) \times 10^{-3}$

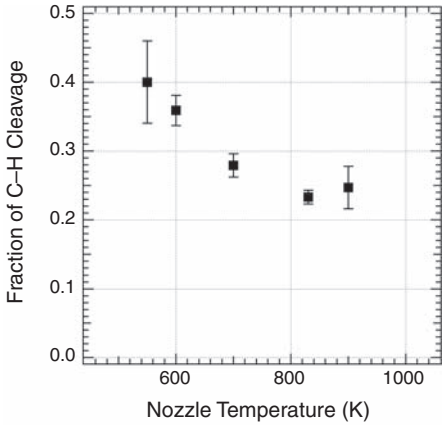


Fig. 2. Fraction of CHD_3 molecules dissociating via the C-H bond cleavage channel as a function of T_{nozzle}. All beams are 2% CHD_3 in He, so both E_{trans} and E_{vib} increase with T_{nozzle}. Error bars are $\pm 2\sigma$ from replicate measurements.

quantified reaction products with Auger electron spectroscopy (AES) of carbon, but AES cannot distinguish CD_3 and CHD_2 (*31*). Instead, we used a temperature-programmed desorption (TPD) method to titrate and quantify surface-bound methyls. Johnson *et al.* showed that surface-bound H (or D) does not react with surface-bound methyl to form methane but that D atoms embedded in the Ni lattice react quantitatively with surface-bound methyl to form methane, which promptly desorbs (*32*). We confirmed that the surface-bound H or D dissociative chemisorption products do not react with methyl and that no isotope scrambling occurs; all desorbing methane originates from deuteration by a single subsurface D atom.

We performed the experiments in a supersonic molecular beam–surface scattering machine (*31*, *33*). We first exposed a clean Ni(111) surface to D atoms, which covered the surface and preloaded subsurface lattice sites with D atoms. Collision-induced recombinative desorption with Xe removed all but 0.03 monolayers (ML) of surface-bound D, created a clean surface for CHD_3 dissociation, and preserved about three ML of subsurface D for post-CHD_3–dose titration. A supersonic molecular beam of CHD_3 then impinged on the surface. State-resolved experiments used a single-mode IR laser to excite CHD_3 to $v' = 1$, $J' = 2$, $K' = 0$ of ν_1 via the R(1)($\Delta K = 0$) transition at 3005.538 cm^{-1} (*34*). The vibrational energy (E_{vib}) is 36 kJ/mol, J' is the total angular momentum of the excited state, and K' is the projection of J' along the unique inertial axis. We held the surface temperature (T_{surf}) below 120 K to stabilize subsurface D and prevent methyl dehydrogenation (*21*). After CHD_3 chemisorption, we gradually raised the surface temperature in a TPD experiment. D atoms that were pre-embedded in the Ni lattice moved to the surface, titrated surface methyls, and formed methane. A mass spectrometer monitored desorption of D_2 (4 atomic mass units), CHD_3 (19 amu), and CD_4 (20 amu) during TPD, as shown in Fig. 1. Near T_{surf} = 190 K, D atoms combined with surface-bound CD_3 and CHD_2 to form CD_4 and CHD_3, respectively. Because CD_4 (20 amu) cannot fragment to the parent mass for CHD_3 (19 amu), integrated peak areas for 19 and 20 amu provided a direct measure of C-D and C-H bond cleavage products. AES measurements related TPD integrals to surface coverage.

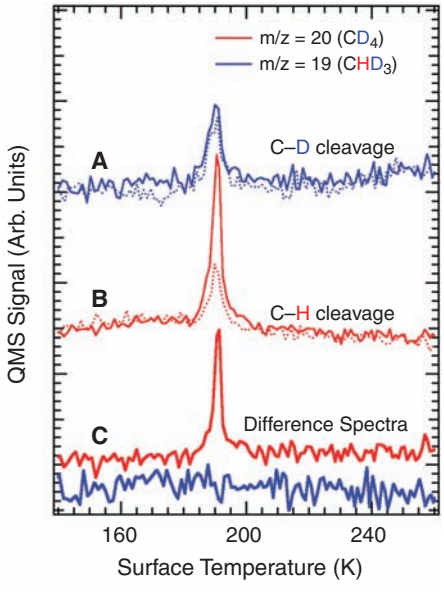


Fig. 3. TPD traces for D-atom titration of CHD_3 reaction products from C-D (**A**) or C-H (**B**) bond cleavage. Data are shown for experiments with (solid lines) and without (dashed lines) laser excitation of the ν_1 C-H stretch (T_{nozzle} = 600 K and E_{trans} = 57 kJ/mol). The difference spectra (**C**) show the net yield of C-H (red line) and C-D (blue line) bond cleavage products that arise solely from the laser-excited molecules. QMS, quadrupole mass spectrometer; Arb. Units, arbitrary units.

To assess whether ν_1 C-H stretch excitation enhances C-H bond cleavage, we first established a point of comparison for our state-resolved measurements. Molecular ensembles that statistically represent the relevant energetic configurations of reagents yield products that are consistent with statistical theories of reactivity, even when the molecular-level requirement for statistical energy redistribution is not met. This likely explains why microcanonical unimolecular rate theory models vibrational state-averaged reactivity of CH_4 on Ni well, even while it fails to predict state-resolved observations of mode-specific or bond-selective chemistry (*35*). Inefficient vibrational cooling in our supersonic expansion produced a nearly thermal (statistical) distribution of vibrational states corresponding to the nozzle source temperature (T_{nozzle}). We measured the product yield of these thermally excited ensembles (without laser excitation) to obtain an experimental estimate of statistical reactivity.

Figure 1 shows TPD data for a CHD_3/He beam expanded at T_{nozzle} = 600 K without laser excitation. We integrated the 19- and 20-amu TPD peaks to obtain the coverage of CHD_2 and CD_3 methyls and calculated the reaction probability summed over the CHD_2 and CD_3 product channels for five different T_{nozzle} values (Table 1). The fraction of dissociation events occurring via C-H bond cleavage is plotted in Fig. 2. The relative yield of C-H cleavage asymptotically approaches 25% as total energy (E_{total}) increases, and C-D cleavage is always the majority channel.

Three factors influenced the relative product yield for experiments with thermally populated vibrations. First, C-H bonds are more reactive than C-D bonds as a result of the kinetic isotope effect and possible tunneling contributions. Second, only 25% of randomly oriented CHD_3 molecules place their C-H bond in the most energetically favorable reaction geometry. Third, the lower frequency of C-D vibrations leads to higher excited-state populations relative to their C-H counterparts. Factor 1 favors C-H cleavage, whereas factors 2 and 3 favor C-D cleavage. The data in Fig. 2 reflect these trends. At low T_{nozzle}, system energy is lowest, isotope effects are most pronounced, and the fraction of C-H cleavage products is highest. As T_{nozzle} increases, translational energy (E_{trans}) and E_{vib} increase, isotope effects diminish, and geometric factors dominate, leading to the geometric limit of 25% C-H bond cleavage.

We now turn to state-resolved reactivity measurements of CHD_3 excited to $v = 1$ of the ν_1 C-H stretch. The laser excites about 1% of CHD_3 to the target eigenstate. We use (i) the ensemble-averaged S_0 for molecular beams with ($S_0^{\text{laser on}}$) and without ($S_0^{\text{laser off}}$) laser excitation, (ii) the fraction of molecules that we excite in the beam (f_{exc}), and (iii) the reactivity of the vibrational ground state ($S_0^{v=0}$) (which is negligible here) to obtain the reaction probability for the laser-excited molecules ($S_0^{v_1}$) (*24*)

$$S_0^{v_1} = \frac{S_0^{\text{laser on}} - S_0^{\text{laser off}}}{f_{exc}} + S_0^{v=0} \qquad (1)$$

Figure 3, A and B, shows TPD traces for mass/charge ratios (m/z) = 19 (D + CHD_2) and 20 (D + CD_3) after experiments with (solid lines) and without (dashed lines) laser excitation of ν_1. The yield of C-H cleavage products detected at 20 amu increases, but the C-D cleavage product is unchanged. (Laser excitation only depletes $v = 0$, which contributes negligibly toward reactivity at T_{nozzle} = 600 K.) Following Eq. 1, we subtracted the laser-off signal from the laser-on signal to obtain a difference signal (Fig. 3C) that reveals the reactivity of the laser-excited molecules. Within our detection limit, laser-excited molecules dissociated exclusively via C-H bond cleavage. Analysis of replicate measurements established a C-H:C-D bond cleavage ratio of at least 30:1 for CHD_3 excited to the ν_1 C-H stretching state.

To assess the extent of bond selectivity, we compared the relative product yield of CHD_3 (ν_1) with that of an isoenergetic ensemble whose vibrations were thermally populated. Table 2 shows that a 2% CHD_3/He beam with T_{nozzle} = 830 K has nearly the same E_{total} as the state-

Table 2. Relative product yield for state-resolved and thermal ensembles of CHD_3.

System	E_{trans} (kJ/mol)	E_{vib} (kJ/mol)	E_{total} (kJ/mol)	C-H:C-D ratio
State-resolved ν_1	57.4	36	93.4	>30:1
Thermal ensemble, T_{nozzle} = 830 K	79.8	14.4	94.2	1:3

resolved beam and a C-H:C-D product ratio of 1:3. Therefore, selective excitation of the ν_1 C-H stretch increases the relative yield of C-H bond cleavage products from laser-excited molecules by at least 90 fold relative to an ensemble of molecules whose vibrations are thermally populated.

Comparative studies of IVR rates in the gas phase and in solution may explain the bond selectivity that we observe. IVR often occurs on multiple time scales, with the most rapid (initial) component corresponding to energy flow into strongly coupled "doorway" states that then relax more slowly into the full vibrational density of states. A series of polyatomic molecules, whose vibrational state densities ranged from 0.7 to 21,000 states cm^{-1}, exhibited initial IVR lifetimes for a C-H stretch that ranged from 4 to 400 ps in the gas phase (*36*). The lifetimes did not correlate with the molecule's state density but instead were likely due to coincidental vibrational resonances that mediated IVR into the doorway states. IVR lifetimes for the same molecules were then measured in a range of solvents. The initial IVR rates barely changed upon solvation, which suggests that the inherent vibrational structure and resonances in a molecule dictate its IVR rates, not the nature, environment, or collision frequency of solvent-solute interactions.

When viewed in the context of our gas-surface scattering studies, we suggest that the gas-surface interaction initiates, but does not necessarily accelerate, IVR. In all collision processes, the approach of a collision partner—here, a highly polarizable metal surface—perturbs the molecule's vibrational structure by altering the potential energy term in the Hamiltonian. Thus, the vibrational eigenstates of CHD_3 evolve as it begins to interact with the surface (*17*); the laser-prepared gas-phase eigenstate becomes a superposition state in the basis of vibrational eigenstates for the CHD_3–surface complex, and IVR ensues. The interaction time of an energetic CHD_3 molecule (E_{trans} = 57 kJ/mol) with the Ni(111) surface is about 0.3 ps. Even large polyatomic molecules have IVR lifetimes for C-H stretching states of 4 ps or greater, so there is simply not enough time for IVR to transfer energy from the C-H stretching state into nonresonant C-D vibrations. As the adsorbate-substrate interaction time increases as a result of reduced collision energy or trapping (physisorption) on the surface, IVR may become more extensive, and the ability to achieve bond-selective control could diminish.

The observation of vibrationally mediated bond-selective chemistry in heterogeneous catalysis is important in several ways. First, it establishes that bond-selective chemistry is possible in a complex chemical system that includes a metal surface with its high vibrational state density (*37*) and efficient electronic channels for vibrational quenching and energy exchange (*38*, *39*). Second, although most examples of bond-selective chemistry occur in low-barrier reactions where E_{vib} provides most of the activation energy, we show that bond-selective surface chemistry persists even when E_{vib} provides less than 40% of the total reagent energy. Third, reduced-dimensionality models for methane dissociation have long modeled CH_4 as a pseudo-diatomic molecule, even though its gas-phase C-H stretches differ substantially from those of an isolated bond stretch (*40*, *41*). The dramatic bond selectivity that we observe suggests that CHD_3 (ν_1) may be an excellent benchmark for comparison with reduced-dimensionality quantum dynamics models. Fourth, whereas previous studies established nonstatistical, vibrational mode-specific behavior in methane's dissociative chemisorption, this work refines the picture of how IVR influences the gas-surface encounter. It suggests that IVR between relatively decoupled motions within a molecule can be slow enough to permit highly selective chemical control. In contrast, fast IVR among the degenerate C-H stretches in CH_4 has been suggested as a plausible explanation for the vibrational mode selectivity observed for ν_1 and ν_3 on Ni(100). Thus, it may be important to specify which IVR process is operative before drawing conclusions about the relative rates of IVR and reaction.

It is interesting to compare our work with previous IETS and IR photodesorption studies of vibrationally mediated chemistry on surfaces. Our beam-surface experiments provide E_{trans} to the reagent, so a single quantum of vibrational excitation is adequate to activate reaction. In contrast, surface-bound species at low T_{surf} lack significant E_{trans} and generally require a combination of multiquanta excitation (*11*) or energy pooling (*12*) to accumulate enough energy to react. The nature of the prepared vibration also differs in the experiments. Adsorption perturbs bonding in the adsorbate and introduces strong coupling to substrate atoms. The presence of a scanning tunneling microscope tip and the electric field that it induces can further perturb an adsorbate's vibrational states. Consequently, the vibrational states of surface adsorbate and the resulting nuclear motions induced upon excitation can differ markedly from the vibrational eigenstates of a gas-phase reagent.

Though previous IETS experiments provide unparalleled control over the local environment of the adsorbates, beam-surface studies offer greater control over the initiation and timing of energy flow during reaction because (i) the energy and vibrational character of the gas-phase excitation is well defined, (ii) IVR is only initiated upon collision, and (iii) the temporal window for reaction is limited by the short surface residence time: an experimental variable that depends on incident angle and speed. Despite their differences, these studies point to the breakdown of the rapid IVR assumption inherent to statistical theories of gas-surface reactivity (*37*), and they reveal both challenges and opportunities for understanding and manipulating surface reactivity.

References and Notes

1. F. F. Crim, *Science* **316**, 1707 (2007).
2. F. F. Crim, *Science* **249**, 1387 (1990).
3. A. Sinha, M. C. Hsaio, F. F. Crim, *J. Chem. Phys.* **92**, 6333 (1990).
4. M. J. Bronikowski, W. R. Simpson, B. Girard, R. N. Zare, *J. Chem. Phys.* **95**, 8647 (1991).
5. Z. H. Kim, H. A. Bechtel, R. N. Zare, *J. Am. Chem. Soc.* **123**, 12714 (2001).
6. S. Yoon, R. J. Holiday, F. F. Crim, *J. Chem. Phys.* **119**, 4755 (2003).
7. S. Yoon, R. J. Holiday, F. F. Crim, *J. Phys. Chem. B* **109**, 8388 (2005).
8. S. Yan, Y. T. Wu, B. L. Zhang, X. F. Yue, K. P. Liu, *Science* **316**, 1723 (2007).
9. B. C. Stipe, M. A. Rezaei, W. Ho, *Phys. Rev. Lett.* **81**, 1263 (1998).
10. T. Komeda, Y. Kim, M. Kawai, B. N. J. Persson, H. Ueba, *Science* **295**, 2055 (2002).
11. J. I. Pascual, N. Lorente, Z. Song, H. Conrad, H. P. Rust, *Nature* **423**, 525 (2003).
12. Z. Liu, L. C. Feldman, N. H. Tolk, Z. Zhang, P. I. Cohen, *Science* **312**, 1024 (2006).
13. F. F. Crim, *J. Phys. Chem.* **100**, 12725 (1996).
14. F. F. Crim, *Acc. Chem. Res.* **32**, 877 (1999).
15. R. N. Zare, *Science* **279**, 1875 (1998).
16. E. Venuti, L. Halonen, R. G. Della Valle, *J. Chem. Phys.* **110**, 7339 (1999).
17. L. Halonen, S. L. Bernasek, D. J. Nesbitt, *J. Chem. Phys.* **115**, 5611 (2001).
18. J. R. Fair, D. Schaefer, R. Kosloff, D. J. Nesbitt, *J. Chem. Phys.* **116**, 1406 (2002).
19. J. P. Camden, H. A. Bechtel, D. J. A. Brown, R. N. Zare, *J. Chem. Phys.* **124**, 034311 (2006).
20. R. J. Holiday, C. H. Kwon, C. J. Annesley, F. F. Crim, *J. Chem. Phys.* **125**, 133101 (2006).
21. M. B. Lee, Q. Y. Yang, S. T. Ceyer, *J. Chem. Phys.* **87**, 2724 (1987).
22. P. Maroni *et al.*, *Phys. Rev. Lett.* **94**, 246104 (2005).
23. R. R. Smith, D. R. Killelea, D. F. DelSesto, A. L. Utz, *Science* **304**, 992 (2004).
24. L. B. F. Juurlink, P. R. McCabe, R. R. Smith, C. L. DiCologero, A. L. Utz, *Phys. Rev. Lett.* **83**, 868 (1999).
25. P. Kratzer, B. Hammer, J. K. Nørskov, *J. Chem. Phys.* **105**, 5595 (1996).
26. G. Henkelman, A. Arnaldsson, H. Jonsson, *J. Chem. Phys.* **124**, 044706 (2006).
27. S. Nave, B. Jackson, *Phys. Rev. Lett.* **98**, 173003 (2007).
28. L. B. F. Juurlink, R. R. Smith, D. R. Killelea, A. L. Utz, *Phys. Rev. Lett.* **94**, 208303 (2005).
29. R. Milot, A. P. J. Jansen, *Phys. Rev. B* **61**, 15657 (2000).
30. R. D. Beck *et al.*, *Science* **302**, 98 (2003).
31. P. R. McCabe, L. B. F. Juurlink, A. L. Utz, *Rev. Sci. Instrum.* **71**, 42 (2000).
32. A. D. Johnson, S. P. Daley, A. L. Utz, S. T. Ceyer, *Science* **257**, 223 (1992).
33. Materials and methods are available as supporting material on *Science* Online.
34. M. Lewerenz, M. Quack, *J. Chem. Phys.* **88**, 5408 (1988).
35. I. Harrison, *Acc. Chem. Res.* **31**, 631 (1998).
36. H. S. Yoo, D. A. McWhorter, B. H. Pate, *J. Phys. Chem. A* **108**, 1380 (2004).
37. H. L. Abbott, A. Bukoski, I. Harrison, *J. Chem. Phys.* **121**, 3792 (2004).
38. J. C. Tully, *Annu. Rev. Phys. Chem.* **51**, 153 (2000).
39. A. M. Wodtke, J. C. Tully, D. J. Auerbach, *Int. Rev. Phys. Chem.* **23**, 513 (2004).
40. M.-N. Carré, B. Jackson, *J. Chem. Phys.* **108**, 3722 (1998).
41. A. C. Luntz, J. Harris, *Surf. Sci.* **258**, 397 (1991).

42. Supported by NSF (grant no. CHE-0415574).

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5864/790/DC1
Materials and Methods
References

9 November 2007; accepted 20 December 2007
10.1126/science.1152819

Colossal Positive and Negative Thermal Expansion in the Framework Material $Ag_3[Co(CN)_6]$

Andrew L. Goodwin,[1]* Mark Calleja,[2] Michael J. Conterio,[1] Martin T. Dove,[1] John S. O. Evans,[3] David A. Keen,[4,5] Lars Peters,[3] Matthew G. Tucker[4]

We show that silver(I) hexacyanocobaltate(III), $Ag_3[Co(CN)_6]$, exhibits positive and negative thermal expansion an order of magnitude greater than that seen in other crystalline materials. This framework material expands along one set of directions at a rate comparable to the most weakly bound solids known. By flexing like lattice fencing, the framework couples this to a contraction along a perpendicular direction. This gives negative thermal expansion that is 14 times larger than in ZrW_2O_8. Density functional theory calculations quantify both the low energy associated with this flexibility and the role of argentophilic ($Ag^+...Ag^+$) interactions. This study illustrates how the mechanical properties of a van der Waals solid might be engineered into a rigid, useable framework.

Thermal expansion in crystalline materials is a relatively well-understood physical process (*1*, *2*). By virtue of the inherent anharmonicity of bond vibrations, the average distance between bonded pairs of atoms increases with temperature, and, in general, this increase is reflected in expansion at the macroscopic scale. The relative rate, α, at which a material expands with increasing temperature usually falls within the range $0 \times 10^{-6}\ K^{-1} < \alpha < 20 \times 10^{-6}\ K^{-1}$ (*2*).

Examples of unconventional thermal expansion behavior are well known, and often these have highlighted some interesting and important physical processes in the respective materials. For example, the balance between lattice thermal expansion and magnetorestriction in invar FeNi alloys gives rise to their well-known and widely exploited near-zero thermal expansion around room temperature (*3*). Similar behavior in the intermetallic conductor YbGaGe has been explained in terms of a continuous electronic valence transition (*4*). A quite different type of atypical thermal expansion is the occurrence of strong negative thermal expansion (NTE) in framework structures such as ZrW_2O_8 (*5*) and $Cd(CN)_2$ (*6*), a consequence of the structural underconstraint of atomic bridging motifs, in these cases, the Zr-O-W and Cd-CN-Cd linkages. The ability to bend these linkages at the O or C and N atoms means that the dominant thermal response is associated with flexing bonds rather than stretching them, and it is this difference that is implicated in their unusual thermodynamic behavior. This flexibility can also lead to pressure-induced amorphization (*7*), elastic constant softening under pressure (*8*), and unusual shear and bulk moduli (*8*, *9*).

We studied the thermal expansion behavior of silver(I) hexacyanocobaltate(III), $Ag_3[Co(CN)_6]$, a framework material assembled from highly underconstrained Co-CN-Ag-NC-Co linkages. Its crystal structure consists of alternating layers of Ag^+ and $[Co(CN)_6]^{3-}$ ions, stacked parallel to the unique axis of its trigonal $P\bar{3}1m$ unit cell (Fig. 1A) (*10*, *11*). Within any given silver-containing layer, the Ag atoms are arranged at the vertices of a Kagome lattice, with the octahedral $[Co(CN)_6]^{3-}$ ions positioned above and below the hexagonal Kagome "holes." These anions are oriented such that each cyanide binds a single neighboring Ag^+ ion, which in turn is bonded to a second $[Co(CN)_6]^{3-}$ ion on the other side of the Kagome sheet. The almost-linear Co-CN-Ag-NC-Co linkages run parallel to the ⟨101⟩ lattice directions (Fig. 1B), forming a set of three identical interpenetrating cubic (α-Po) networks (*11*).

There are a number of points of interest concerning this structure. First, the separation between neighboring Ag atoms is circa (ca.) 3.5 Å, which is only marginally greater than the van der Waals limit of 3.4 Å (*12*), despite the Coulombic repulsion. Moreover, this close Ag...Ag approach is unsupported by the covalent lattice. Such a feature is strongly characteristic of so-called $d^{10}...d^{10}$ metallophilicity, whereby multipolar dispersive interactions produce a relatively weak "bond" between d^{10} centers of ca. 30 kJ mol^{-1} (*13*, *14*). Second, the cobalt atoms are present in the low-spin d^6 ($S = 0$) electronic configuration and so are both diamagnetic and coordinationally inert (*15*). Third, the Ag atoms are N-bound by the cyanide ions. This assignment has been questioned (*16*) precisely because it is unusual in cyanide-containing silver salts, where there is a strong chemical preference for C-bound Ag centers (*15*).

Having prepared a sample of $Ag_3[Co(CN)_6]$ as described in (*10*), we performed x-ray diffraction measurements over the temperature range from 16 to 500 K. The sample appeared to have decomposed fully by 500 K to give a product whose diffraction profile matched that of elemen-

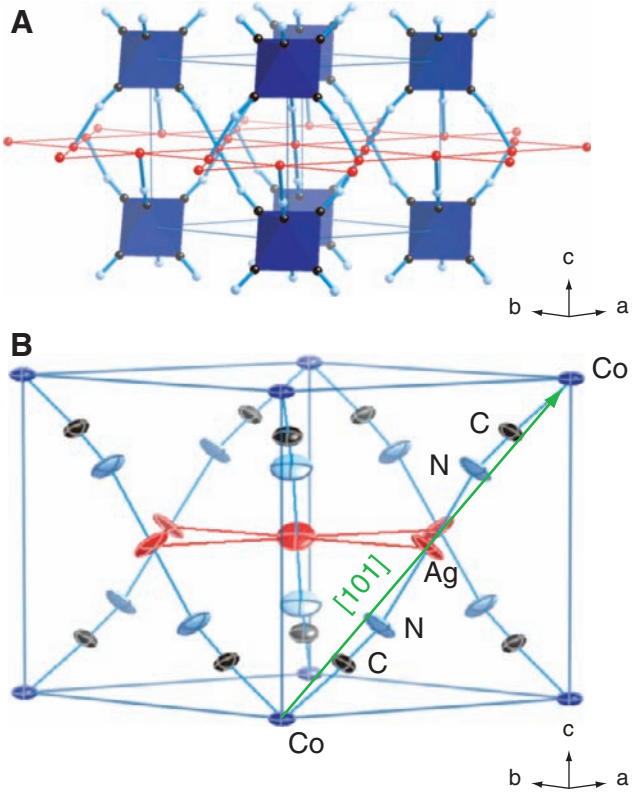


Fig. 1. Representations of the crystal structure of $Ag_3[Co(CN)_6]$. (**A**) The trigonal lattice consists of alternating layers of octahedral $[Co(CN)_6]^{3-}$ ions ($[CoC_6]$ octahedra shown in blue) and Ag^+ cations (red spheres arranged on the central Kagome-type lattice). (**B**) The crystallographic unit cell at 300 K as determined by structural refinement of our neutron diffraction data. The strongest bonding interactions, which occur within Co-CN-Ag-NC-Co linkages, all lie parallel to the crystallographic ⟨101⟩ directions. The refined anisotropic thermal ellipsoids (shown here at 50% probability) indicate that the dominant type of atomic displacement at 300 K involves translations of the C, N, and Ag atoms perpendicular to these linkages.

[1]Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK. [2]Cambridge eScience Centre, Wilberforce Road, Cambridge CB3 0WA, UK. [3]Department of Chemistry, University Science Laboratories, University of Durham, South Road, Durham DH1 3LE, UK. [4]ISIS Facility, Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0QX, UK. [5]Department of Physics, Oxford University, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK.

*To whom correspondence should be addressed. E-mail: alg44@cam.ac.uk

tal Ag. We also collected neutron time-of-flight total scattering patterns at temperatures of 10, 50, 150, and 300 K using the GEM diffractometer at ISIS (*17*, *18*). The room-temperature powder diffraction patterns were readily indexed according to the previously published crystallographic unit cell (*11*), and Rietveld refinement of the neutron data was used to produce a structural model (Fig. 1B) using the program GSAS (*19*). By virtue of the scattering contrast between C and N atoms attained in neutron experiments, we were able to confirm that the cyanide ions are entirely C-bound to each Co atom, thus addressing the concerns raised in (*16*). Crystallographic details and refined values of the anisotropic thermal displacement parameters are given in tables S1 and S2, respectively.

On cooling from room temperature, both the x-ray and the neutron diffraction patterns were affected in two particular ways (Fig. 2, A and B). First, the *d*-spacing values of most of the reflections changed substantially (with some values increasing and others decreasing). Second, we observed an increasingly severe anisotropic peak broadening effect; this effect, which disappeared on reheating, was strongest for those peaks with the greatest thermal shift in *d* spacing. We were able to model the unusual peak shape variation by using spherical harmonics to account for anisotropic strain broadening within the TOPAS structural refinement package (*20*) or, equivalently, by refining empirical lattice parameter distributions using the GSAS program (*19*, *18*) (Fig. 2, C to F). Both approaches gave entirely consistent values of the strain-free lattice parameters.

The quantitative thermal variations in these strain-free lattice parameters were determined from the x-ray diffraction data (*18*) (Fig. 3) (raw values are listed in tables S3 and S4). Our results show $Ag_3[Co(CN)_6]$ exhibits essentially linear thermal expansion behavior over its entire temperature stability range that is an order of magnitude stronger than that typically observed for framework materials: The uniaxial coefficients of thermal expansion were found to lie in the ranges $+130 \times 10^{-6}\ K^{-1} < \alpha_a < +150 \times 10^{-6}\ K^{-1}$ and $-120 \times 10^{-6}\ K^{-1} > \alpha_c > -130 \times 10^{-6}\ K^{-1}$ over much of the temperature range studied. Slightly more moderate effects appeared to occur at the very lowest temperatures, although we note that the uncertainty in the calculated values of α is substantially larger for these terminal data points. There was no appreciable hysteresis in cell parameters nor any evidence for the existence of structural phase transitions (the slight discontinuity at 300 K is due to the different experimental conditions used to measure data above and below room temperature). The absence of any sharp features in differential scanning calorimetric measurements of the specific heat supported this finding.

We see that the uniaxial NTE behavior observed in this study is larger than that reported for the isotropic materials ZrW_2O_8 ($\alpha_a = -9 \times 10^{-6}\ K^{-1}$) (*5*) and $Cd(CN)_2$ ($\alpha_a = -20 \times 10^{-6}\ K^{-1}$) (*6*) and greater also than the calculated behavior for various metal-organic frameworks (MOFs) ($-27 \times 10^{-6}\ K^{-1} \leq \alpha_a \leq -11 \times 10^{-6}\ K^{-1}$) (*21*), whereas the positive thermal expansion (PTE) effect along the *a* and *b* crystallographic axes is matched in magnitude only by the most weakly bound solids (*2*), for example, Xe at 50 K ($\alpha_a \cong +200 \times 10^{-6}\ K^{-1}$) (*22*). In order to highlight these fundamental differences from typical framework behavior, we are suggesting the use of the term "colossal" to signify $|\alpha| \geq 100 \times 10^{-6}\ K^{-1}$.

We now consider the relationship of the thermal expansion to the interatomic separations in the crystal structure. The magnitude of NTE and PTE effects means that many of these separations change substantially with temperature, but there also exists a set of directions along which the thermal expansion coefficient

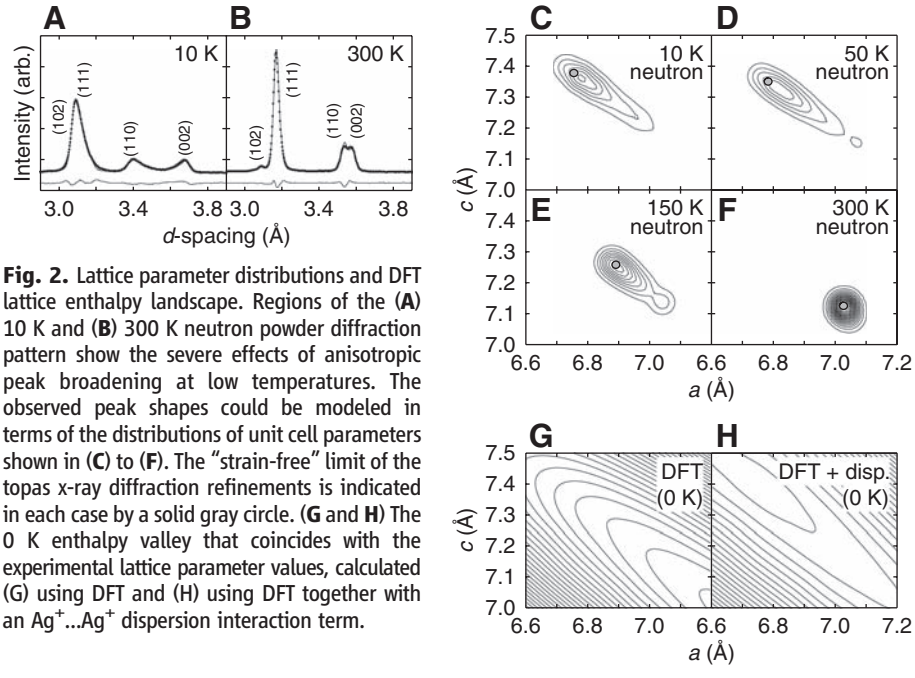


Fig. 2. Lattice parameter distributions and DFT lattice enthalpy landscape. Regions of the (**A**) 10 K and (**B**) 300 K neutron powder diffraction pattern show the severe effects of anisotropic peak broadening at low temperatures. The observed peak shapes could be modeled in terms of the distributions of unit cell parameters shown in (**C**) to (**F**). The "strain-free" limit of the topas x-ray diffraction refinements is indicated in each case by a solid gray circle. (**G** and **H**) The 0 K enthalpy valley that coincides with the experimental lattice parameter values, calculated (G) using DFT and (H) using DFT together with an $Ag^+ \cdots Ag^+$ dispersion interaction term.

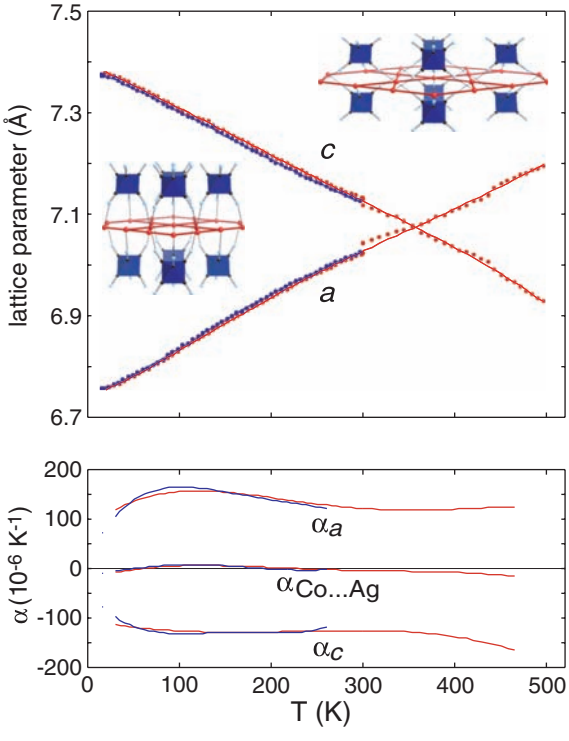


Fig. 3. Thermal expansion behavior of $Ag_3[Co(CN)_6]$ as determined from x-ray powder diffraction. (**Top**) Thermal variation of the lattice parameters *a* and *c* measured on cooling (300 to 16 K; blue) and heating (20 to 300 K, 300 to 500 K; red). The large changes in these parameters correspond to a substantial expansion of the Ag...Ag and $[Co(CN)_6] \cdots [Co(CN)_6]$ contacts but a similarly strong collapse in the separation between successive $(Ag^+)_n$ and $[Co(CN)_6]$ layers (accentuated in the schematics). The slight discontinuity at 300 K is due to the different experimental conditions used to measure data above and below room temperature. (**Bottom**) The coefficients of thermal expansion determined by a smooth polynomial fit ($n = 5$) to the raw lattice parameter data, together with $\alpha_{Co...Ag}$, the coefficient of thermal expansion along the ⟨101⟩ crystal axes. Values of α at the highest and lowest temperature points in each data set (where the first derivative of the polynomial fit is poorly constrained) have been omitted for clarity.

vanishes. This set includes the ⟨101⟩ axes, the directions along which the Co-CN-Ag-NC-Co linkages are oriented. As such, despite the large change in lattice dimensions, the data show that the separation between connected Co...Ag...Co atoms remains essentially constant across the temperature range studied (see the $\alpha_{Co...Ag}$ curve in the bottom graph of Fig. 3). This is, of course, entirely consistent with intuition for a strongly bound chain of atoms. There is also a small concomitant reorientation of the CN ions that preserves the average C-Co-C and N-Ag-N angles such that the geometries of the transition metal coordination polyhedra are largely unaffected by the thermal expansion. This reorientation means that the thermal expansion behavior of the metal-cyanide linkages in $Ag_3[Co(CN)_6]$ is more complex than that in $Cd(CN)_2$, for example, where NTE is caused by vibrational motion of the linkages alone. However, the transverse CN and Ag displacements evident in the atomic displacement parameters of Fig. 1B do resemble the typical behavior observed in other NTE frameworks, and it is likely that in $Ag_3[Co(CN)_6]$ local transverse vibrations at least help reduce thermal expansion along the ⟨101⟩ directions.

On the other hand, interatomic separations parallel and perpendicular to the trigonal axis are strongly affected by temperature. The separation between silver- and hexacyanocobaltate-containing layers, which corresponds to $c/2$, decreases quickly as the material is heated, whereas the average Ag...Ag and $[Co(CN)_6]$...$[Co(CN)_6]$ distances, which correspond to $a/2$, increase just as rapidly. The strong coupling between these two lattice parameters is caused by the Co-CN-Ag-NC-Co linkages: If the structural integrity of these linkages is to be preserved, then an increase in a must be accompanied by a decrease in c of comparable magnitude (*23*). As such, the crystal structure behaves like a sheet of garden lattice fencing, whereby an expansion of the lattice along one axis forces a contraction in the perpendicular direction.

This picture of geometric flexibility explains the coupling between a and c, but the question remains as to what in particular is responsible for the large magnitude of the changes in these lattice parameters. An interesting comparison can be made with the related compound $H_3[Co(CN)_6]$, which shares the same structure as $Ag_3[Co(CN)_6]$ but with H atoms replacing Ag atoms (*11*, *24*). At room temperature, the deuterated analog $D_3[Co(CN)_6]$ crystallizes with lattice dimensions $a = 6.431 \pm 0.004$ [or 6.431(4)] Å and $c = 5.710(4)$ Å (*25*) so that the $[Co(CN)_6]^{3-}$ ions are closer together than they are in $Ag_3[Co(CN)_6]$. At 77 K, the lattice parameters are $a = 6.411(4)$ Å and $c = 5.715(4)$ Å (*25*), reflecting a more moderate thermal expansion behavior with $\alpha_a = +14(6) \times 10^{-6}$ K^{-1} and $\alpha_c = -4(6) \times 10^{-6}$ K^{-1}. Colossal thermal expansion is thus not an inherent property of the framework topology.

We obtained a more quantitative insight into the thermomechanical properties of $Ag_3[Co(CN)_6]$ by using the density functional theory (DFT) code CASTEP (*26*) to calculate the variation in lattice enthalpy for different values of the unit cell parameters (*18*). What emerges from these calculations is that there exists a pronounced low-enthalpy valley that connects the various lattice parameter values measured crystallographically (Fig. 2G). The cell dimensions can vary along the "floor" of this valley with minimal cost in lattice enthalpy. Because excited electronic states are not taken into account at the DFT level, CASTEP omits any contribution from argentophilic dispersion interactions. However, the inclusion of a modest additional dispersive (r^{-6}) term to the DFT enthalpy values is sufficient to shift the overall minimum from its original DFT-only position (a = 7.65 Å) to one closer to the expected 0 K value (≅6.74 Å), producing an even flatter minimum along the valley floor in the process (Figs. 2H and 4A). So, to a very large extent, the cell parameter a appears to be determined by these dispersion forces. This is highly unusual for a framework material and hints at why the observed PTE effect might share more in common with the sort of values observed for van der Waals solids such as Xe.

Exploring this enthalpy valley more closely, we note that reasonably large changes in a and c can carry enthalpy penalties so low that they are comparable to the subtle thermal changes in phonon frequencies often found in framework materials. This means that phonon contributions to the free energy may be responsible for driving the large changes in lattice parameters. A reciprocal-space analysis of our neutron scattering data (*27*), allowed us to calculate the partial phonon densities of states for Ag, Co, and CN species as a function of temperature (*18*). What we found was that the average phonon energy associated with Ag vibrations decreased noticeably with increasing temperature (and hence with increasing a), whereas there was very little change in the relative energies of Co and CN vibrations (Fig. 4B). In terms of the energy profile illustrated in Fig. 4A, this means that, at higher temperatures (for which the phonon contribution to the overall energy becomes more important), the change in phonon frequencies associated with increasing a is commensurate with the small lattice enthalpy penalty involved. The decrease in Ag vibrational energies of ca. 5 meV between 10 and 300 K corresponds to an additional contribution of about 0.04 eV per unit cell to the lattice free energy at 300 K, comparable to the difference in lattice enthalpies at the corresponding values of a. Consequently, those phonon modes involving Ag displacements contribute most strongly to the overall (positive) Grüneisen parameter, and hence it is the anharmonicity of Ag vibrations that appears to drive the thermal expansion of the flexible lattice. We note that this anharmonicity is by no means colossal in itself, but because α is inversely proportional to material stiffness a modest value of the Grüneisen parameter can still produce a large value of α.

Consequently, it is the geometric flexibility of the $Ag_3[Co(CN)_6]$ lattice, that is, the shallowness of the enthalpy valley shown in Fig. 2H, that allows very weak Ag^+...Ag^+ interactions to produce such an unusually large PTE effect, translated via flexing of Co-CN-Ag-NC-Co linkages into an equally strong NTE effect along the trigonal crystal axis. This unusual behavior serves to illustrate how very weak bonding interactions can actually have a profound effect on crystal thermodynamics in sufficiently flexible systems. In this case, not only do we observe unusual thermal expansion, but the lattice parameter distributions shown in Fig. 2 are a direct reflection of an atypically soft lattice (more specifically, one with a large and negative elastic compliance S_{13}); the crystallites appear to

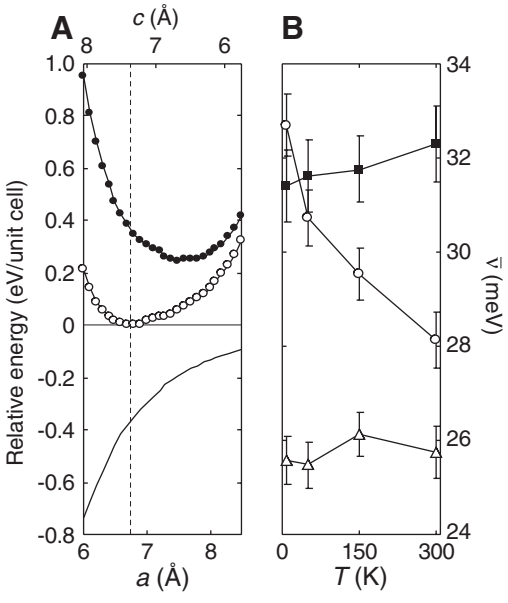


Fig. 4. Lattice enthalpy and phonon energies in $Ag_3[Co(CN)_6]$. (**A**) DFT lattice enthalpies (solid circles) along the floor of the enthalpy valley found in (a, c) space. Addition of a dispersive r^{-6} term (solid line) produces a modified enthalpy curve (open circles) with a particularly shallow minimum positioned closer to the experimental 0 K value (dashed vertical line). (**B**) Thermal variation in mean partial phonon frequencies for Ag (open circles), Co (open triangles), and CN (solid squares) species as calculated from our neutron scattering data. Values were calculated from all 48 phonon modes across a grid of 245 points in reciprocal space (giving 11,760 data points in total); error bars correspond to the standard error in the calculated mean frequencies. The energies of phonon modes involving Ag displacements decrease with an increase in temperature (and hence a), providing a mechanism to overcome the very small lattice enthalpy differences near the minimum in (A).

deform readily in response to thermal and/or mechanical stress. There is a strong interest in uniaxial NTE materials of such extraordinary properties. For example, precision optical devices used on satellites are often highly sensitive to even very minor changes in dimension, and this sensitivity is exacerbated by the large temperature gradients over which they are forced to operate. Very thin aligned coatings of a material such as $Ag_3[Co(CN)_6]$ could provide an intrinsic protection against this thermal variation, avoiding the present reliance on mechanical adjustment. More generally, the concept of exploiting geometric flexibility to amplify the role of low-energy interactions in thermodynamic behavior offers a method of preparing materials with a range of hitherto unanticipated physical properties.

References and Notes

1. T. H. K. Barron, G. K. White, *Heat Capacity and Thermal Expansion at Low Temperatures* (Kluwer Academic, New York, 1999).
2. R. S. Krishnan, R. Srinivasan, S. Devanarayanan, *Thermal Expansion of Crystals* (Pergamon, Oxford, 1979).
3. E. F. Wasserman, in *Ferromagnetic Materials*, K. H. J. Buschow, E. P. Wohlfarth, Eds. (North-Holland, Amsterdam, 1990), vol. V.
4. J. R. Salvador, F. Guo, T. Hogan, M. G. Kanatzidis, *Nature* **425**, 702 (2003).
5. T. A. Mary, J. S. O. Evans, T. Vogt, A. W. Sleight, *Science* **272**, 90 (1996).
6. A. L. Goodwin, C. J. Kepert, *Phys. Rev. B* **71**, 140301 (2005).
7. D. A. Keen *et al.*, *Phys. Rev. Lett.* **98**, 225501 (2007).
8. C. A. Pantea *et al.*, *Phys. Rev. B* **73**, 214118 (2006).
9. T. R. Ravindran, A. K. Arora, S. Chandra, M. C. Valsakumar, N. V. Chandra Shekar, *Phys. Rev. B* **76**, 054302 (2007).
10. A. Ludi, H. U. Güdel, V. Dvorák, *Helv. Chim. Acta* **50**, 2035 (1967).
11. L. Pauling, P. Pauling, *Proc. Natl. Acad. Sci. U.S.A.* **60**, 362 (1968).
12. A. Bondi, *J. Phys. Chem.* **68**, 441 (1964).
13. H. Schmidbaur, *Nature* **413**, 31 (2001).
14. P. Pyykkö, *Angew. Chem. Int. Ed.* **43**, 4412 (2004).
15. A. G. Sharpe, *The Chemistry of Cyano Complexes of the Transition Metals* (Academic Press, London, 1976).
16. U. Geiser, J. A. Schlueter, *Acta Crystallogr.* **C59**, i21 (2003).
17. A. C. Hannon, *Nucl. Instrum. Methods Phys. Res. A* **551**, 88 (2005).
18. Materials and methods are available as supporting material on *Science* Online.
19. A. C. Larson, R. B. von Dreele, "General structure analysis system (GSAS)," Los Alamos National Laboratory Report No. LAUR 86-748 (2000).
20. A. A. Coelho, TOPAS v2.0: General profile and structure analysis software for powder diffraction data, Karlsruhe (2000).
21. D. Dubbeldam, K. S. Walton, D. E. Ellis, R. Q. Snurr, *Angew. Chem. Int. Ed.* **46**, 4496 (2007).
22. D. R. Sears, H. P. Klug, *J. Chem. Phys.* **37**, 3002 (1962).
23. It is straightforward to show that the precise condition is given by $\alpha_c = -a^2\alpha_a/c^2$. Because in this case $a \cong c$, the condition reduces to $\alpha_c \cong -\alpha_a$.
24. H. U. Güdel, A. Ludi, P. Fischer, W. Hälg, *J. Chem. Phys.* **53**, 1917 (1970).
25. H. U. Güdel, A. Ludi, P. Fischer, *J. Chem. Phys.* **56**, 674 (1972).
26. S. J. Clark, *et al.*, *Z. Kristallogr.* **220**, 567 (2005).
27. A. L. Goodwin, M. G. Tucker, M. T. Dove, D. A. Keen, *Phys. Rev. Lett.* **93**, 075502 (2004).

28. We gratefully acknowledge the assistance of M. Giot (Rutherford Appleton Laboratory) in performing heat capacity measurements, D. Wilson (J. W. Goethe University, Frankfurt am Main) for valuable discussions and for providing the pseudopotentials used in our DFT analysis, the University of Cambridge's CamGrid infrastructure for computational resources, Engineering and Physical Science Research Council (UK) for funding to J.S.O.E. and L.P. under ep/c538927/1, and Trinity College, Cambridge for the provision of financial support to M.J.C. and A.L.G.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5864/794/DC1
Materials and Methods
Tables S1 to S4
References

9 October 2007; accepted 12 December 2007
10.1126/science.1151442

Elastic Anisotropy of Earth's Inner Core

Anatoly B. Belonoshko,[1,2]* Natalia V. Skorodumova,[3] Anders Rosengren,[2,4] Börje Johansson[1,3,5]

Earth's solid-iron inner core is elastically anisotropic. Sound waves propagate faster along Earth's spin axis than in the equatorial plane. This anisotropy has previously been explained by a preferred orientation of the iron alloy hexagonal crystals. However, hexagonal iron becomes increasingly isotropic on increasing temperature at pressures of the inner core and is therefore unlikely to cause the anisotropy. An alternative explanation, supported by diamond anvil cell experiments, is that iron adopts a body-centered cubic form in the inner core. We show, by molecular dynamics simulations, that the body-centered cubic iron phase is extremely anisotropic to sound waves despite its high symmetry. Direct simulations of seismic wave propagation reveal an anisotropy of 12%, a value adequate to explain the anisotropy of the inner core.

It has been proposed (*1*) and lately confirmed (*2*) that Earth's inner core (*3*) (IC), a spherical body in the center of Earth with a radius of about 1200 km, is elastically anisotropic. Comparing the time residuals for the different paths of seismic waves, Creager (*2*) found that seismic waves propagate by 3 to 4% faster in the direction of Earth's spin axis than in the direction aligned with the equatorial plane. Furthermore, the anisotropy has a rather complicated structure (*4*–*9*). The uppermost layer of the core, a few dozen kilometers thick but varying between hemispheres, is nearly elastically isotropic (*4*, *6*). In addition, the innermost part of the inner core (IMIC) (*7*), with a radius of 300 km (*7*) to 500 km (*5*), seems to possess a different form of anisotropy (*8*).

A number of mechanisms explaining the anisotropy have been proposed (*10*–*15*). Although the mechanisms vary in detail, the major consensus is that the anisotropy is due to a lattice preferred orientation of the IC material. This conclusion is also supported by the correlation between the attenuation and velocities of seismic waves (*5*). Therefore, it is important to understand the elastic anisotropy of the IC material under representative conditions. It has long been established, on the basis of the equation of state as compared to seismic data and the abundance of iron, that Earth's IC mainly consists of iron (*16*–*19*). Because iron at high pressure is stable in the hexagonal close-packed (hcp) structure, it has been suggested that the anisotropy is due to the preferred orientation of the hcp lattice (*20*). As it was difficult at that time to compute the elastic properties of iron simultaneously at high pressure and temperature, the elastic properties were calculated at high pressure and zero temperature. These calculations (*20*) showed that the anisotropy of the IC would be approximately reproduced if a single hcp iron crystal, oriented with its "fast" *c* axis along the spin axis of Earth, was assumed to exist in the center of the planet.

This, of course, would be the case if the high temperature of Earth's IC does not change the nonideal *c/a* ratio (*21*) existing at low temperature. However, it has lately been shown (*22*, *23*) that the *c/a* ratio in hcp iron becomes almost ideal at the high temperature of the IC. Therefore, the elastic anisotropy of the hcp phase practically vanishes at high temperature. This means, in turn, that even the assumption that the IC is a single hcp Fe crystal does not allow us to reproduce the measured time residuals (*2*).

Therefore, one must search for another, elastically anisotropic phase of iron (or iron alloy) to explain the IC anisotropy. The body-centered cubic (bcc) Fe phase has been suggested (*22*, *24*–*27*) and has recently received experimental confirmation (*28*). Therefore, it is of interest to determine the speed of sound wave propagation in oriented monocrystalline as well as polycrystalline bcc Fe samples.

Because the difficulty of reproducing the conditions of the IC prevents a suitable experiment at present, we have simulated sound wave propagation in the material by means of the molecular dynam-

[1]Applied Materials Physics, Department of Materials Science and Engineering, Royal Institute of Technology, SE-100 44 Stockholm, Sweden. [2]Condensed Matter Theory, Department of Theoretical Physics, AlbaNova University Center, Royal Institute of Technology, SE-106 91 Stockholm, Sweden. [3]Condensed Matter Theory Group, Department of Physics, Uppsala University, Uppsala Box 530, Sweden. [4]NORDITA, AlbaNova University Center, SE-106 91 Stockholm, Sweden. [5]School of Physics and Optoelectronic Technology and College of Advanced Science and Technology, Dalian University of Technology, Dalian 116024, China.

*To whom correspondence should be addressed. E-mail: anatoly.belonoshko@fysik.uu.se

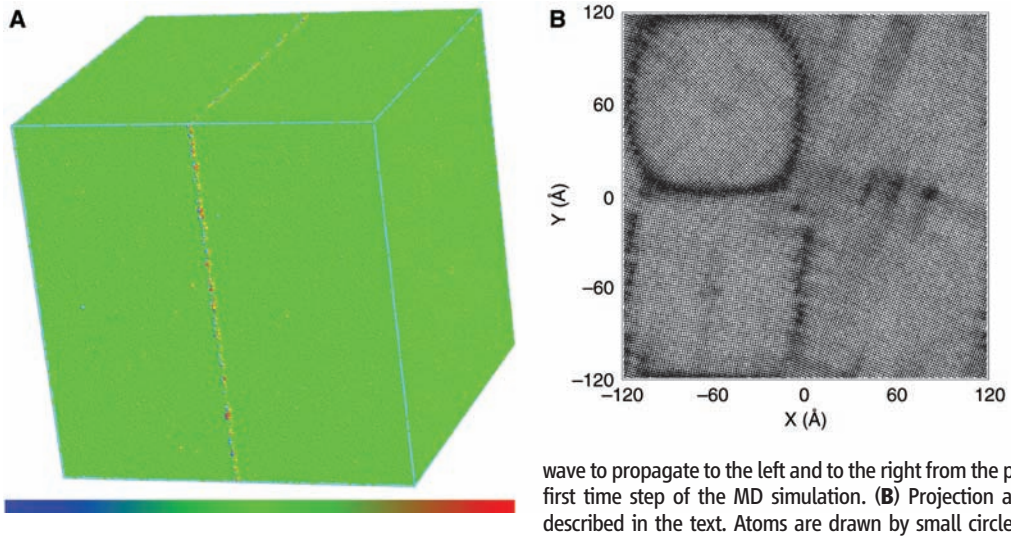


Fig. 1. Some of the simulated samples. (**A**) Body-centered cubic sample, equilibrated at $P = 360.0$ GPa and $T = 6600$ K. The sound wave is initiated as described in the text. The length of the cube edge is 238.03 Å. The atoms are shown by spheres of radius 1.1 Å, which are colored according to the X component of the force vector acting on the atoms. Most of the atoms are green, which corresponds to a comparably small force acting on them (see the color bar in the bottom part of the figure). The plane of dark-blue (large negative X component of the force vector) and red (large positive X component of the force vector) atoms in the middle section represents the initial energy pulse, which causes a sound wave to propagate to the left and to the right from the plane (see Fig. 2A). The sample is shown at the first time step of the MD simulation. (**B**) Projection along the Z axis, showing the frozen sample described in the text. Atoms are drawn by small circles. Defects and grain boundaries are seen as dark areas. The sample is referred to as a polycrystal in the text and Table 1.

Table 1. Calculated velocities of sound waves under *PT* conditions of inner core.

Sample*	*P* (GPa)	*T* (K)	Direction†	Velocity (km/s)
bcc, single	361.0	6589	100	12.05
bcc, single	360.3	6587	010	12.02
bcc, single	360.2	6595	001	12.16
bcc, poly	366.7	6918	*X*	12.86
bcc, poly	364.8	6929	*Y*	12.60
bcc, poly	364.3	6850	*Z*	12.02
bcc, single	364.8	6644	110	13.06
bcc, single	362.7	6640	112	12.90
bcc, single	364.6	6924	111	13.60
hcp, single	368.5	6860	001	13.89
hcp, single	369.1	7002	100	13.89
hcp, single	364.5	6726	110	13.57

*The sample description provides crystal structure (hcp or bcc) and whether the crystal is single or polycrystal. †The direction of the initiated sound wave is given by providing crystallographic direction *hkl*, where *hkl* are the integer vector components. For example, 111 in the bcc crystal means direction of the big diagonal in the cubic unit cell, and 001 means direction along the edge of the cell.

ics (MD) method. The iron was modeled with the embedded-atom method (EAM) (*22*, *29*–*32*), which describes the interaction between atoms as a combination of short-range repulsion and many-body attraction terms. The latter represents the energy of embedding an atom into the electron gas. The model was parameterized to reproduce the results of first-principles calculations (*29*). This EAM model has been exceptionally well tested (*22*, *29*–*32*), including a calculation of the elastic constants of the bcc iron crystal and comparing them to the results from one of the most precise ab initio methods (the projector-augmented wave method) (*32*). The successful application of our EAM model (*22*, *29*–*32*) allowed us to rely on the EAM model for calculating the speed of sound in iron.

Overall, we studied three samples of bcc iron: two single crystals and one polycrystal. One of the single crystals was chosen in standard crystallographic orientation where the unit cell contains two atoms. Another crystal was chosen in such a way that the big diagonal ([111] direction) was oriented along the Z axis. The unit cell for this sample contained 12 atoms. Axes X and Y coincided with the directions [110] and [112], respectively. The preparation of the third, polycrystalline sample included the following steps. The bcc sample containing 2,000,000 atoms was heated to 10,000 K to ensure melting. Then, four bcc crystals were embedded into the four quadrants of the box containing the melted iron. The [001] axes of these crystals were aligned with the Z axis. The embedded crystals were randomly rotated around the [001] axis passing through the centers of quadrants in the XY plane. This sample was then crystallized at $T = 6800$ K and $P = 3.6$ Mbar for 300,000 time steps (Fig. 1B). The samples containing single crystals were equilibrated at about the same values of P and T. The typical length of the side of the sample was roughly 240 Å (Fig. 1A). Such a large sample was necessary to simulate the sound wave propagation without the interference of sound waves attributable to the periodic boundary conditions.

We have studied the propagation of sound waves in single-crystal bcc iron along the [100], [010], [001], [110], [112], and [111] crystallographic directions and along the X, Y, and Z directions in the case of the polycrystalline bcc Fe samples (Figs. 1 and 2). We also studied sound waves in single-crystal hcp iron along the [100], [001], and [110] directions. The sound waves were initiated (Fig. 1A) by shifting the two halves of the prepared samples toward each other by 0.3 Å along the X axis. The shifting of the two halves toward each other created an energy pulse concentrated in the very thin layer of about 15% of the interatomic distance. Starting from such a configuration, we performed simulations in the NVE ensemble (where N is the number of particles, E is the total energy, and V is the volume). The propagation of the initial energy pulse was followed by calculating the profile of the characteristic

$$A_i^x = \frac{1}{N_i^{\text{atoms}}} \sum_{\text{atom} \in i\text{th layer}} f_{\text{atom}}^x \qquad (1)$$

where N_i^{atoms} is the number of atoms within the ith layer, f_{atom}^x is the X component of the force acting on an atom, and X is the direction of wave propagation.

This method was chosen because at high pressure, force is highly sensitive to the distance change. The average force acting on an atom in equilibrium is zero. When a sound wave passes through the sample, it creates a wave of densification immediately followed by the rarefaction wave. Both rarefaction and densification are subtle. However, because the force in the EAM model is roughly an 8th power of the inverse distance, force changes markedly within these waves. This makes it comparably easy to detect. One can see how the sound wave propagates with time along different directions (Fig. 2). Having the positions of the sound wave at different moments in time, one can precisely determine the speed of the sound wave.

Our results are summarized in Table 1. The temperatures listed are close to the most reliable

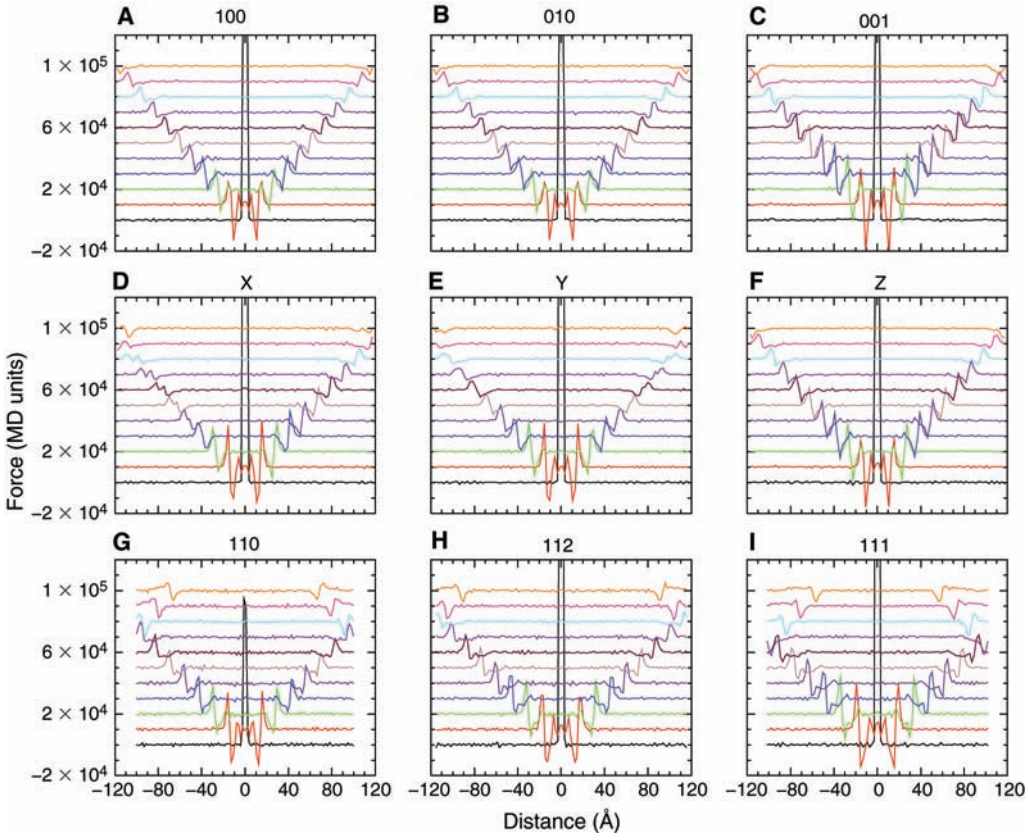


Fig. 2. Positions of sound waves for a number of samples and directions. For each sample and direction we calculated at least 11 profiles of the quantity in Eq. 1 (which is an average of the forces on atoms in the given layer). The profiles have been calculated for configurations separated in time by 0.1 ps. For convenience, each calculated profile is plotted by a different color and shifted upward (starting from profile shown in red) by 10,000 MD units of force (in MD units, the unit of time is the picosecond, the unit of distance is the angstrom, and the unit of mass is the atomic mass unit). (**A**) bcc sample, [100] direction; (**B**) bcc sample, [010] direction; (**C**) bcc sample, [001] direction; (**D**) poly-"XY" crystal, *X* direction; (**E**) poly-"XY" crystal, *Y* direction; (**F**) poly-"XY" crystal, *Z* direction; (**G**) bcc sample, [110] direction; (**H**) bcc sample, [112] direction; (**I**) bcc sample, [111] direction. Some of the profiles show waves that have traveled through the boundary of the sample (all of the samples are modeled with periodic boundary conditions) and entered the sample from the opposite side. The speed of the sound wave is calculated by dividing the distance between two waves by 2 and the time interval the waves traveled. Because the procedure is identical for all samples and directions, the relative differences between velocities are highly precise.

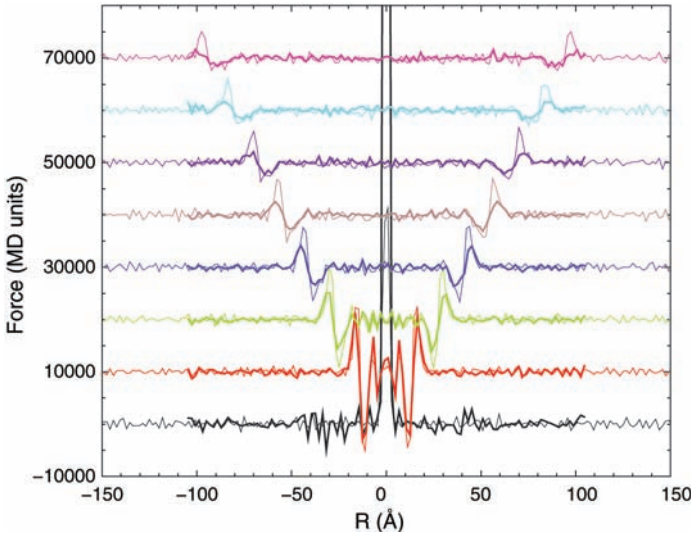


Fig. 3. Positions of sound waves propagating in the hcp single crystal. The positions of waves propagating along the *c* axis (thin line) and *a* axis (thick line) coincide. The profiles for each subsequent time are shifted upward by 10,000 MD units, similar to Fig. 2.

present estimates (*22*). From the computed velocities we learn the following. First, the precision of our procedure for computing velocity is on the order of 0.1 km/s. This follows from the comparison of the first three lines (Table 1) that contain velocities for the single bcc crystal and equivalent directions. Second, the ⟨001⟩ directions are the slowest ones, whereas the ⟨111⟩ directions are the fastest ones. Any deviation from ⟨001⟩ directions leads to increase of the sound wave velocity (compare velocities in lines 4 and 5 with velocity in line 6 for polycrystal; direction *Z* is equivalent to [001], whereas directions *X* and *Y* are approximately equivalent to a random polycrystal). Third, defects have practically no impact on the sound wave velocity (compare the first three lines and line 6). Fourth, the difference between the slowest (along ⟨001⟩) and fastest (along ⟨111⟩) velocities of sound waves in bcc iron is 1.53 ± 0.1 km/s. Fifth, the hcp phase of iron at the IC conditions becomes essentially isotropic (compare the last three lines; see also Fig. 3). The anisotropy of hcp, largest between the [001] and [100] directions at low temperature, vanishes at high temperature. The nonaxial direction [110] differs from the axial directions [001] and [100] by less than 2.5%, clearly insufficient to match the experimental observations (*2*).

What is the reason for such a different elastic behavior of the bcc and hcp iron phases? At high temperature, an iron atom in the hcp structure is surrounded by 12 identical neighbors. An iron atom in the bcc structure is surrounded by eight atoms at a distance slightly shorter than that in the hcp structure and by six atoms located at a slightly larger distance compared to the nearest-neighbor separation in the hcp structure. This is exactly what makes bcc so anisotropic: two different kinds of nearest neighbors, whereas in hcp there is just one.

Such a large difference in the velocities explains the anisotropy of the core. One of the proposed mechanisms for the formation of a lattice preferred orientation (*10*–*15*) (or likely some combination of them) is responsible for the formation of the bcc crystals with [111] aligned with the spin axis. Given the large difference of the velocities in different crystallographic directions along with the subtle impact of defects on the velocity, about 30% of crystals being preferably

oriented would produce the observed seismic signals. This is a much weaker requirement than the one suggested in (*20*), where even at zero temperature the whole single crystal (100%) has to be oriented with one of its axes along the spin axis of the planet. In passing, we note that bcc velocities (Table 1) are a better match to the seismic velocities in Earth's center [11.26 km/s (*18*)] than are hcp velocities.

Recently, it was suggested that the IC consists of two parts; the central part seems to possess an anisotropy different from the rest of the core (*7*). Later, it was proposed that the innermost IC (IMIC) is elastically less anisotropic (*8*). Such a different elastic behavior can be explained in terms of two iron phases, one of which (bcc) is anisotropic and the other (hcp) isotropic. If the *PT* slope of the hcp-bcc boundary and the geotherm in the core cross each other, then the different elastic behavior might be explained by the bcc-hcp phase transition, where the outer anisotropic part of the IC consists of the anisotropic bcc phase and the IMIC consists of a mixture of bcc and hcp iron alloys. The exact position of the hcp-bcc alloy *PT* boundary is unknown. However, if the field of the bcc stability is narrow (and it most likely is), then such a transition is probable. An alternative explanation could be that during the formation of the IMIC, the processes responsible for the formation of a lattice preferred orientation were weaker, and therefore in this part of the core a random orientation of bcc crystals is dominant. In either case, the bcc phase is an indisputable favorite over the hcp phase to be responsible for the anisotropy of the IC. This is strong evidence for the presence of a bcc iron/iron alloy phase in Earth's inner core.

References and Notes

1. A. Morelli, A. M. Dziewonski, J. H. Woodhouse, *Geophys. Res. Lett.* **13**, 1545 (1986).
2. K. C. Creager, *Nature* **356**, 309 (1992).
3. I. Lehmann, *Bur. Cent. Seismol. Int.* **14**, 3 (1936).
4. A. Cao, B. Romanowicz, *Geophys. Res. Lett.* **34**, L08303 (2007).
5. A. Souriau, B. Romanowicz, *Phys. Earth Planet. Inter.* **101**, 33 (1997).
6. X. Song, D. V. Helmberger, *J. Geophys. Res.* **100**, 9805 (1995).
7. M. Ishii, A. Dziewonski, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14026 (2002).
8. C. Beghein, J. Trampert, *Science* **299**, 552 (2003).
9. B. Romanowicz, X.-D. Li, J. Durek, *Science* **274**, 963 (1996).
10. S. Karato, *Science* **262**, 1708 (1993).
11. M. I. Bergman, *Nature* **389**, 60 (1997).
12. S. Yoshida, I. Sumita, M. Kumazawa, *J. Phys. Condens. Matter* **10**, 11215 (1998).
13. B. A. Buffett, *Science* **288**, 2007 (2000).
14. S. C. Singh, M. A. J. Taylor, J. P. Montagner, *Science* **287**, 2471 (2000).
15. B. A. Buffett, H.-R. Wenk, *Nature* **413**, 60 (2001).
16. E. Birch, *J. Geophys. Res.* **57**, 227 (1952).
17. L. V. Altshuler, K. K. Krupnikov, B. N. Ledenev, V. I. Zhuchikhin, M. I. Brazhnik, *Zhur. Eksp. Teor. Fiz.* **34**, 874 (1958).
18. A. M. Dziewonski, D. L. Anderson, *Phys. Earth Planet. Inter.* **25**, 297 (1981).
19. R. J. Hemley, H.-K. Mao, *Int. Geol. Rev.* **43**, 1 (2001).
20. L. Stixrude, R. E. Cohen, *Science* **267**, 1972 (1995).
21. We are aware of a paper (*33*) that claims to solve the problem we deal with in our paper. The authors (*33*) reported a computationally obtained anomalous change of *c/a* ratio (*c* is the size of the hexagonal unit cell along the [001] direction, and *a* is the corresponding size in the [100] and [010] directions on the basal plane; all atoms occupy identical positions when *c/a* = 1.6299, the so-called ideal ratio) in hexagonal iron on increasing temperature. That could save the hexagonal phase as the IC material. However, one of the authors of that paper (*33*) has recently published another paper (*34*) on the same subject, where it is demonstrated that the result in the previous publication (*33*) is an artifact of erroneous calculations. Therefore, we do not discuss those results in our paper.
22. A. B. Belonoshko, R. Ahuja, B. Johansson, *Nature* **424**, 1032 (2003).
23. C. Gannarelli, D. Alfé, M. J. Gillan, *Phys. Earth Planet. Inter.* **152**, 67 (2005).
24. J. M. Brown, R. G. McQueen, *J. Geophys. Res.* **91**, 7485 (1986).
25. M. Ross, D. A. Young, R. J. Grover, *J. Geophys. Res.* **95**, 21713 (1990).
26. M. Matsui, O. L. Anderson, *Phys. Earth Planet. Inter.* **103**, 55 (1997).
27. L. Vočadlo *et al.*, *Nature* **424**, 536 (2003).
28. L. Dubrovinsky *et al.*, *Science* **316**, 1880 (2007).
29. A. B. Belonoshko, R. Ahuja, B. Johansson, *Phys. Rev. Lett.* **84**, 3638 (2000).
30. L. Koci, A. B. Belonoshko, R. Ahuja, *Phys. Rev. B* **73**, 224113 (2006).
31. A. B. Belonoshko, E. I. Isaev, N. V. Skorodumova, B. Johansson, *Phys. Rev. B* **74**, 214102 (2006).
32. A. B. Belonoshko *et al.*, *Science* **316**, 1603 (2007).
33. G. Steinle-Neumann, R. E. Cohen, L. Stixrude, O. Gulseren, *Nature* **413**, 57 (2001).
34. X. W. Sha, R. E. Cohen, *Phys. Rev. B* **74**, 064103 (2006).

35. We thank B. Smith and I. Todorov for the DL_POLY package and K. Kadau for the package MD_render. Computations were performed at the Parallel Computer Center (PDC) in Stockholm, the National Supercomputing Center in Linköping, and Los Alamos National Laboratory (New Mexico, USA). Supported by the Swedish Research Council and the Swedish Foundation of Strategic Research.

10 September 2007; accepted 20 December 2007
10.1126/science.1150302

The Spatial Pattern and Mechanisms of Heat-Content Change in the North Atlantic

M. Susan Lozier,[1]* Susan Leadbetter,[2] Richard G. Williams,[2]* Vassil Roussenov,[2] Mark S. C. Reed,[1] Nathan J. Moore[1]†

The total heat gained by the North Atlantic Ocean over the past 50 years is equivalent to a basinwide increase in the flux of heat across the ocean surface of 0.4 ± 0.05 watts per square meter. We show, however, that this basin has not warmed uniformly: Although the tropics and subtropics have warmed, the subpolar ocean has cooled. These regional differences require local surface heat flux changes (±4 watts per square meter) much larger than the basinwide average. Model investigations show that these regional differences can be explained by large-scale, decadal variability in wind and buoyancy forcing as measured by the North Atlantic Oscillation index. Whether the overall heat gain is due to anthropogenic warming is difficult to confirm because strong natural variability in this ocean basin is potentially masking such input at the present time.

Recent evidence that the world's oceans have warmed over the past 50 years (*1*, *2*), with the attendant increase in the ocean's heat content an order of magnitude larger than the increase in the atmospheric and cryospheric heat content (*2*, *3*), has underscored the importance of the ocean as a heat reservoir for Earth's climate system. To facilitate predictions of future oceanic heat uptake, however, an understanding of how the ocean has warmed in response to long-term natural and/or anthropogenic forcing is important. A mechanistic understanding of this warming, the goal of this study, begins with an examination of the spatial variability of the observed warming, followed by modeling experiments designed to isolate the mechanisms responsible for the observed pattern. This study focuses on the North Atlantic, a basin with strong, documented climate variability (*4*–*8*) as well as unparalleled data density (*9*).

To establish the spatial pattern of heat-content change in the North Atlantic, we analyzed historical hydrographic station data from the National Oceanic Data Center World Ocean Database 2005 (*10*). Temperature data from 1950 to 2000 were binned into 2° horizontal grids for 11 constant depth layers spanning the sea surface to the ocean floor. The constraints imposed by data density and our choice of 2° spatial resolution restricted our analysis of temporal change to two time periods, 1950 to 1970 and 1980 to 2000 (fig. S1). Thus, we determine whether the ocean's heat content at

[1]Division of Earth and Ocean Sciences, Nicholas School of the Environment and Earth Science, Duke University, Durham, NC 27708, USA. [2]Department of Earth and Ocean Sciences, Liverpool University, Liverpool L69 3GP, UK.

*To whom correspondence should be addressed. E-mail: mslozier@duke.edu (M.S.L.); ric@liverpool.ac.uk (R.G.W.)
†Present address: Department of Geography, Michigan State University, East Lansing, MI 48823, USA.

the end of the 20th century was significantly different from what it was near the mid-century mark.

The integration of the 11 temperature layers leads to a significant heat-content gain for the basin as a whole [$1.610 \times 10^{22} \pm 0.19 \times 10^{22}$ J (SE); table S1] that compares relatively well to an earlier analysis of basin-integrated change (*1*). This heat-content gain between the two 20-year periods, primarily concentrated in the upper 2 km, requires an equivalent surface heat influx of 0.42 ± 0.05 W m^{-2} over the entire North Atlantic. A striking pattern to the heat-content change in the North Atlantic over the past 50 years (Fig. 1A) reveals that this basin has not experienced a uniform warming trend. Although the subtropics and tropics show an overall gain, the subpolar region has experienced a significant loss (table S1). Regional heat-content changes can be as large as $\pm 1.5 \times 10^{20}$ J between the two 20-year periods (Fig. 1A), changes that would have required a surface heat flux increase of about ±4 W m^{-2} from the former to the latter time period, an estimate that is an order of magnitude larger than the basin-averaged heat flux of 0.42 ± 0.05 W m^{-2}.

To explore mechanisms responsible for the observed North Atlantic heat-content changes, we conducted a modeling study. The ocean model experiments were run with the use of a well-documented ocean general circulation model (*11*) with 1.4° horizontal resolution spun up from rest with climatological monthly forcing fields and then integrated by using realistic surface forcing fields, winds, and buoyancy over the time period from 1950 to 2000. Given the uncertainty in surface forcing fields, all model experiments were run with reanalysis products from two agencies: European Centre for Medium-Range Weather Forecasts (ECMWF) (*12*) and National Center for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) (*13*). Because the products yielded qualitatively similar results, NCEP/NCAR-forced model results are shown as Supporting Online Material (SOM) figures. To test the model's skill at reproducing the observed heat-content changes, we used model temperature data from two 20-year model integrations, one from 1950 to 1970 and another from 1980 to 2000, to compute the modeled heat-content changes from the latter time period to the former. For both sets of forcing fields, there is a heat-content gain over the tropics and much of the subtropical gyre, yet a heat loss over the subpolar gyre (Fig. 2A and fig. S3A), a pattern broadly reflective of the observed fields (Fig. 1A), including a clear subtropical-subpolar gyre boundary separating the regions of heat gain and loss. There are differences, however, with the model integrations gaining insufficient heat over the Sargasso Sea and retaining more heat over the western side of the tropics, as well as having a loss of heat along 24°N, compared with the data. Another important difference is that the modeled heat-content changes are generally of larger amplitude than the observed changes (table S1). It is unclear whether this difference results from the inherent

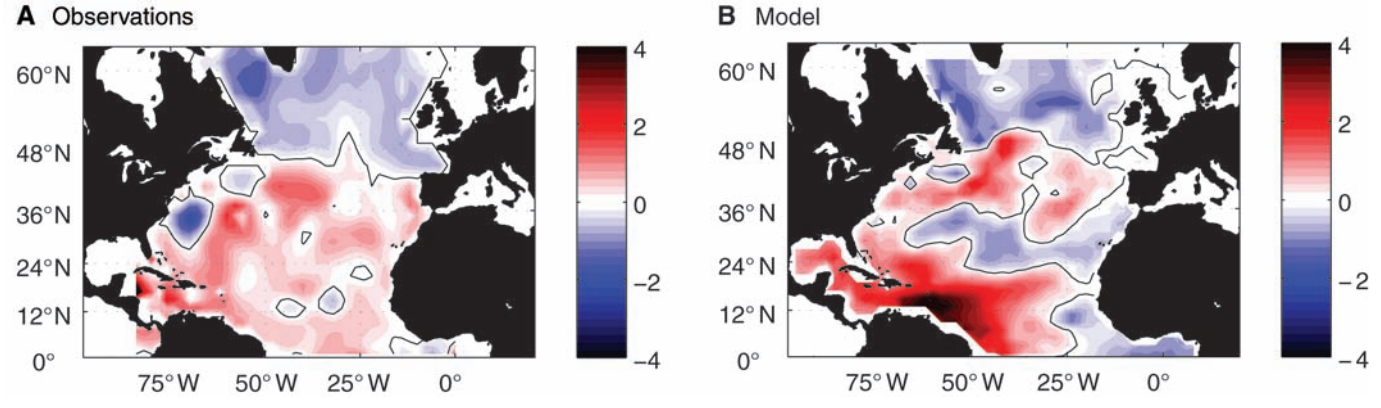


Fig. 1. Change in ocean heat content (color bar indicates units of 10^{20} J; red represents a gain in heat for the later period) between the 20-year periods 1950–1970 and 1980–2000 diagnosed from (**A**) historical data integrated over the water column and (**B**) 1.4° ocean model output using realistic ECMWF wind and buoyancy forcing after a 60-year spin-up. Observations and model data were binned onto the same 2° grid. The observations reveal an overall gain in heat in the tropics and subtropics and a loss of heat in the subpolar ocean.

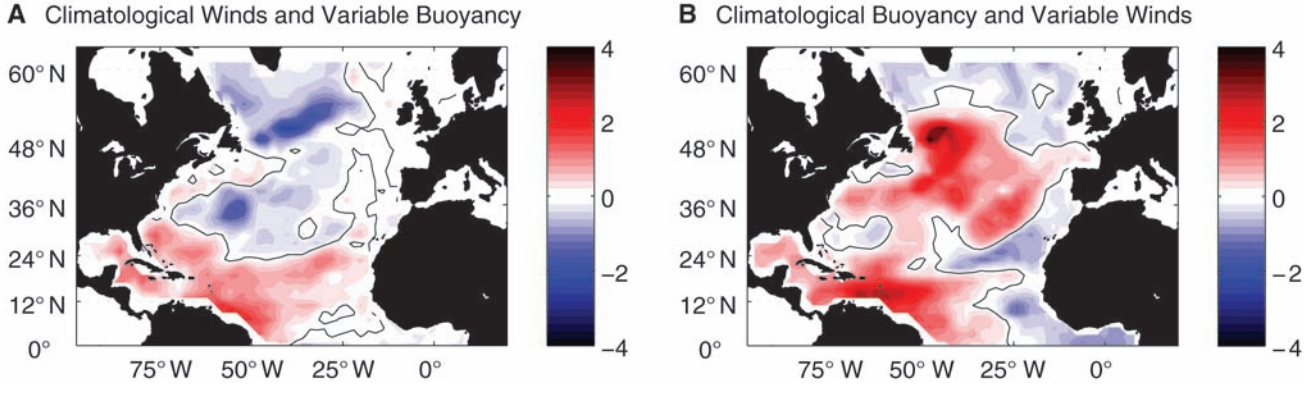


Fig. 2. Change in ocean heat content (color bar indicates 10^{20} J; red represents a gain in heat for the later period) between the 20-year periods 1950–1970 and 1980–2000 diagnosed from idealized model experiments (**A**) using realistic surface buoyancy fluxes from ECMWF, but with climatological winds, and (**B**) using realistic winds from ECMWF, but with climatological surface buoyancy fluxes. Because the ECMWF reanalysis product is available from 1958 only, the ECMWF study uses reconstructed 1950–1957 forcing from the average of 1958–1970. For this comparison, two 20-year integrations were conducted: one with forcing from 1950 to 1970 and another with forcing from 1980 to 2000. In order to separate the mechanical and thermal effects of the wind, we forced the variable-wind runs with climatological latent and sensible heat fluxes.

smoothing that results from the use of hydrographic data irregularly distributed in space and time and/or whether the difference results from a poor knowledge of heat and freshwater fluxes and boundary conditions for the model. Lastly, we note that the broad pattern of heat-content change remained unaltered when the same modeling experiments were repeated at 0.5° horizontal grid spacing (fig. S3B).

Given the broad correspondence between the large-scale patterns of heat content from the model and the data, we investigated the extent to which these patterns are controlled either by surface buoyancy fluxes or by wind-induced circulation changes. First, the model experiments were repeated with the same surface heat and freshwater fluxes described above but with climatological winds. In this case (Fig. 2A), the model integrations again lead to a gain in heat content over the tropics and a loss over the subpolar gyre, but this loss of heat extends unrealistically over the central part of the subtropical gyre, a clear mismatch with the observations. In a second experiment, the model was forced by climatological surface heat and freshwater fluxes, but with realistic winds. In this opposing case, the gain in heat content extends over the tropics and much of the subtropical gyre, but there is insufficient loss of heat over the subpolar gyre (Fig. 2B). From these two model experiments, we conclude that the gain in heat content over the tropics and loss over the subpolar gyre is a consequence of changes in air-sea heat and freshwater fluxes, whereas the gain in heat content over the subtropical gyre is primarily a consequence of wind-induced circulation changes.

The distinction in forcing mechanisms for the subtropical and subpolar regions seen in the model is largely consistent with observations: Analyses of temperature and salinity data have revealed that changes in the subpolar gyre are a reflection of density-compensated water-mass changes, whereas the subtropical changes are due to the deepening of density surfaces,

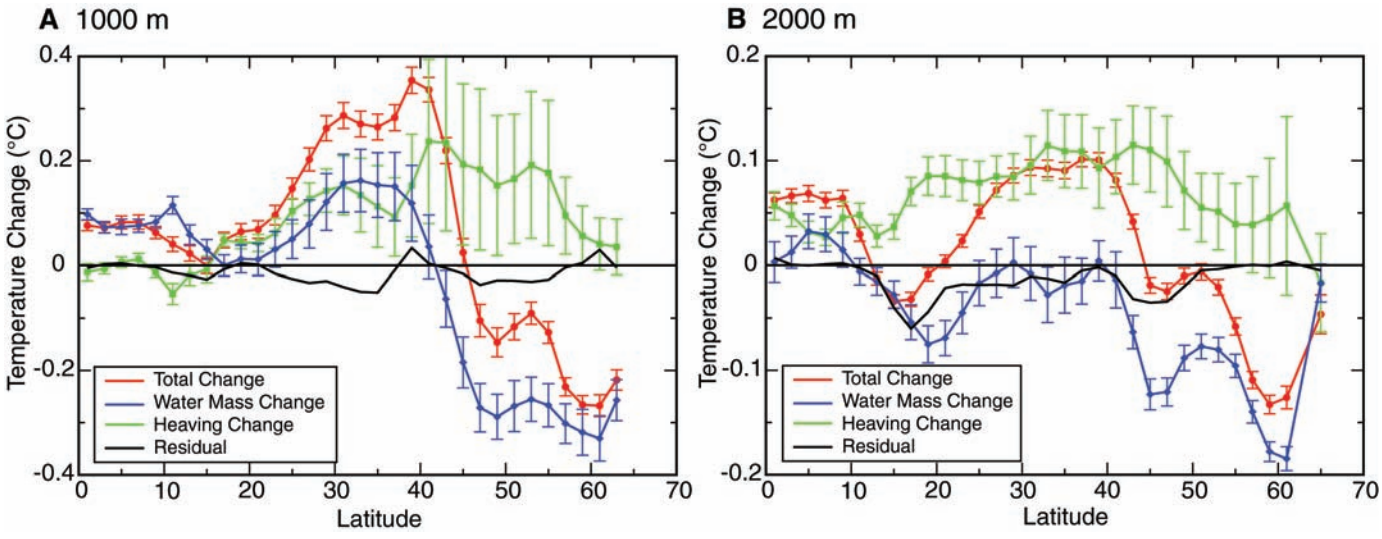


Fig. 3. Observed zonally averaged temperature change (°C) between the 20-year periods 1950–1970 and 1980–2000 along a constant depth (red) of (**A**) 1000 m or (**B**) 2000 m, revealing a warming in the tropics and subtropics but a cooling north of 45°N. This temperature change at constant depth is separated into changes along an isopycnal (blue) and the temperature change from a deepening of isopycnals (green) and a residual from the misfit of these independent estimates (black). Over the whole basin, there is a warming from the deepening of isopycnals (heave, green), which is opposed by a cooling of the water mass along the isopycnals (blue) in the subpolar gyre. The vertical displacement of the isopycnals is related to changes in the wind forcing. Error bars indicate standard errors.

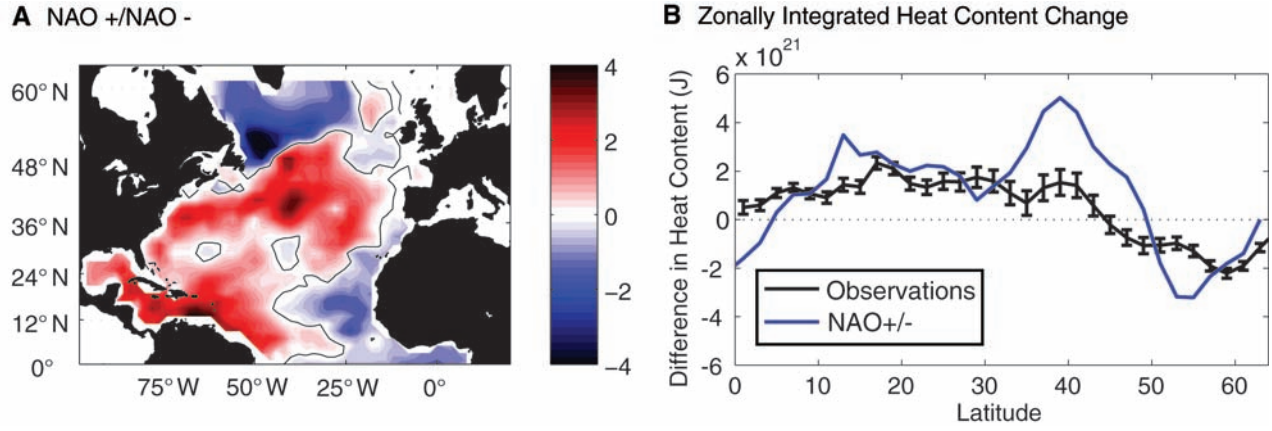


Fig. 4. (**A**) Change in ocean heat content (color bar indicates 10^{20}]; red is a gain) from a model run using NAO– composite forcing to a model run using NAO+ composite forcing. (**B**) Zonal integral in heat-content change for the observations (black; error bars represent standard errors) and for the NAO model runs (blue). Sixteen years were averaged to create a composite NAO– index of –2.34; 20 years were averaged to create a composite NAO+ index of 2.37. Model integrations were also made in which the NAO+ composite forcing was defined as the forcing fields of the 5 individual years with the most positive NAO indices over the 50-year record, randomly ordered, and then repeatedly integrated for 20 years; a NAO– state was similarly defined. Model results were similar to those shown above.

in agreement with a past study that analyzed differences from repeated hydrographic cross sections (*4*). A decomposition of the total temperature change (*14*, *15*) at a particular depth into that which is created by temperature change along a surface of constant density, attributable to water mass change, and that which is created when isotherms move vertically past a depth horizon, attributable generally to wind-forced displacement or heaving, reveals this gyre-specific response (Fig. 3). In the subpolar gyre, the cooling at both 1000 m (the base of the thermocline) and 2000 m (representative of the deep waters below the thermocline) results from colder waters residing on the density surfaces at those depths, suggesting that water-mass changes play a dominant role. In contrast, for the subtropical gyre there is a more complicated response, with the warming at 2000 m due to a deepening of isotherms and the warming at 1000 m due to both a deepening of isotherms and water-mass changes. Ironically, this simple, basinwide decomposition of temperature change over the past 50 years reveals a complexity that requires a more nuanced explanation than simple basinwide warming.

These observations of temperature changes compare well with the model results: An analysis of model output reveals that the wind-induced increase in heat content over the subtropical gyre is achieved through a redistribution of heat associated with an enhanced wind-induced pumping down of the thermocline and an increased northward transport of heat from the tropics to the northern subtropical gyre. This interpretation of how the subtropical warming is largely controlled is also supported from a comparison of hydrographic sections along 36°N conducted in 1959, 1981, and 2005. Similar heave/water-mass diagnostics applied to these data (*16*) reveal that a warming over the past 2 decades in the upper 900 m can be attributed to a wind-induced thickening of the thermocline, whereas temperature and salinity changes in the deeper waters can be accounted for by changes in the water masses spreading from the Mediterranean and Labrador Seas.

Given the prominent role of the surface forcing, we explored whether the large-scale atmospheric variability, as expressed by the North Atlantic Oscillation (NAO) index, can impart the observed oceanic heat-content change. Within the 50-year span of our observations, there has been a well-recognized shift in the NAO from a low index from 1950 to 1970 to a high index from 1980 to 2000 (*17*). To address the role of this change in large-scale atmospheric forcing, we abandoned the integration of the model with sequential forcing and instead simply ran the model for a 20-year period with atmospheric forcing from a composite of NAO– years and for a 20-year period with a composite of NAO+ years. A computation of the model differences in heat content (Fig. 4 and fig. S3C) reveals that the NAO+ ocean has a larger heat content in the subtropics and tropics compared with the NAO– ocean, whereas the opposite is true for the subpolar region. Such a contrast matches the general pattern for the observed heat-content changes (Fig. 1A), as well as that for the difference between the two model runs forced with 1950–1970 conditions and 1980–2000 conditions (Fig. 1B). A comparison of the zonally integrated heat-content changes as a function of latitude (Fig. 4B) confirms that the NAO difference can largely account for the observed gyre-specific heat-content changes over the past 50 years, although there are some notable differences in the latitudinal band from 35° to 45°N. Thus, we suggest that the large-scale, decadal changes in wind and buoyancy forcing associated with the NAO is primarily responsible for the ocean heat-content changes in the North Atlantic over the past 50 years.

This data and modeling study of heat-content changes for the North Atlantic provides several cautionary notes for the investigation and interpretation of climate signals in the ocean. First, an examination of the spatial pattern associated with the reported heat-content changes illustrates that basin-averaged changes can mask important spatial differences. Secondly, there is not a single attribution for the observed changes in the North Atlantic heat content: Changes in surface buoyancy forcing lead to tropical warming and high-latitude cooling, whereas wind-induced redistribution of heat leads to subtropical warming. Thirdly, the broad pattern of heat-content change can be accounted for by changes in the large-scale atmospheric forcing over the past 50 years.

When viewed in isolation, the net heat gain for the North Atlantic basin (+0.4 W m^{-2}) is likely explained as a small residual from the cancellation of the larger regional heat gains and losses (±4 W m^{-2}). Any anthropogenic warming would presently be masked by such strong natural variability. However, given the reported heat gain for each of the other world ocean basins (*1*, *2*, *18*) and the rising air temperatures, the relatively small basinwide heat gain is plausibly attributable to anthropogenic forcing. The overall North Atlantic heat-content change, equivalent to an average increase in the surface heat flux of +0.4 W m^{-2}, is the same sign yet slightly below the lower estimates of anthropogenic-induced radiative heating, ranging from +0.6 to +2.4 W m^{-2} since 1750 (*19*). Presumably, other parts of the global ocean and climate system have taken up the remainder of the excess heat input.

Lastly, the positive trend in the winter NAO index in the 1990s has been attributed to anthropogenic forcing (*17*), implying that the NAO could be the route by which anthropogenic warming is imprinted on the ocean. However, although most climate models show a slight strengthening of the NAO index with anthropogenic forcing, the climate models also underestimate the strength of the recent decadal trend in the NAO, raising doubts as to the viability of the connection between the NAO and anthropogenic forcing in climate models (*20*, *21*). Hence, although the change in ocean heat content over the North Atlantic can be connected to the decadal trend in the NAO, it is premature to conclusively attribute these regional patterns of heat gain to greenhouse warming. Continued long-term monitoring of North Atlantic temperatures is needed to answer the question of whether the basin-average warming is reflecting anthropogenic forcing and/or natural variability.

References and Notes

1. S. Levitus, J. I. Antonov, T. P. Boyer, C. Stephens, *Science* **287**, 2225 (2000).
2. S. Levitus, J. I. Antonov, T. P. Boyer, *Geophys. Res. Lett.* **32**, L02604 (2005).
3. S. Levitus *et al.*, *Science* **292**, 267 (2001).
4. B. K. Arbic, W. B. Owens, *J. Clim.* **14**, 4091 (2001).
5. H. L. Bryden *et al.*, *J. Clim.* **9**, 3162 (1996).
6. T. M. Joyce, R. S. Pickart, R. C. Millard, *Deep Sea Res.* **46**, 245 (1999).
7. J. R. Lazier, in *Natural Climate Variability on Decade-to-Century Time Scales* (National Academy Press, Washington, DC, 1995), p. 295.
8. R. Curry, B. Dickson, I. Yashayaev, *Nature* **426**, 826 (2003).
9. M. S. Lozier, W. B. Owens, R. G. Curry, *Prog. Oceanogr.* **36**, 1 (1995).
10. T. P. Boyer *et al.*, *World Ocean Database 2005*, S. Levitus, Ed. (National Oceanic and Atmospheric Administration Atlas National Environmental Satellite, Data, and Information Service 60, Government Printing Office, Washington, DC, 2006).
11. R. Bleck, L. T. Smith, *J. Geophys. Res.* **95**, 3273 (1990).
12. S. M. Uppala *et al.*, *Q. J. R. Meteorol. Soc.* **131**, 2961 (2005).
13. E. Kalnay *et al.*, *Bull. Am. Meteorol. Soc.* **77**, 437 (1996).
14. N. L. Bindoff, T. J. McDougall, *J. Phys. Oceanogr.* **24**, 1137 (1994).
15. The decomposition is expressed mathematically as $\frac{\partial T}{\partial t}|_z = \frac{\partial T}{\partial t}|_\rho + \frac{\partial T}{\partial z}\frac{\partial z}{\partial t}|_\rho$.
16. S. J. Leadbetter, R. G. Williams, E. L. McDonagh, B. A. King, *Geophys. Res. Lett.* **34**, L12608 (2007).
17. J. W. Hurrell, *Science* **269**, 676 (1995).
18. N. L. Bindoff *et al.*, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon *et al.*, Eds. (Cambridge Univ. Press, Cambridge, 2007).
19. P. Forster *et al.*, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon *et al.* Eds. (Cambridge Univ. Press, Cambridge, 2007).
20. N. P. Gillett, H. F. Graf, T. J. Osborn, *Geophys. Monogr. Am. Geophys. Union 134* (American Geophysical Union, Washington, DC, 2003).
21. D. B. Stephenson *et al.*, *Clim. Dyn.* **27**, 401 (2006).

22. This work was supported by grants from NSF and the UK Natural Environment Research Council. Atmospheric data sets were obtained from the British Atmospheric Data Centre.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1146436/DC1
Materials and Methods
Figs. S1 to S3
Table S1
References

13 June 2007; accepted 17 December 2007
Published online 3 January 2008;
10.1126/science.1146436
Include this information when citing this paper.

Direct and Indirect Effects of Resource Quality on Food Web Structure

Tibor Bukovinszky,[1]*† F. J. Frank van Veen,[2]† Yde Jongema,[1] Marcel Dicke[1]

The diversity and complexity of food webs (the networks of feeding relationships within an ecological community) are considered to be important factors determining ecosystem function and stability. However, the biological processes driving these factors are poorly understood. Resource quality affects species interactions by limiting energy transfer to consumers and their predators, affecting life history and morphological traits. We show that differences in plant traits affect the structure of an entire food web through a series of direct and indirect effects. Three trophic levels of consumers were influenced by plant quality, as shown by quantitative herbivore–parasitoid–secondary parasitoid food webs. We conclude, on the basis of our data, that changes in the food web are dependent on both trait- and density-mediated interactions among species.

Food webs depict networks of trophic relationships (interactions across levels of consumers) in ecological communities. A challenge in ecology is to understand the interplay between bottom-up (resource-based) and top-down (consumer-based) forces that structure food webs (*1*–*3*) and to uncover the link between food web structure and ecosystem function and stability (*4*, *5*). The species richness (diversity) and complexity of food webs are key aspects (*4*, *6*, *7*); for a given diversity and complexity, other food web attributes can be predicted with simple rules of predation created on the basis of body size (*8*).

The diversity of communities is determined by past evolutionary processes and immigration and extinction, which may be driven by both direct and indirect interactions (*9*, *10*). In food webs, complexity is usually measured as connectance: the fraction of all possible trophic links that is realized. It is determined by the diet breadth of the species in the web. Optimal foraging theory suggests that observed connectance relies on the body sizes of predators and prey (*11*). Therefore, food web structure may be influenced by variation in the traits of component species. For example, insect host–parasitoid systems are influenced by variation in plant traits (*12*). Plant nutritional quality directly affects trophic interactions by influencing the morphology, behavior, and life history of insects (*13*–*15*); but how this affects food web structure remains unknown. We studied how differences in resource quality influenced structural properties of a multitrophic food web through density- and size-mediated interactions.

Aphids feed by piercing the phloem of their food plant and are sensitive to changes in plant quality (*12*). Aphid communities formed and developed naturally for one season on 24 replicate plots of 144 plants of either a feral or domesticated (Brussels sprouts, var. *gemmifera*) population of *Brassica oleracea* L. Cultivated

[1]Laboratory of Entomology, Wageningen University, Post Office Box 8031, 6700 EH Wageningen, Netherlands. [2]Natural Environment Research Council Centre for Population Biology, Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK.

*To whom correspondence should be addressed. E-mail: tibor.bukovinszky@wur.nl

†These authors contributed equally to this work.

Table 1. Statistics of interaction webs constructed in plots of the feral *Brassica* line and Brussels sprouts and final model selection of variables explaining food web metrics. The results of analysis were derived from data presented in table S1. prop., proportion; NS, not significant.

Variable		Mean sprout (SEM)	Mean feral (SEM)	Final model	*F* value (d.f.)	*P* value
Aphid	Density (per plant)	619 (65)	2232 (318)	*Brassica* line	24.7 (1,22)	<0.001
	Relative density (prop. of *M. persicae*)	0.01 (0.003)	0.06 (0.01)	*Brassica* line + aphid density	40.3 (2,21)	<0.001
	Prop. parasitised	0.18 (0.02)	0.1 (0.01)	*Brassica* line	19.8 (1,22)	<0.001
Primary parasitoids (Pp)	Mummy density	134.5 (17)	222.9 (31.9)	Aphid density	22.4 (1,22)	<0.001
	Adult density	21.6 (1.9)	20.6 (3.5)	*Brassica* line + aphid density	5.7 (2,21)	0.011
	Diversity	1.11 (0.03)	1.32 (0.09)	*Brassica* line	4.5 (1,22)	<0.05
	Link diversity	1.36 (0.06)	1.95 (0.16)	Pp diversity + Pp density + aphid density + prop. of *M. persicae*	67.3 (4,19)	<0.001
	Connectance	0.35 (0.04)	0.38 (0.03)	NS		
	Prop. parasitised (secondary)	0.82 (0.01)	0.9 (0.01)	Mummy density + *Brassica* line	14.6 (2,21)	<0.001
	Mummy size (mm^2)	1.22 (0.01)	1.49 (0.01)	*Brassica* line	152.4 (1,22)	<0.001
Secondary parasitoids (Sp)	Adult density	113 (15)	202 (31)	Aphid density	21.1 (1,22)	<0.001
	Hyperparasitoid density	77 (10)	67 (9)	NS		
	Prop. of mummy parasitoids (of Sp.)	0.32 (0.02)	0.66 (0.01)	*Brassica* line	149.9 (1,22)	<0.001
	Diversity	2.7 (0.13)	4.4 (0.1)	*Brassica* line	103.4 (1,22)	<0.001
	Link diversity	2.8 (0.13)	4.9 (0.15)	Sp diversity (96%) + *Brassica* line + aphid density + Sp density	295.1 (4,19)	<0.001
	Connectance	0.24 (0.01)	0.35 (0.01)	Prop. of mummy parasitoids + aphid density	41.4 (2,21)	<0.001

Brassicas and their wild relatives show considerable variation in primary and secondary chemistry and morphology (*15*, *16*). The studied Brassicas differed in several traits, such as secondary chemistry, leaf thickness, and architecture. In all plots, two aphid species, *Brevicoryne brassicae* and *Myzus persicae* (Homoptera: Aphididae), were found. Differences in aphid body size (Table 1, Pp mummy size) and densities (Fig. 1 and Table 1) show that the nutritional quality of feral *Brassica* was higher than that of domesticated *Brassica*.

The aphids in these plots were commonly attacked by parasitoid wasps. Primary parasitoids (Pp's) attacking aphids are in turn parasitized by two guilds of secondary parasitoids (Sp's) that differ in the way they feed on their host. First, hyperparasitoids are koinobiont endoparasitoids (*17*). They lay their egg inside the developing primary parasitoid larva inside the living aphid host and delay development until the primary parasitoid larva has killed and mummified the aphid. Second, mummy parasitoids are idiobiont ectoparasitoids that attack the mummified aphid (fig. S1) and deposit their egg on the primary parasitoid larva or pre-pupa, on which the mummy parasitoid larva feeds externally. Because of the more direct interaction with host defenses, hyperparasitoids are typically more specialized than the generalist mummy parasitoids (*18*, *19*) [supporting online material (SOM) text]. Thus, we studied four tropic levels; plants (feral and domestic *B. oleracea*), aphid herbivores, primary parasitoids, and secondary parasitoids. Aphid-parasitoid systems are particularly suited to construction of quantitative food webs, because the parasitoid pupates inside the aphid integument, forming a so-called mummy, allowing identification of both the host and its parasitoid. Investigations of quantitative food webs, where densities and link strengths are known, have the advantage that they allow for the detection of changes in the structure that binary webs would miss (*20*) and are far less sensitive to the effects of rare species and links (*21*, *22*).

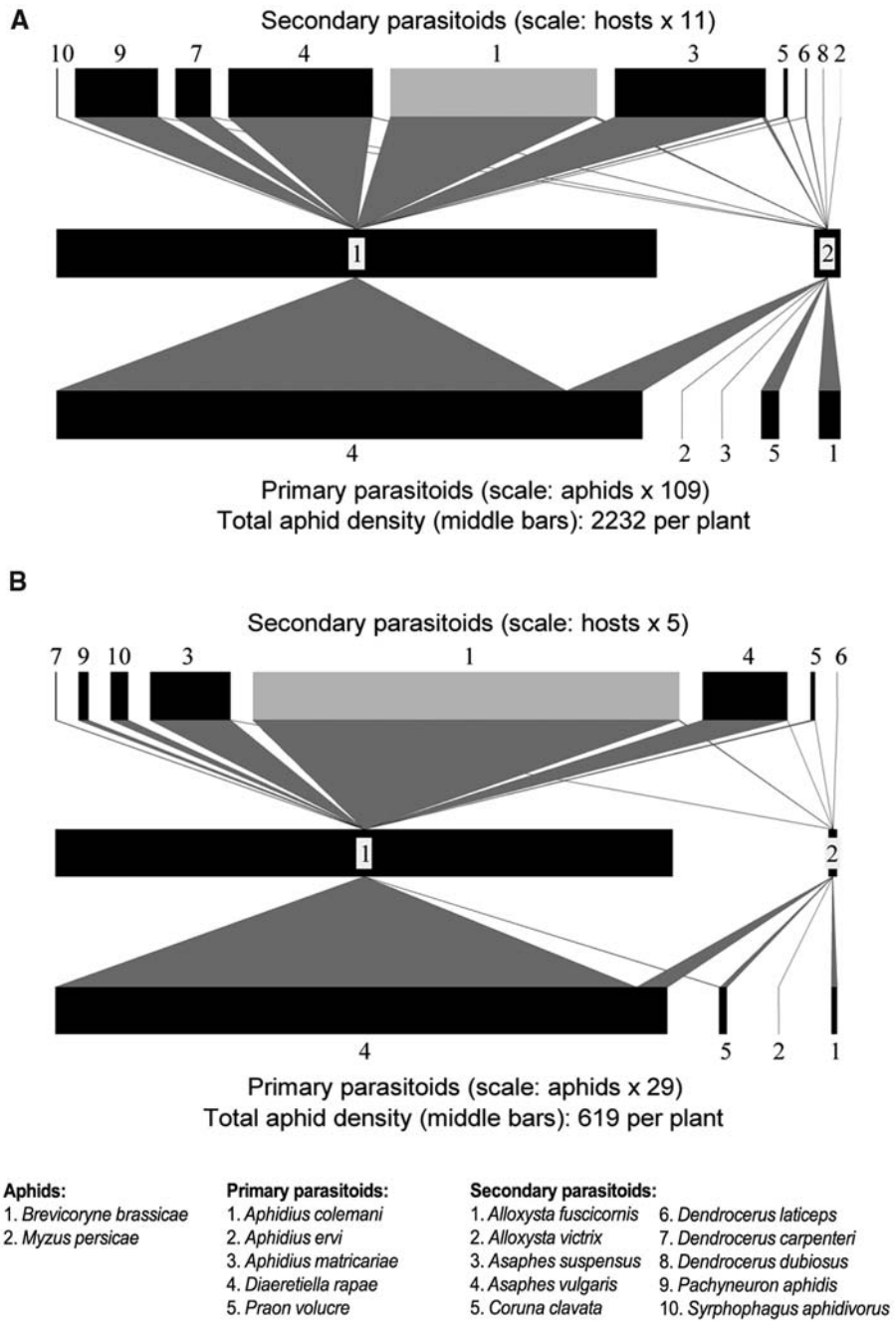


Fig. 1. Aphid-parasitoid interaction webs on (**A**) feral *Brassica* and (**B**) domestic Brussels sprouts pooled across replicated plots (fig. S2). The width of the bars is proportional to abundance. Middle bars are aphids, bottom bars are primary parasitoids, and top bars are secondary parasitoids, with gray and black bars depicting hyperparasitoids and mummy parasitoids, respectively. Wedges indicate the relative contribution of a host to a parasitoid population. Numbers refer to species identity. Note the predominance of mummy parasitoids on feral *Brassica* and associated density of links.

Aphids were attacked by five species of primary parasitoid, all of which were found on the plots of feral *Brassica*, whereas four of them were present on the Brussels sprouts (Fig. 1). The number of parasitized aphids was higher on feral *Brassica* (Table 1, Pp mummy density) but the proportion of parasitism was lower (Table 1, aphid proportion parasitized). However, the densities of emerging primary parasitoids on the two types of plants were very similar (Table 1, Pp adult density). The higher resource flow from aphid to primary parasitoids on feral *Brassica* was passed on to the next trophic level, which was reflected in a significant 1.8-fold increase in the density of the secondary parasitoids (Table 1, Sp adult density). We calculated a diversity index, taking into account the relative densities (see materials and methods in SOM). This shows that diversity among primary parasitoids was significantly higher on feral *Brassica* (Table 1, Pp diversity) and resulted in a greater diversity of aphid–primary parasitoid links. (Table 1, Pp link diversity).

Primary parasitoids were attacked by 10 species of secondary parasitoid, all of which were present on feral *Brassica* and 8 of which were present on Brussels sprouts (Fig. 1). Species diversity and link diversity within secondary parasitoid food webs were respectively 1.6 and 1.8 times higher in plots of feral *Brassica* (Table 1, Sp diversity and Sp link diversity). The diversity measure reflects both the mean number (±SEM) of species (7.1 ± 0.3 versus 5.9 ± 0.2 in feral and domestic *Brassica*, respectively) and the evenness of species abundances (0.63 ± 0.02 versus 0.46 ± 0.02). The connectance for secondary parasitoids with

aphids was 1.5 times greater in feral than in domestic *Brassica* (Table 1, Sp connectance), with 74% (R^2 from general linear model analysis) of the variation explained by the proportion of mummy parasitoids among secondary parasitoids, which was twice as high in feral *Brassica* (Table 1 and Fig. 1). We saw no evidence that the host *Brassica* affected the density of specialist hyperparasitoids, suggesting that the differences in secondary parasitoid density and diversity were entirely due to the generalist mummy parasitoids.

Our data show that variation in plant quality has cascading effects across trophic levels as mediated by changes in the abundance and size of resource items (Fig. 2). The bottom of the food chain—the plant type—affects the herbivorous aphids, which in turn controls the quality and abundance of the primary parasitoids, and finally affects the top of the food web, the secondary parasitoids. These direct and indirect interactions are mediated by both the densities and traits (here size) of herbivores. Most interesting is the effect of aphid size on the non-adjacent secondary parasitoid trophic level. Larger aphids result in larger mummies, leading to greater representation of the generalist mummy parasitoid guild (fig. S3), resulting in greater link diversity and connectance of secondary parasitoids. The strength of the pathways indicates this importance of trait-mediated effects (mummy size) in affecting secondary parasitoid communities (Fig. 2). Further evidence for trait-mediated effects was that larger mummies were more likely to yield mummy parasitoids and this effect was stronger in the feral *Brassica* (χ^2_3 = 9.41, $P < 0.05$).

We also investigated the effects of host plants on size (as a proxy for fitness) and population structure (sex ratio) of primary and secondary parasitoids. For the four most common parasitoids in our study, adult size significantly depended on aphid host size ($F_{1, 793} = 571.13$, $P < 0.001$), affecting fitness by increasing fecundity and longevity (*19*). Host size was greatest on feral *Brassica* (Fig. 3A; $F_{1, 794} = 164.2$, $P < 0.001$), resulting in significantly larger parasitoid offspring (Fig. 3B; $F_{1, 797} = 78.73$, $P < 0.001$). Furthermore, we observed 11% more females of the primary parasitoid *Diaeretiella rapae* ($\chi^2_1 = 4.36$, $P < 0.05$) and the mummy parasitoid *Asaphes vulgaris* ($\chi^2_1 = 7$, $P < 0.01$) on the feral *Brassica* (fig. S4). Female parasitoids emerged from larger hosts than did males ($F_{3, 794} = 59.28$, $P < 0.001$), suggesting that sex allocation in parasitoids is a function of host quality and the quality of their primary nutrient source.

On the basis of these data, we concluded that the feral *Brassica* was a better host for the aphids and, indirectly, the parasitoids than was the domestic line. It is likely that this increased size, and therefore fitness, in both aphids and the primary and secondary parasitoids was the result of differences in plant traits such as plant metabolites, defense chemicals, and architecture, which we hypothesize caused the observed differences in food web structure. Our

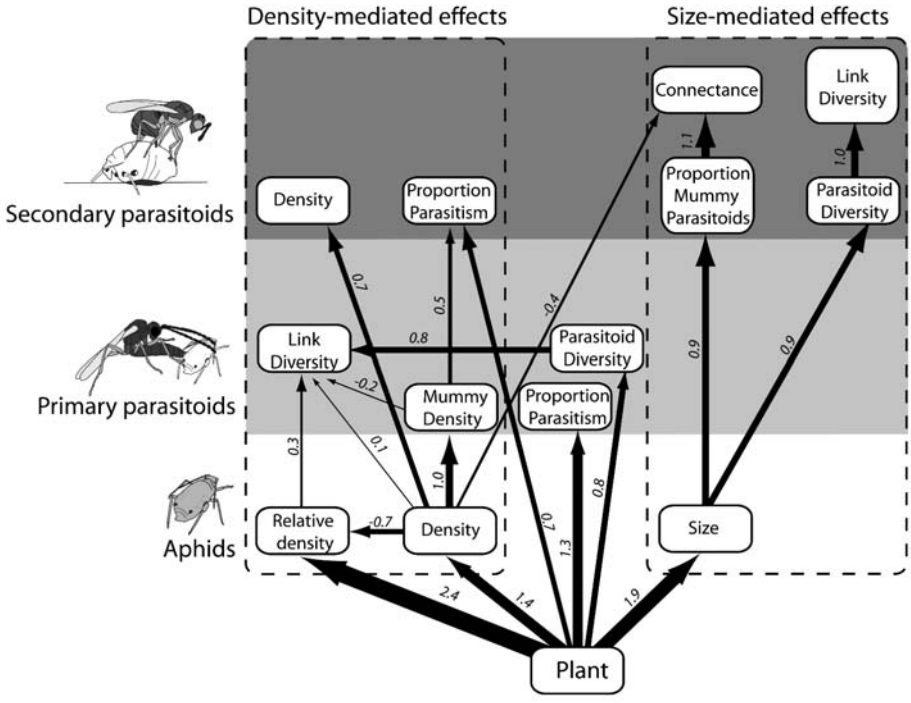


Fig. 2. Summary diagram of direct and indirect effects of plant quality on the structure of aphid-parasitoid communities. Arrow thickness is scaled to standardized coefficients from path analysis to illustrate the relative strength of effects.

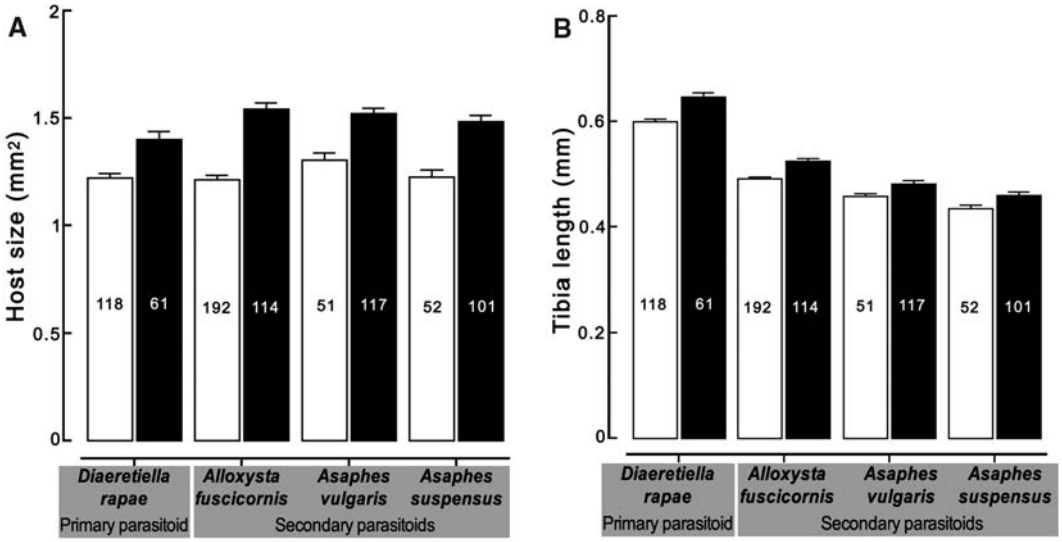


Fig. 3. Mean (+SEM) (**A**) host size and (**B**) hind tibia length of the four dominant parasitoid species in plots of the feral (black bars) and domestic (white bars) *Brassica*. Numbers show sample sizes.

study indicates that changes in resource traits influence food web diversity and complexity by interacting with foraging biology (indicated by size-dependent parasitism and sex allocation) of consumers across several trophic levels through a cascade of density- and trait-mediated effects (Fig. 2), with implications for food web stability and ecosystem functioning.

References and Notes

1. M. D. Hunter, P. W. Price, *Ecology* **73**, 724 (1992).
2. O. J. Schmitz, P. A. Hambäck, A. P. Beckerman, *Am. Nat.* **155**, 141 (2000).
3. L. Oksanen, S. D. Fretwell, J. Arruda, P. Niemalä, *Am. Nat.* **118**, 240 (1981).
4. J. M. Montoya, S. L. Pimm, R. V. Solé, *Nature* **442**, 259 (2006).
5. A. R. Ives, S. R. Carpenter, *Science* **317**, 58 (2007).
6. K. S. McCann, *Nature* **405**, 228 (2000).
7. A. R. Ives, B. J. Cardinale, *Nature* **429**, 174 (2004).
8. R. J. Williams, N. D. Martinez, *Nature* **404**, 180 (2000).
9. F. J. F. Van Veen, R. J. Morris, H. C. J. Godfray, *Annu. Rev. Entomol.* **51**, 187 (2006).
10. R. J. Morris, O. T. Lewis, H. C. J. Godfray, *Nature* **428**, 310 (2004).
11. A. P. Beckerman, O. L. Petchey, P. H. Warren, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13745 (2006).
12. M. Omacini, E. J. Chaneton, C. M. Ghersa, C. B. Muller, *Nature* **409**, 78 (2001).
13. N. Underwood, M. D. Rausher, *Ecology* **81**, 1565 (2000).
14. P. J. Ode, *Annu. Rev. Entomol.* **51**, 163 (2006).
15. J. A. Harvey, N. M. Van Dam, R. Gols, *J. Anim. Ecol.* **72**, 520 (2003).
16. C. Gómez-Campo, S. Prakash, in *Developments in Plant Genetics and Breeding*, C. Gomez-Campo, Ed. (Elsevier, Amsterdam, 1999), chap. 4.
17. H. C. J. Godfray, *Parasitoids: Behavioral and Evolutionary Ecology* (Princeton Univ. Press, Princeton, NJ, ed. 1, 1994), pp. 9–10.
18. C. B. Müller, I. C. T. Adriaanse, R. Belshaw, H. C. J. Godfray, *J. Anim. Ecol.* **68**, 346 (1999).
19. D. J. Sullivan, W. Völkl, *Annu. Rev. Entomol.* **44**, 291 (1999).
20. J. M. Tylianakis, T. Tscharntke, O. T. Lewis, *Nature* **445**, 202 (2007).
21. F. Bersier, C. Banasek-Richter, M. F. Cattin, *Ecology* **83**, 2394 (2002).
22. F. J. F. van Veen, C. B. Müller, J. K. Pell, H. C. J. Godfray, *J. Anim. Ecol.* **77**, 191 (2008).

23. The farm of Wageningen University prepared the field layout. G. Bukovinszkine Kiss and V. Taravel helped in collecting data. J. A. Harvey provided feral *Brassica* seeds. Food webs were drawn with code by H. C. J. Godfray. M.D. and T.B. were funded by the Netherlands Organisation for Scientific Research–Earth and Life Sciences Council (NWO-ALW, VICI grant 865.03.002) and F.J.F.vV. by the Natural Environment Research Council, UK. Voucher specimens were deposited at the Laboratory of Entomology (Wageningen University, reference number Buko2005.001).

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5864/804/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
Table S1

25 July 2007; accepted 27 December 2007
10.1126/science.1148310

Biomechanical Energy Harvesting: Generating Electricity During Walking with Minimal User Effort

J. M. Donelan,[1]* Q. Li,[1] V. Naing,[1] J. A. Hoffer,[1] D. J. Weber,[2] A. D. Kuo[3]

We have developed a biomechanical energy harvester that generates electricity during human walking with little extra effort. Unlike conventional human-powered generators that use positive muscle work, our technology assists muscles in performing negative work, analogous to regenerative braking in hybrid cars, where energy normally dissipated during braking drives a generator instead. The energy harvester mounts at the knee and selectively engages power generation at the end of the swing phase, thus assisting deceleration of the joint. Test subjects walking with one device on each leg produced an average of 5 watts of electricity, which is about 10 times that of shoe-mounted devices. The cost of harvesting—the additional metabolic power required to produce 1 watt of electricity—is less than one-eighth of that for conventional human power generation. Producing substantial electricity with little extra effort makes this method well-suited for charging powered prosthetic limbs and other portable medical devices.

Humans are a rich source of energy. An average-sized person stores as much energy in fat as a 1000-kg battery (*1*, *2*). People use muscle to convert this stored chemical energy into positive mechanical work with peak efficiencies of about 25% (*3*). This work can be performed at a high rate, with 100 W easily sustainable (*1*). Many devices take advantage of human power capacity to produce electricity, including hand-crank generators as well as wind-up flashlights, radios, and mobile phone chargers (*4*). A limitation of these conventional methods is that users must focus their attention on power generation at the expense of other activities, typically resulting in short bouts of generation. For electrical power generation over longer durations, it would be desirable to harvest energy from everyday activities such as walking.

It is a challenge, however, to produce substantial electricity from walking. Most energy-harvesting research has focused on generating electricity from the compression of the shoe sole, with the best devices generating 0.8 W (*4*). A noteworthy departure is a spring-loaded backpack (*5*) that harnesses the vertical oscillations of a 38-kg load to generate as much as 7.4 W of electricity during fast walking. This device has a markedly low "cost of harvesting" (COH), a dimensionless quantity defined as the additional metabolic power in watts required to generate 1 W of electrical power

$$\mathrm{COH} = \frac{\Delta \text{ metabolic power}}{\Delta \text{ electrical power}} \qquad (1)$$

where Δ refers to the difference between walking while harvesting energy and walking while carrying the device but without harvesting energy. The COH for conventional power generation is simply related to the efficiency with which (i) the device converts mechanical work to electricity and (ii) muscles convert chemical energy into positive work

$$\mathrm{COH}_{\text{conventional generation}} = \frac{\Delta \text{ metabolic power}}{\Delta \text{ electrical power}} = \frac{1}{\text{device eff} \times \text{muscle eff}} \qquad (2)$$

The backpack's device efficiency is about 31% (*5*), and muscle's peak efficiency is about 25% (*3*), yielding an expected COH of 12.9. But the backpack's actual COH of 4.8 ± 3.0 (mean ± SD) is less than 40% of the expected amount. Its economy appears to arise from reducing the energy expenditure of walking with loads (*6*, *7*). No device has yet approached the power generation of the backpack without the need to carry a heavy load.

We propose that a key feature of how humans walk may provide another means of economical energy harvesting. Muscles cyclically perform positive and negative mechanical work within each stride (Fig. 1A) (*8*). Mechanical work is required to redirect the body's center of mass between steps (*9*, *10*) and simply to move the legs back and forth (*11*, *12*). Even though the average mechanical work performed on the body over an entire stride is zero, walking exacts a metabolic cost because both

[1]School of Kinesiology, Simon Fraser University (SFU), Burnaby, BC V5A 1S6, Canada. [2]Department of Physical Medicine and Rehabilitation, University of Pittsburgh, Pittsburgh, PA 15213, USA. [3]Departments of Mechanical Engineering and Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA.

*To whom correspondence should be addressed. E-mail: mdonelan@sfu.ca

positive and negative muscle work require metabolic energy (*3*). Coupling a generator to leg motion would generate electricity throughout each cycle, increasing the load on the muscles during acceleration but assisting them during deceleration (Fig. 1B). Although generating electricity during the acceleration phase would exact a substantial metabolic cost, doing so during the deceleration phase would not, resulting in a lower COH than for conventional generation. An even lower COH could be achieved by selectively engaging the generator only during deceleration (Fig. 1C), similar to how regenerative braking generates power while decelerating a hybrid car (*13*). Here, "generative braking" produces electricity without requiring additional positive muscle power (*14*). If implemented effectively, metabolic cost could be about the same as that for normal walking, so energy would be harvested with no extra user effort (*15*).

We developed a wearable, knee-mounted prototype energy harvester to test the generative-braking concept (Fig. 2). Although other joints might suffice, we focused on the knee because it performs mostly negative work during walking (*16*). The harvester comprises an orthopedic knee brace configured so that knee motion drives a gear train through a one-way clutch, transmitting only knee extension motion at speeds suitable for a dc brushless motor that serves as the generator (*17*). For convenient testing, generated electrical power is then dissipated with a load resistor rather than being used to charge a battery. The device efficiency, defined as the ratio of the electrical power output to the mechanical power input, was empirically estimated to be no greater than 63%, yielding an estimated COH for conventional generation of 6.4 (Eq. 2). A potentiometer senses knee angle, which is fed back to a computer controlling a relay switch in series with the load resistor, allowing the electrical load to be selectively disconnected in real time. For generative braking, we programmed the harvester to engage only during the end of the swing phase (Fig. 3), producing electrical power while simultaneously assisting the knee flexor muscles in decelerating the knee. We compared this mode against a continuous-generation mode that harvests energy whenever the knee is extending (*18*). We could also manually disengage the clutch and completely decouple the gear train and generator from knee motion. This disengaged mode served as a control condition to estimate the metabolic cost of carrying the harvester mass, independent of the cost of generating electricity.

Energy-harvesting performance was tested on six male subjects who wore a device on each leg while walking on a treadmill at 1.5 m s^{-1}. We estimated metabolic cost using a standard respirometry system and measured the electrical power output of the generator (Fig. 3C). In the continuous-generation mode (Fig. 4A), subjects generated 7.0 ± 0.7 W of electricity with an insignificant 18 ± 24 W (P = 0.07) increase in metabolic cost over that of the control condition (*19*). In the generative-braking mode (Fig. 4B), subjects generated 4.8 ± 0.8 W of electricity with an insignificant 5 ± 21 W increase in metabolic cost as compared with that of the control condition (P = 0.6). For context, this electricity is sufficient to power 10 typical cell phones simultaneously (*5*). The results dem-

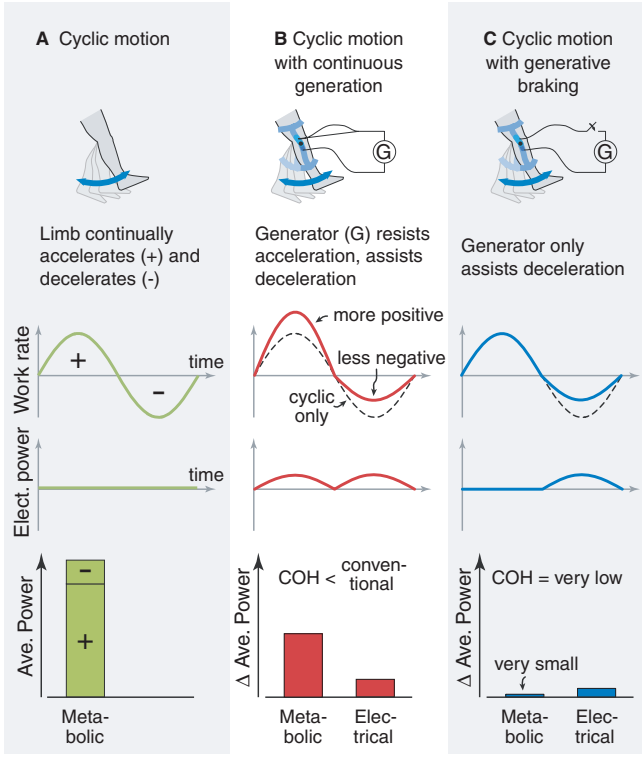


Fig. 1. Theoretical advantages of generative braking during cyclic motion, comparing the back-and-forth motion of the knee joint without power generation (**A**) against a generator operating continuously (**B**) and against a generator operating only during braking (**C**). Each column of plots shows the rate of work performed by muscles (work rate) and the electricity (elect. power) generated over time, as well as the average metabolic power expended by the human and the resulting average electrical power (ave. power bar graphs). In (B) and (C), work rate is compared against that for (A), denoted by dashed lines, and average power is shown as the difference (Δ ave. power) with respect to (A). COH is defined as the ratio of the electrical to metabolic Δ ave. powers.

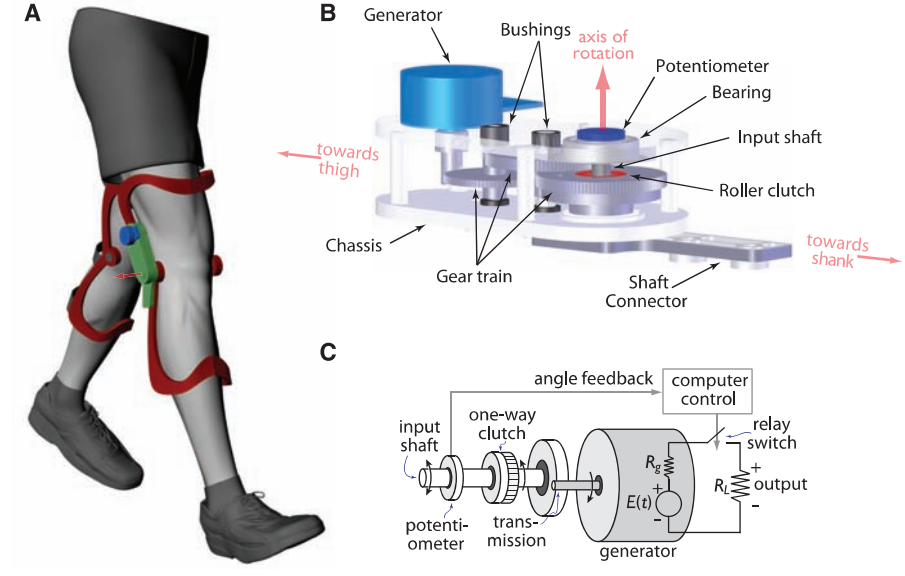


Fig. 2. Biomechanical energy harvester. (**A**) The device has an aluminum chassis (green) and generator (blue) mounted on a customized orthopedic knee brace (red), totaling 1.6-kg mass, with one worn on each leg. (**B**) The chassis contains a gear train that converts low velocity and high torque at the knee into high velocity and low torque for the generator, with a one-way roller clutch that allows for selective engagement of the gear train during knee extension only and no engagement during knee flexion. (**C**) The schematic diagram shows how a computer-controlled feedback system determines when to generate power using knee-angle feedback, measured with a potentiometer mounted on the input shaft. Generated power is dissipated in resistors. R_g, generator internal resistance; R_L, output load resistance; $E(t)$, generated voltage.

onstrate that substantial electricity could be generated with minimal increase in user effort.

The corresponding COH values highlight the advantage of generative braking (Fig. 4). Average COH in generative braking was only 0.7 ± 4.4; less than 1 W of metabolic power was required to generate 1 W of electricity. This is significantly less than the COH of 6.4 expected for conventional generation ($P = 0.01$). The COH in continuous generation, 2.3 ± 3.0, was also significantly lower than that for conventional generation ($P = 0.01$), indicating that the former mode also generated some of its electricity from the deceleration of the knee. The difference between the two modes, 2.2 ± 0.7 W of electricity, came at a difference in metabolic cost of 13 ± 12 W ($P = 0.05$). A COH taken from the average ratio of these differences yields 5.7 ± 6.2, which is nearly the same as that expected of conventional generation ($P = 0.4$). This indicates that continuous generation of power at the knee during walking produces electricity partially by conventional generation with a high COH and partially by generative braking with a very low COH. But generative braking, with less than one-eighth the COH of conventional generation, benefits almost entirely from the deceleration of the knee.

This preliminary demonstration could be improved substantially. We constructed the prototype for convenient experimentation, leading to a control condition about 20% more metabolically costly than normal walking: The disengaged-clutch mode required an average metabolic power of 366 ± 63 W as compared with 307 ± 64 W for walking without wearing the devices. The increase in cost is due mainly to the additional mass and its location, because the lower a given mass is placed, the more expensive it is to carry (*20*, *21*). Although the current increase in metabolic cost is unacceptably high for most practical implementations, revisions to improve the fit, weight, and efficiency of the device can not only reduce the cost but also increase the generated electricity. A generator designed specifically for this application could have lower internal losses and require a smaller, lighter gear train. Commercially available gear trains can have much lower friction and higher efficiency, in more compact and lightweight forms. Relocating the device components higher would decrease the metabolic cost of carrying that mass. A more refined device would also benefit from a more form-fitting knee brace made out of a more lightweight material such as carbon fiber.

Several potential applications are especially suited for generative braking. These include lighting and communications needs for the quarter of the world's population who currently live without electricity supply (*22*). Innovative prosthetic knees and ankles use motors to assist walking, but battery technology limits their power and working time (*23*–*25*). Energy harvesters worn on human joints may prove useful for powering the robotic artificial joints. In implantable devices, such as neurostimulators and drug pumps, battery power limits device sophistication, and battery replacement requires surgery (*26*). A future energy harvester might be implanted alongside such a device, perhaps in parallel with a muscle, and use generative braking to provide substantial power indefinitely. Generative braking might then find practical applications in forms very different from that demonstrated here.

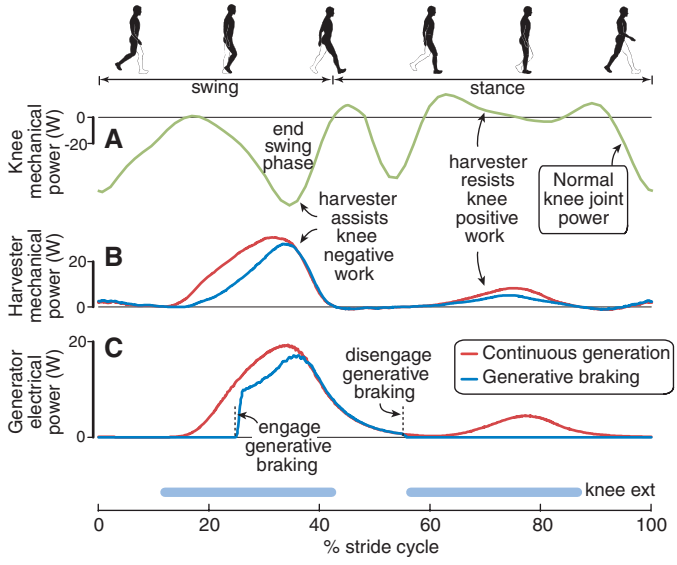


Fig. 3. Timing of power generation during walking. Time within a stride cycle, beginning with the swing phase, is shown at the bottom. The shaded bars indicate when the knee is extending and the energy harvester's clutch is engaged. (**A**) The pattern of knee mechanical power during normal walking illustrates that the knee typically generates a large amount of negative power at the end of the swing phase (*16*). (**B**) Mechanical power performed on the harvester over time, shown for continuous generation (red line) and generative braking (blue line). (**C**) Generated electrical power over time, also for both types of generation.

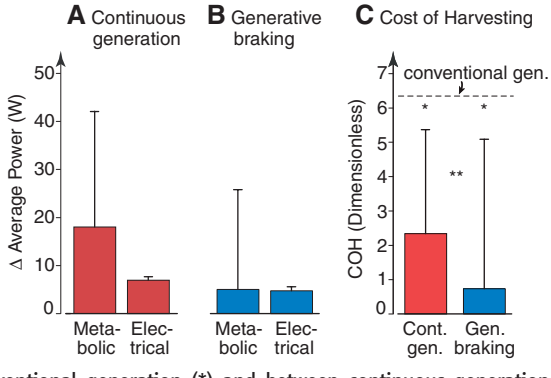


Fig. 4. Average metabolic cost and generated electricity for continuous generation (**A**) and generative braking (**B**), with change in metabolic cost (Δ average power) shown relative to the control condition. (**C**) COH (see Fig. 1) for continuous generation and generative braking as compared against that for conventional generation (dashed line). In both modes, a fraction of the harvested energy is generated from the deceleration of the knee rather than directly from muscle action. Error bars in (A) to (C) indicate SD. Asterisks indicate significant differences with conventional generation (*) and between continuous generation and generative braking (**) ($P < 0.05$ for all comparisons).

References and Notes

1. G. A. Brooks, T. D. Fahey, K. M. Baldwin, *Exercise Physiology: Human Bioenergetics and Its Applications* (McGraw-Hill, Boston, ed. 4, 2005).
2. T. Starner, *IBM Syst. J.* **35**, 618 (1996).
3. R. Margaria, *Int. Z. Angew. Physiol.* **25**, 339 (1968).
4. J. A. Paradiso, T. Starner, *IEEE Pervasive Comput.* **4**, 18 (2005).
5. L. C. Rome, L. Flynn, E. M. Goldman, T. D. Yoo, *Science* **309**, 1725 (2005).
6. A. D. Kuo, *Science* **309**, 1686 (2005).
7. L. C. Rome, L. Flynn, T. D. Yoo, *Nature* **444**, 1023 (2006).
8. H. Elftman, *Am. J. Physiol.* **125**, 339 (1939).
9. A. D. Kuo, J. M. Donelan, A. Ruina, *Exerc. Sport Sci. Rev.* **33**, 88 (2005).
10. R. Margaria, *Biomechanics and Energetics of Muscular Exercise* (Clarendon, Oxford, 1976).
11. J. Doke, J. M. Donelan, A. D. Kuo, *J. Exp. Biol.* **208**, 439 (2005).
12. R. L. Marsh, D. J. Ellerby, J. A. Carr, H. T. Henry, C. I. Buchanan, *Science* **303**, 80 (2004).
13. N. Demirdoven, J. Deutch, *Science* **305**, 974 (2004).
14. We have chosen terminology that distinguishes generative braking from regenerative braking because the electricity produced in regenerative braking is reused to power the motion of the hybrid automobile. In generative braking, the electricity is not reused to power walking.
15. Knee-joint power has contributions from forces generated by muscle fibers, tendons, connective tissue, and other passive soft tissues. The actual change in metabolic cost with generative braking depends on the relative contribution of muscle fibers to decelerating the knee joint. If muscle fibers are generating the negative power, a reduction in metabolic cost is expected. This is also true if muscle fibers are active but isometric. If the deceleration is due entirely to passive forces from elastic and plastic deformations of soft tissues, no change in metabolic cost is expected. What actually occurs at the knee during the end of the swing phase is unclear, precluding a quantitative prediction of the change in metabolic cost.
16. D. A. Winter, *Biomechanics and Motor Control of Human Movement* (Wiley, New York, ed. 2, 1990).
17. Additional methodological details, results, and videos are available as supporting material on *Science* Online.
18. The one-way clutch prevents the device from generating electricity from flexion. Nevertheless, we refer to this mode as continuous generation because, unlike the generative-braking mode, electricity is continually generated from extension regardless of whether the motion is accelerating or decelerating.

19. We used *t* tests to determine whether there was a statistical difference between conditions with an alpha level of 0.05. All electrical power comparisons were statistically significant.
20. R. C. Browning, J. R. Modica, R. Kram, A. Goswami, *Med. Sci. Sports Exerc.* **39**, 515 (2007).
21. R. G. Soule, R. F. Goldman, *J. Appl. Physiol.* **27**, 687 (1969).
22. International Energy Agency, *World Energy Outlook* (IEA Books, Paris, 2006).
23. D. Berry, *Phys. Med. Rehabil. Clin. N. Am.* **17**, 91 (2006).
24. J. L. Johansson, D. M. Sherrill, P. O. Riley, P. Bonato, H. Herr, *Am. J. Phys. Med. Rehabil.* **84**, 563 (2005).
25. R. Seymour *et al*., *Prosthet. Orthot. Int.* **31**, 51 (2007).
26. O. Soykan, in *Business Briefing: Medical Device Manufacturing and Technology*, E. Cooper, Ed. (World Markets Research Centre, London, 2002), pp. 76–80.
27. Supported by a Natural Sciences and Engineering Research Council (NSERC) grant I2IPJ/326586-05 to J.M.D. and J.A.H., a Michael Smith Foundation for Health Research (MSFHR) Scholar Award to J.M.D., a Canadian Institutes of Health Research New Investigator Award to J.M.D., a MSFHR Postdoctoral Trainee Award to Q.L, and an NSERC Undergraduate Student Researcher Award to V.N. We thank Ossur for providing the knee braces, as well as S. H. Collins, R. Kram, A. Ruina, and the SFU Locomotion Lab for their helpful comments and suggestions. J.M.D. is chief science officer and board member of Bionic Power, Incorporated. J.M.D., Q.L., J.A.H., D.J.W., and A.D.K. have equity interest in Bionic Power, Incorporated, which performs research and development on the energy-harvesting technology reported in this paper.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5864/807/DC1
Materials and Methods
Figs. S1 to S3
Table S1
References
Movies S1 to S4

29 August 2007; accepted 3 January 2008
10.1126/science.1149860

Three-Dimensional Super-Resolution Imaging by Stochastic Optical Reconstruction Microscopy

Bo Huang,[1,2] Wenqin Wang,[3] Mark Bates,[4] Xiaowei Zhuang[1,2,3]*

Recent advances in far-field fluorescence microscopy have led to substantial improvements in image resolution, achieving a near-molecular resolution of 20 to 30 nanometers in the two lateral dimensions. Three-dimensional (3D) nanoscale-resolution imaging, however, remains a challenge. We demonstrated 3D stochastic optical reconstruction microscopy (STORM) by using optical astigmatism to determine both axial and lateral positions of individual fluorophores with nanometer accuracy. Iterative, stochastic activation of photoswitchable probes enables high-precision 3D localization of each probe, and thus the construction of a 3D image, without scanning the sample. Using this approach, we achieved an image resolution of 20 to 30 nanometers in the lateral dimensions and 50 to 60 nanometers in the axial dimension. This development allowed us to resolve the 3D morphology of nanoscopic cellular structures.

Far-field optical microscopy offers three-dimensional (3D) imaging of biological specimens with minimal perturbation and biomolecular specificity when combined with fluorescent labeling. These advantages make fluorescence microscopy one of the most widely used imaging methods in biology. The diffraction barrier, however, limits the imaging resolution of conventional light microscopy to 200 to 300 nm in the lateral dimensions, leaving many intracellular organelles and molecular structures unresolvable. Recently, the diffraction limit has been surpassed and lateral imaging resolutions of 20 to 50 nm have been achieved by several "super-resolution" far-field microscopy techniques, including stimulated emission depletion (STED) and its related RESOLFT (reversible saturable optically linear fluorescent transitions) microscopy (*1*, *2*); saturated structured illumination microscopy (SSIM) (*3*); stochastic optical reconstruction microscopy (STORM) (*4*, *5*); photoactivated localization microscopy (PALM) (*6*, *7*); and other methods using similar principles (*8*–*10*).

Although these techniques have improved 2D image resolution, most organelles and cellular structures cannot be resolved without high-resolution imaging in all three dimensions. Three-dimensional fluorescence imaging is most commonly performed using confocal or multiphoton microscopy, the axial resolution of which is typically in the range of 500 to 800 nm (*11*, *12*). The axial imaging resolution can be improved to roughly 100 nm by 4Pi and I^{5}M microscopy (*13*–*15*). Furthermore, an axial resolution as high as 30 to 50 nm has been obtained with STED along the axial direction using the 4Pi illumination geometry, but the same imaging scheme does not provide super resolution in the lateral dimensions (*1*).

Here, we demonstrate 3D STORM imaging with a spatial resolution that is 10 times better than the diffraction limit in all three dimensions without invoking sample or optical-beam scanning. STORM and PALM rely on single-molecule detection (*16*) and exploit the photoswitchable nature of certain fluorophores to temporally separate the otherwise spatially overlapping images of numerous molecules, thereby allowing the high-precision localization of individual molecules (*4*–*7*, *9*). Limited only by the number of photons detected (*17*), localization accuracies as high as 1 nm can be achieved in the lateral dimensions for a single fluorescent dye at ambient conditions (*18*). Not only can the lateral position of a particle be determined from the centroid of its image (*19*, *20*), the shape of the image also contains information about the particle's axial (*z*) position. Nanoscale localization accuracy has been achieved in the *z* dimension by introducing defocusing (*21*–*24*) or astigmatism (*25*, *26*) into the image, without substantially compromising the lateral positioning capability.

In this work, we used the astigmatism imaging method to achieve 3D STORM imaging. To this end, a weak cylindrical lens was introduced into the imaging path to create two slightly different focal planes for the *x* and *y* directions (Fig. 1A) (*25*, *26*). As a result, the ellipticity and orientation of a fluorophore's image varied as its position changed in *z* (Fig. 1A): When the fluorophore was in the average focal plane [approximately halfway between the *x* and *y* focal planes where the point spread function (PSF) has equal widths in the *x* and *y* directions], the image appeared round; when the fluorophore was above the average focal plane, its image was more focused in the *y* direction than in the *x* direction and thus appeared ellipsoidal with its long axis along *x*; conversely, when the fluorophore was below the average focal plane, the image appeared ellipsoidal with its long axis along *y*. By fitting the image with a 2D elliptical Gaussian function, we obtained the *x* and *y* coordinates of the peak position as well as the peak widths w_x and w_y, which in turn allowed the *z* coordinate of the fluorophore to be unambiguously determined.

To experimentally generate a calibration curve of w_x and w_y as a function of *z*, we immobilized Alexa 647–labeled streptavidin molecules or quantum dots on a glass surface and imaged individual molecules to determine the w_x and w_y values as the sample was scanned in *z* (Fig. 1B). In 3D STORM analysis, the *z* coordinate of each photoactivated fluorophore was then determined by comparing the measured w_x and w_y values of its image with the calibration curves. In addition, for samples immersed in aqueous solution on a glass substrate, all *z* localizations were rescaled by a factor

[1]Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA. [2]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. [3]Department of Physics, Harvard University, Cambridge, MA 02138, USA. [4]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: zhuang@chemistry.harvard.edu

of 0.79 to account for the refractive index mismatch between glass and water [see (*27*) for a detailed description of the analysis procedures].

The 3D resolution of STORM is limited by the accuracy with which individual photoactivated fluorophores can be localized in all three dimensions during a switching cycle. We recently discovered a family of photoswitchable cyanine dyes (Cy5, Cy5.5, Cy7, and Alexa 647) that can be reversibly cycled between a fluorescent and a dark state by light of different wavelengths. The reactivation efficiency of these photoswitchable "reporters" depends critically on the proximity of an "activator" dye, which can be any one of a variety of dye molecules (e.g., Cy3, Cy2, Alexa 405) (*5*, *28*). We used Cy3 and Alexa 647 as the activator and reporter pair to perform 3D STORM imaging. A red laser (657 nm) was used to image Alexa 647 molecules and deactivate them to the dark state; a green laser (532 nm) was used to reactivate Alexa 647

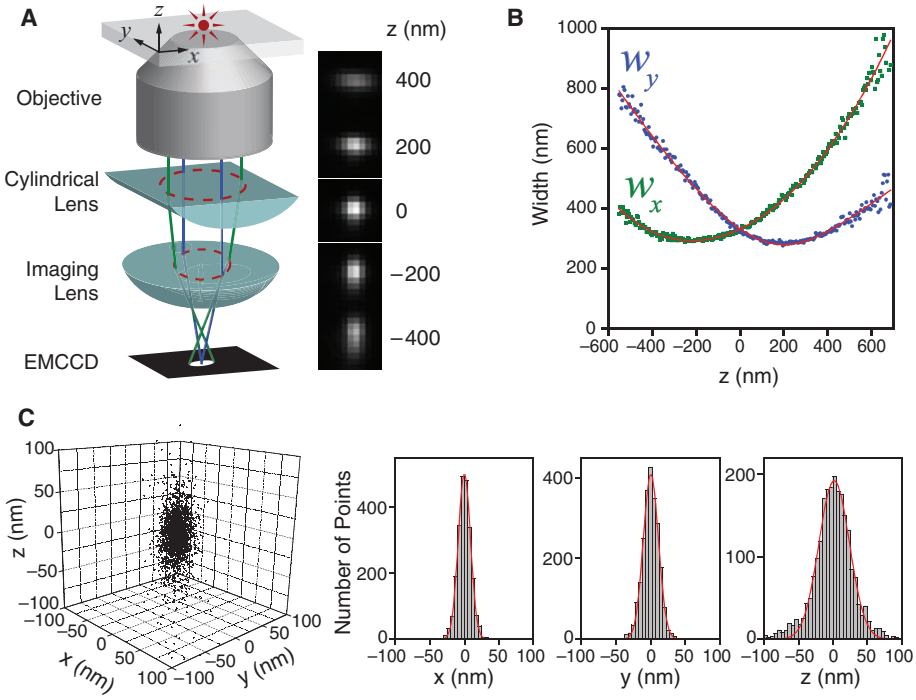


Fig. 1. The scheme of 3D STORM. (**A**) Three-dimensional localization of individual fluorophores. The simplified optical diagram illustrates the principle of determining the *z* coordinate of a fluorescent object from the ellipticity of its image by introducing a cylindrical lens into the imaging path. The right panel shows images of a fluorophore at various *z* positions. EMCCD, electron-multiplying charge-coupled device. (**B**) Calibration curve of image widths w_x and w_y as a function of *z* obtained from single Alexa 647 molecules. Each data point represents the average value obtained from six molecules. The data were fit to a defocusing function (red curve) as described in (*27*). (**C**) Three-dimensional localization distribution of single molecules. Each molecule gives a cluster of localizations due to repetitive activation of the same molecule. Localizations from 145 clusters were aligned by their center of mass to generate the overall 3D presentation of the localization distribution (left panel). Histograms of the distribution in *x*, *y*, and *z* (right panels) were fit to a Gaussian function, yielding standard deviations of 9 nm in *x*, 11 nm in *y*, and 22 nm in *z*.

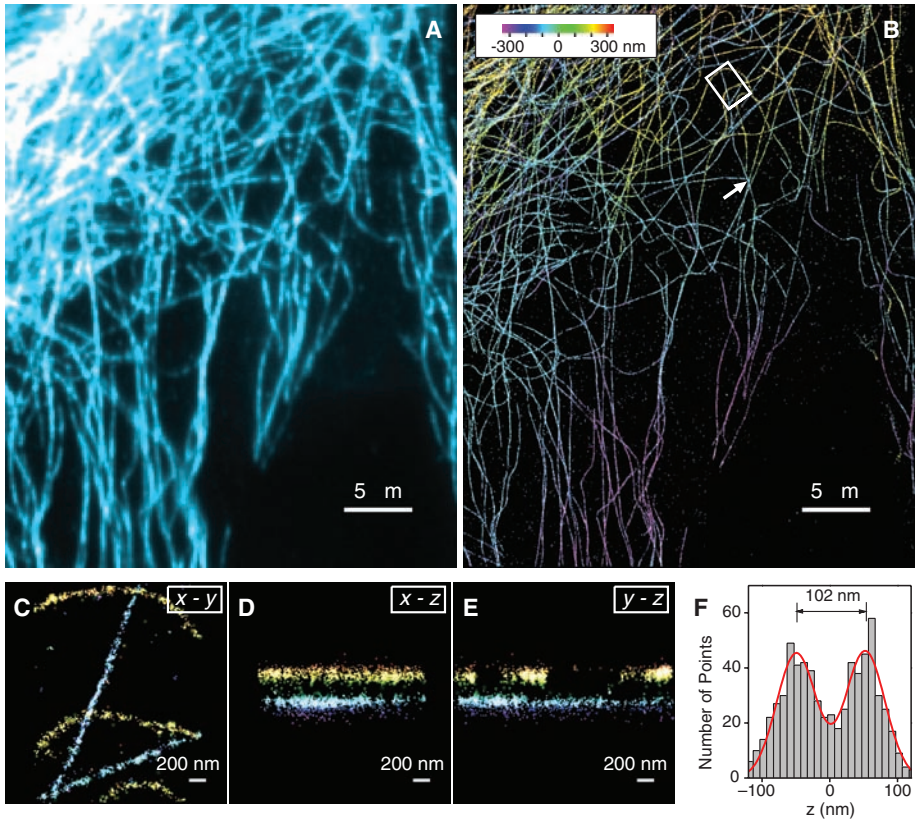


Fig. 2. Three-dimensional STORM imaging of microtubules in a cell. (**A**) Conventional indirect immunofluorescence image of microtubules in a large area of a BS-C-1 cell. (**B**) The 3D STORM image of the same area, with the *z*-position information color-coded according to the color scale bar. Each localization is depicted in the STORM image as a Gaussian peak, the width of which is determined by the number of photons detected (*5*). (**C** to **E**) The *x*-*y*, *x*-*z*, and *y*-*z* cross sections of a small region of the cell outlined by the white box in (B), showing five microtubule filaments. Movie S1 shows the 3D representation of this region, with the viewing angle rotated to show different perspectives (*27*). (**F**) The *z* profile of two microtubules crossing in the *x*-*y* projection but separated by 102 nm in *z*, from a region indicated by the arrow in (B). The histogram shows the distribution of *z* coordinates of the localizations, fit to two Gaussians with identical widths (FWHM = 66 nm) and a separation of 102 nm (red curve). The apparent width of 66 nm agrees quantitatively with the convolution of our imaging resolution in *z* (represented by a Gaussian function with FWHM of 55 nm) and the previously measured width of antibody-coated microtubules (represented by a uniform distribution with a width of 56 nm) (*5*).

in a Cy3-dependent manner (*5*, *28*). Each activator-reporter pair could be cycled on and off hundreds of times before permanent photobleaching occurred. An average of 6000 photons were detected per switching cycle by means of objective-type total internal reflection fluorescence or epifluorescence imaging geometry. This reversible switching behavior provided an internal control to measure the localization accuracy. To this end, we immobilized streptavidin molecules doubly labeled with Cy3 and Alexa 647 on a glass surface (*27*). The molecules were then switched on and off for multiple cycles, and their *x*, *y*, and *z* coordinates were determined for each switching cycle (*27*). This procedure resulted in a cluster of localizations for each molecule (Fig. 1C). The standard deviations of the localization distribution obtained within 100 nm of the average focal plane were 9 nm in *x*, 11 nm in *y*, and 22 nm in *z*, and the corresponding full width at half maximum (FWHM) values were 21 nm, 26 nm, and 52 nm, providing a quantitative measure of the localization accuracy in 3D (Fig. 1C). The localization accuracies in the two lateral dimensions were similar to our previous 2D STORM resolution obtained without the cylindrical lens (*5*). The localization accuracy in *z* was approximately twice the localization accuracy measured in *x* and *y*. Because the image width increases as the fluorophore moves away from the focal plane, the localization accuracy decreases with increasing absolute values of *z*, especially in the lateral dimensions. Therefore, we typically chose a *z* imaging depth of about 600 nm near the focal plane, within which the lateral and axial localization accuracies varied by factors of <1.6 and <1.3, respectively, relative to the values obtained at the average focal plane. The imaging depth may, however, be increased by the use of *z* scanning in future experiments.

As an initial test of 3D STORM, we imaged a model bead sample prepared by immobilizing 200-nm biotinylated polystyrene beads on a glass surface and then incubating the sample with Cy3- and Alexa 647–labeled streptavidin to coat the beads with photoswitchable probes (*27*). Three-dimensional STORM images of the beads were obtained by iterative, stochastic activation of sparse subsets of optically resolvable Alexa 647 molecules, allowing the *x*, *y*, and *z* coordinates of individual molecules to be determined. Over the course of multiple activation cycles, the positions of numerous fluorophores were determined and used to construct a full 3D image (*27*). The projections of the bead images appeared approximately spherical when viewed along all three directions, with average diameters of 210 ± 16, 225 ± 25, and 228 ± 25 nm in *x*, *y*, and *z*, respectively (fig. S1) (*27*), indicating accurate localization in all three dimensions. Because the image of each fluorophore simultaneously encodes its *x*, *y*, and *z* coordinates, no additional time was required to localize each molecule in 3D STORM as compared with 2D STORM imaging.

Applying 3D STORM to cell imaging, we next performed indirect immunofluorescence imaging of the microtubule network in green monkey kidney epithelial (BS-C-1) cells. Cells were immunostained with primary antibodies and then with secondary antibodies doubly labeled with Cy3 and Alexa 647 (*27*). The 3D STORM image not only showed a substantial improvement in resolution over the conventional wide-field fluorescence image (Fig. 2, A and B), but also provided the *z*-dimension information (color-coded in Fig. 2B) that was not available in the conventional image. Multiple layers of microtubule filaments were clearly visible in the

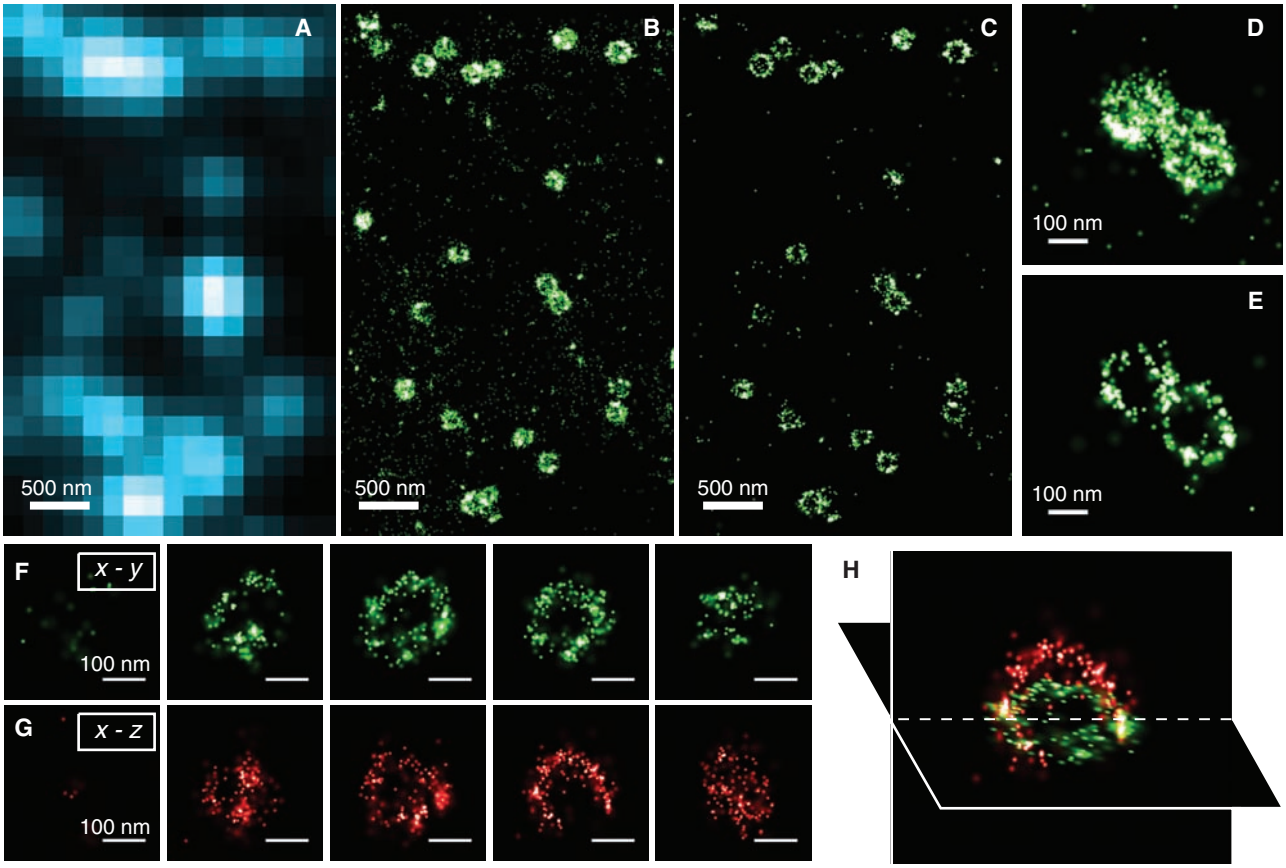


Fig. 3. Three-dimensional STORM imaging of clathrin-coated pits in a cell. (**A**) Conventional direct immunofluorescence image of clathrin in a region of a BS-C-1 cell. (**B**) The 2D STORM image of the same area, with all localizations at different *z* positions included. (**C**) An *x*-*y* cross section (50 nm thick in *z*) of the same area, showing the ring-like structure of the periphery of the CCPs at the plasma membrane. (**D** and **E**) Magnified view of two nearby CCPs in 2D STORM (D) and their *x*-*y* cross section (100 nm thick) in the 3D image (E). (**F** to **H**) Serial *x*-*y* cross sections (each 50 nm thick in *z*) (F) and *x*-*z* cross sections (each 50 nm thick in *y*) (G) of a CCP, and an *x*-*y* and *x*-*z* cross section presented in 3D perspective (H), showing the half-spherical cage-like structure of the pit.

x-y, *x-z*, and *y-z* cross sections of the cell (Fig. 2, C to E, and movie S1) (*27*).

To characterize our cell imaging resolution more quantitatively, we identified point-like objects in the cell that appeared as small clusters of localizations away from any discernible microtubule filaments. These clusters likely represent individual antibodies nonspecifically attached to the cell. The FWHM values of these clusters, which were randomly chosen over the entire measured *z*-range of the cell, were 22 nm in *x*, 28 nm in *y*, and 55 nm in *z* (fig. S2) (*27*), similar to those determined for individual molecules immobilized on a glass surface (compare fig. S2 with Fig. 1C). Two microtubule filaments separated by 100 nm in *z* appeared well separated in the 3D STORM image (Fig. 2F). The apparent width of the microtubule filaments in the *z* dimension was 66 nm, slightly larger than our intrinsic imaging resolution in *z* and in quantitative agreement with the convolution of the imaging resolution and the independently measured width of the antibody-coated microtubule (Fig. 2F). Because the effective resolution is determined by a combination of the intrinsic imaging resolution (as characterized above) and the size of the labels (e.g., antibodies), improved resolution may be achieved by using direct immunofluorescence to remove one layer of antibody labeling, as we show in the next example, or by using Fab fragments or genetically encoded peptide tags (*29*, *30*) in place of antibodies.

Finally, to demonstrate that 3D STORM can resolve the 3D morphology of nanoscopic structures in cells, we imaged clathrin-coated pits (CCPs) in BS-C-1 cells. CCPs are spherical cage-like structures, about 150 to 200 nm in size, assembled from clathrin and cofactors on the cytoplasmic side of the cell membrane to facilitate endocytosis (*31*). To image CCPs, we adopted a direct immunofluorescence scheme using primary antibodies against clathrin doubly labeled with Cy3 and Alexa 647 (*27*). When imaged by conventional fluorescence microscopy, all CCPs appeared as nearly diffraction-limited spots with no discernible structure (Fig. 3A). In 2D STORM images in which the *z*-dimension information was discarded, the round shape of CCPs was clearly seen (Fig. 3, B and D). The size distribution of CCPs measured from the 2D projection image, 180 ± 40 nm, agrees quantitatively with the size distribution determined using electron microscopy (EM) (*32*). Including the *z*-dimension information allowed us to clearly visualize the 3D structure of the pits (Fig. 3, C and E to H). Figures 3C and 3E show the *x-y* cross sections of the image, taken from a region near the opening of the pits at the cell surface. The circular ring-like structure of the pit periphery was unambiguously resolved. Consecutive *x-y* and *x-z* cross sections of the pits (Fig. 3, F to H) clearly revealed the half-spherical cage-like morphology of these nanoscopic structures that was not observable in the 2D images. These experiments demonstrate the ability of 3D STORM to resolve nanoscopic features of cellular structures with molecular specificity under ambient conditions.

References and Notes

1. S. W. Hell, *Nat. Biotechnol.* **21**, 1347 (2003).
2. S. W. Hell, *Science* **316**, 1153 (2007).
3. M. G. L. Gustafsson, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13081 (2005).
4. M. J. Rust, M. Bates, X. Zhuang, *Nat. Methods* **3**, 793 (2006).
5. M. Bates, B. Huang, G. T. Dempsey, X. Zhuang, *Science* **317**, 1749 (2007); published online 15 August 2007 (10.1126/science.1146598).
6. E. Betzig *et al.*, *Science* **313**, 1642 (2006); published online 9 August 2006 (10.1126/science.1127344).
7. S. T. Hess, T. P. K. Girirajan, M. D. Mason, *Biophys. J.* **91**, 4258 (2006).
8. A. Sharonov, R. M. Hochstrasser, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18911 (2006).
9. A. Egner *et al.*, *Biophys. J.* **93**, 3285 (2007).
10. H. Bock *et al.*, *Appl. Phys. B* **88**, 161 (2007).
11. P. Torok, T. Wilson, *Opt. Commun.* **137**, 127 (1997).
12. W. R. Zipfel, R. M. Williams, W. W. Webb, *Nat. Biotechnol.* **21**, 1369 (2003).
13. M. Nagorni, S. W. Hell, *J. Struct. Biol.* **123**, 236 (1998).
14. M. G. L. Gustafsson, D. A. Agard, J. W. Sedat, *J. Microsc.* **195**, 10 (1999).
15. A. Egner, S. W. Hell, *Trends Cell Biol.* **15**, 207 (2005).
16. W. E. Moerner, M. Orrit, *Science* **283**, 1670 (1999).
17. R. E. Thompson, D. R. Larson, W. W. Webb, *Biophys. J.* **82**, 2775 (2002).
18. A. Yildiz *et al.*, *Science* **300**, 2061 (2003); published online 5 June 2003 (10.1126/science.1084398).
19. L. S. Barak, W. W. Webb, *J. Cell Biol.* **90**, 595 (1981).
20. J. Gelles, B. J. Schnapp, M. P. Sheetz, *Nature* **331**, 450 (1988).
21. A. M. van Oijen, J. Kohler, J. Schmidt, M. Muller, G. J. Brakenhoff, *Chem. Phys. Lett.* **292**, 183 (1998).
22. M. Speidel, A. Jonas, E. L. Florin, *Opt. Lett.* **28**, 69 (2003).
23. P. Prabhat, S. Ram, E. S. Ward, R. J. Ober, *Proc. SPIE* **6090**, 60900L (2006).
24. E. Toprak, H. Balci, B. H. Behm, P. R. Selvin, *Nano Lett.* **7**, 2043 (2007).
25. H. P. Kao, A. S. Verkman, *Biophys. J.* **67**, 1291 (1994).
26. L. Holtzer, T. Meckel, T. Schmidt, *Appl. Phys. Lett.* **90**, 053902 (2007).
27. See supporting material on *Science* Online.
28. M. Bates, T. R. Blosser, X. Zhuang, *Phys. Rev. Lett.* **94**, 108101 (2005).
29. I. Chen, A. Y. Ting, *Curr. Opin. Biotechnol.* **16**, 35 (2005).
30. B. N. G. Giepmans, S. R. Adams, M. H. Ellisman, R. Y. Tsien, *Science* **312**, 217 (2006).
31. V. I. Slepnev, P. De Camilli, *Nat. Rev. Neurosci.* **1**, 161 (2000).
32. J. E. Heuser, R. G. W. Anderson, *J. Cell Biol.* **108**, 389 (1989).

33. Supported in part by NIH grant GM 068518. X.Z. is a Howard Hughes Medical Institute Investigator.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1153529/DC1
Materials and Methods
Figs. S1 and S2
Movie S1
References

28 November 2007; accepted 17 December 2007
Published online 3 January 2008;
10.1126/science.1153529
Include this information when citing this paper.

An Association Between the Kinship and Fertility of Human Couples

Agnar Helgason,[1,2]* Snæbjörn Pálsson,[1,3] Daníel F. Guðbjartsson,[1] Þórður Kristjánsson,[1] Kári Stefánsson[1,4]

Previous studies have reported that related human couples tend to produce more children than unrelated couples but have been unable to determine whether this difference is biological or stems from socioeconomic variables. Our results, drawn from all known couples of the Icelandic population born between 1800 and 1965, show a significant positive association between kinship and fertility, with the greatest reproductive success observed for couples related at the level of third and fourth cousins. Owing to the relative socioeconomic homogeneity of Icelanders, and the observation of highly significant differences in the fertility of couples separated by very fine intervals of kinship, we conclude that this association is likely to have a biological basis.

There has been long-standing uncertainty about the impact of kinship or consanguinity between spouses on the total number of offspring they produce (completed fertility). Consanguineous unions among humans increase the probability of a zygote receiving the same deleterious recessive alleles from both parents, with a possible adverse effect on fertility through an increased rate of miscarriage, infant mortality, and morbidity (*1*–*3*). Conversely, consanguineous unions may confer greater completed fertility through earlier age at marriage, as well as the socioeconomic advantages associated with preserving land and wealth within extended families. (*4*, *5*). In other species, lower fitness has been observed in offspring of distantly related individuals, which appears to be a result of the breakdown of coadapted gene complexes (*6*).

Previous studies examining the relationship between kinship and fertility in humans have focused on relatively close relationships between couples, rarely evaluating relationships more distant than second cousins (who share two great-grandparents) (*4*). Such studies have tended to be

[1]deCODE Genetics, Sturlugata 8, 101 Reykjavik, Iceland. [2]Department of Anthropology, University of Iceland, 101 Reykjavik, Iceland. [3]Department of Biology, University of Iceland, 101 Reykjavik, Iceland. [4]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland.

*To whom correspondence should be addressed. E-mail: agnar@decode.is

Table 1. A summary of kinship and fertility in 25-year intervals from 1800 to 1965. Shown are descriptive statistics for kinship coefficients and three variables that reflect the completed fertility (the total number of offspring) and reproductive success (the total number of children who reproduce and the total number of grandchildren) of the couples.

	All couples					Couples with $q_\Phi > 0$	
Birth year of female	*N*	Mean q_Φ (SE)*	Mean number of offspring per couple (SE)	Mean number of offspring that reproduce per couple (SE)	Mean number of grandchildren per couple (SE)	*N*	Mean kinship $\Phi \times 1000$ (25–75 percentiles)
1800–1824	8,673	0.426 (0.0021)	3.610 (0.0359)	1.765 (0.0195)	7.901 (0.1051)	8,362	4.93 (0.004–1.012)
1825–1849	14,338	0.514 (0.0013)	3.468 (0.0254)	1.639 (0.0146)	7.384 (0.0768)	14,109	5.45 (0.029–1.195)
1850–1874	15,863	0.606 (0.0011)	3.221 (0.0234)	1.749 (0.0157)	7.193 (0.0733)	15,575	4.76 (0.054–1.257)
1875–1899	16,691	0.672 (0.0012)	3.392 (0.0231)	2.430 (0.0180)	9.053 (0.0758)	16,268	3.70 (0.043–1.050)
1900–1924	24,732	0.721 (0.0011)	2.791 (0.0143)	2.360 (0.0127)	7.467 (0.0450)	23,799	2.01 (0.024–0.562)
1925–1949	39,635	0.759 (0.0010)	2.547 (0.0087)	1.996 (0.0083)	4.983 (0.0237)	37,762	0.82 (0.022–0.336)
1950–1965	40,879	0.782 (0.0012)	2.004 (0.0058)	0.501 (0.0038)	0.864 (0.0075)	38,336	0.50 (0.033–0.306)
Total	160,811	0.695 (0.0005)	2.740 (0.0058)	1.648 (0.0046)	5.330 (0.0182)	154,211	2.22 (0.029–0.526)

*q_Φ is a weighted measure of the genealogical information available to calculate the kinship of couples, with values between 0 (when at least one spouse has no known ancestor) and 1 (when all ancestors are known for both spouses). See (*10*) for more details.

performed in populations with relatively high rates of consanguineous marriages, such as those of India, Pakistan, and the Middle East (*4*, *7*–*9*); however, these populations also tend to be characterized by large socioeconomic disparities.

To explore the relationship between fertility and kinship in humans, we examined 160,811 Icelandic couples from the deCODE Genetics genealogical database born between 1800 and 1965 (*10*). The advantage of using the Icelandic data set lies in this population being small and one of the most socioeconomically and culturally homogeneous societies in the world (*11*), with little variation in family size, use of contraceptives, and marriage practices (*12*), in contrast with most previously studied populations (*4*, *7*–*9*). By estimating kinship based on a depth of up to 10 generations from each couple, we were able to assess differences in fertility across a fine scale of kinship values. Our data indicated that there has been a decrease by a factor of 10 in mean kinship between Icelandic couples during the past two centuries, from 0.005 for couples with females born 1800 to 1824 to 0.0005 for those born 1950 to 1965 (Table 1). This is equivalent to a change from couples being related on average between the level of third and fourth cousins to couples being related on average at the level of fifth cousins. The primary cause is probably a demographic transition from a poor agricultural society to an affluent industrial society, involving extensive migration from rural regions to urban centers, accompanied by a rapid expansion in population size (*13*). The outcome of this transition is an expansion of the pool of potential mates for contemporary Icelanders, particularly those who are distantly related. Typically, this kind of demographic transition results in a drop in the average number of children per couple with time (Table 1). However, this relationship is not monotonic for the Icelandic data (fig. S1). To compare the kinship and fertility of couples born between 1800 and 1965, we standardized the variables documenting kinship, the number of children per couple, and other measures of reproductive success (*10*).

A monotonic positive relationship was observed between the degree of kinship among spouses and the number of children they produced (Fig. 1A). Furthermore, the reproductive success of the couples, as reflected by the number of their children who reproduced (Fig. 1B), followed an n-shaped curve from the relatively low reproductive success of couples related at the level of second cousins or closer, to the maximum for couples related at the level of third and fourth cousins, after which there is a steady decrease in reproductive success with diminishing kinship between spouses. A similar picture emerges when the number of grandchildren per couple is examined (Fig. 1C).

These results are based on couples born during a period of almost 200 years, in the course of which there was a marked decline both in the mean fertility and in kinship between couples (Table 1). Nonetheless, the same general relationship between kinship and reproductive outcome was observed within each 25-year subinterval (fig. S2). We evaluated the correlation between the standardized variables of kinship and reproductive outcome for all couples and for each time interval separately (Table 2), adjusting for the impact of geographical differences in the kinship and fertility of couples within Iceland (*10*). Each test revealed a significant association with kinship, with correlation coefficients of 0.063 ($P = 1.5 \times 10^{-129}$) for the number of children, 0.045 ($P = 3.6 \times 10^{-66}$) for the number of children who reproduced, and 0.042 ($P = 7.6 \times 10^{-58}$) for the number of grandchildren. To assess the potential impact of q_Φ (the amount of information available to calculate the kinship coefficient, Φ, for each couple) on the key variables of kinship and reproductive outcome, we also performed the correlation analyses for the subset of 112,683 couples for whom all ancestors are known four generations back in time (Table 2). Almost identical results were obtained for couples born after 1850. For couples born before 1850, the association with fertility was statistically significant, but not with the two indicators of reproductive success (i.e., children and grandchildren), primarily because so many couples with incompletely known ancestral genealogies had to be omitted from the analysis.

Although the general pattern is one of both greater fertility and reproductive success with increasing kinship between spouses, there was a notable deficit in the reproductive success of couples related at the level of second cousins or closer (Fig. 1, B and C). Figure 1D shows that this deficit was partly accounted for by a shorter average life span of children produced by such couples (see also fig. S3). However, because there was still a strong monotonic relationship between kinship and fertility of couples when we restricted analysis to the number of children who survived to the age of 30 years, the lower reproductive success of the most related couples may also stem from greater morbidity or mortality of their offspring during adulthood (fig. S4). We do not find evidence for a sex difference in such reproductive costs among offspring (fig. S5).

Although Icelanders have experienced a socioeconomic transformation from 1800 to the present (*14*, *15*), accompanied by a reduction in family size and decreasing kinship between couples (Table 1), essentially the same relationship between kinship and fertility was observed at the beginning and end of this 200-year period (fig. S2). By estimating kinship between spouses at a genealogical depth of up to 10 generations, it was possible to examine the association with fertility and reproductive success at a very fine scale. Thus, for example, there is a statistically significant difference in the number of children produced by couples related at the level of sixth versus seventh cousins ($P = 1.4 \times 10^{-7}$). Relationships at this genealogical distance are rarely known to the couples or their families and acquaintances in their social environment and are unlikely to influence factors such as age at the commencement of reproduction or the practice of consanguineous unions to preserve family property (*4*, *16*).

Although some interaction of fertility and kinship with socioeconomic factors cannot be

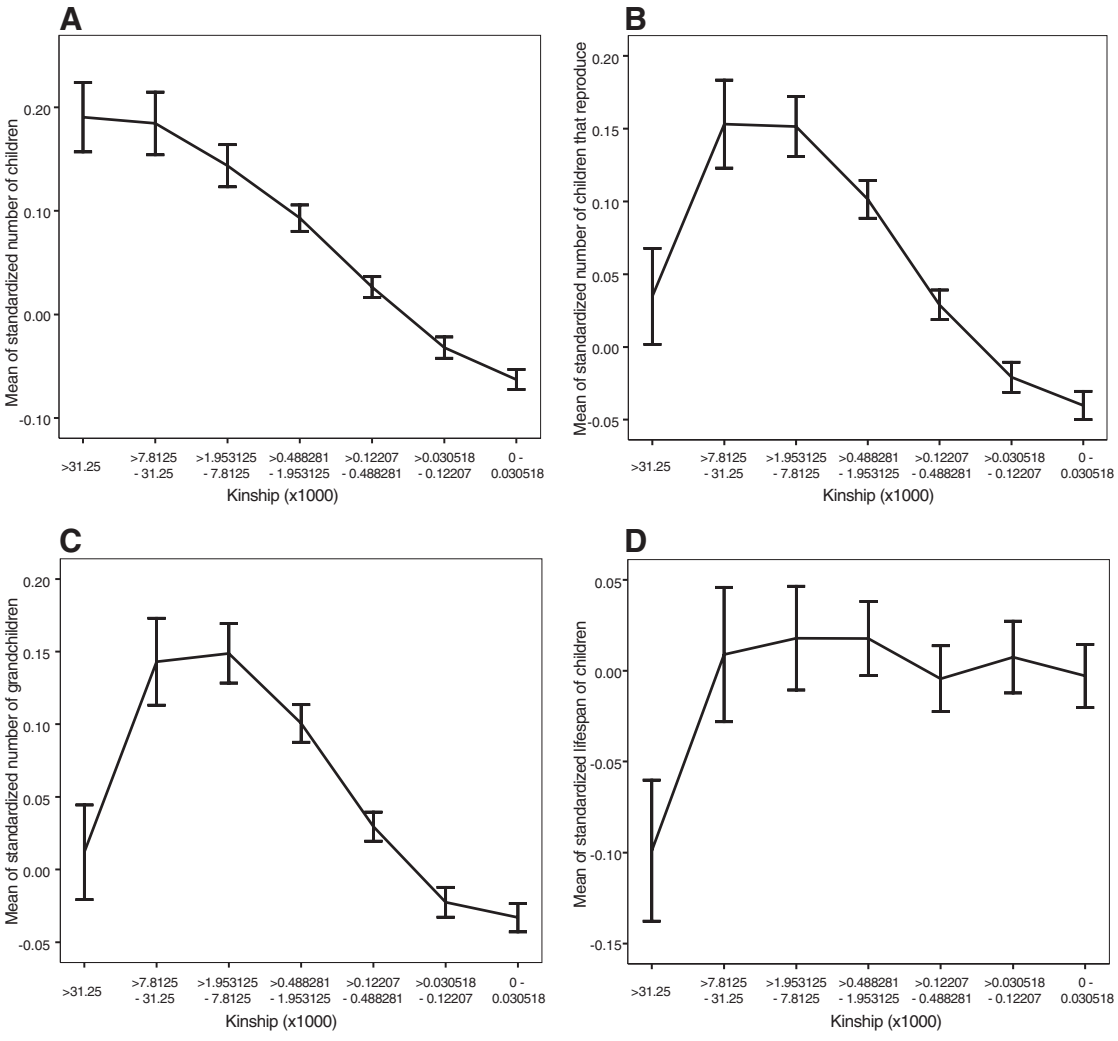


Fig. 1. The relationship between kinship and reproduction among Icelandic couples. The four panels show means and 95% confidence intervals of standardized variables relating to the reproductive outcome of Icelandic couples as a function of seven intervals of kinship. (**A**) shows the total number of children, (**B**) the number of children who reproduced, (**C**) the number of grandchildren, and (**D**) the mean life expectancy of children. The first interval of kinship represents all couples related at the level of second cousins or closer, the second interval represents couples related at the level of third cousins and up to the level of second cousins, and so on, with each subsequent category representing steps to fourth, fifth, sixth, and seventh cousins and the final category representing couples with no known relationship and those with relationships up to the level of eighth cousins.

Table 2. The correlation between kinship and reproductive outcome.

	All couples				Couples with all ancestors known four generations back			
Birth year of female	*N*	Number of children Pearson's *r* (*P* value)	Children who reproduce Pearson's *r* (*P* value)	Grandchildren Pearson's *r* (*P* value)	*N*	Number of children Pearson's *r* (*P* value)	Children who reproduce Pearson's *r* (*P* value)	Grandchildren Pearson's *r* (*P* value)
1800–1824	8,346	0.071 (4.7×10^{-10})	0.054 (1.6×10^{-6})	0.057 (3.8×10^{-7})	1,401	0.130 (5.1×10^{-3})	0.022 (6.2×10^{-1})	0.022 (6.3×10^{-1})
1825–1849	14,050	0.085 (8.8×10^{-23})	0.042 (1.5×10^{-6})	0.037 (1.8×10^{-5})	4,783	0.088 (1.4×10^{-5})	0.027 (1.7×10^{-1})	0.013 (5.2×10^{-1})
1850–1874	15,440	0.088 (2.2×10^{-26})	0.053 (1.2×10^{-10})	0.045 (5.4×10^{-8})	10,568	0.097 (2.4×10^{-18})	0.046 (3.1×10^{-5})	0.037 (7.9×10^{-4})
1875–1899	16,150	0.080 (7.2×10^{-23})	0.053 (4.4×10^{-11})	0.047 (7.0×10^{-9})	13,563	0.080 (3.4×10^{-18})	0.049 (8.4×10^{-8})	0.042 (3.5×10^{-6})
1900–1924	23,740	0.072 (2.3×10^{-27})	0.067 (4.9×10^{-24})	0.065 (6.0×10^{-23})	21,022	0.076 (1.5×10^{-25})	0.068 (6.0×10^{-21})	0.067 (2.1×10^{-20})
1925–1949	36,733	0.056 (4.0×10^{-27})	0.049 (4.1×10^{-21})	0.047 (2.1×10^{-19})	31,510	0.059 (2.8×10^{-23})	0.051 (4.4×10^{-18})	0.05 (1.1×10^{-17})
1950–1965	36,510	0.034 (6.4×10^{-11})	0.020 (1.5×10^{-4})	0.018 (6.5×10^{-4})	29,836	0.034 (1.6×10^{-8})	0.027 (2.3×10^{-5})	0.025 (9.8×10^{-5})
All	150,969	0.063 (1.5×10^{-129})	0.045 (3.6×10^{-66})	0.042 (7.6×10^{-58})	112,683	0.063 (2.1×10^{-86})	0.046 (2.8×10^{-46})	0.043 (6.1×10^{-41})

ruled out, our results support the hypothesis that the positive association between kinship and fertility has a basis in reproductive biology. A positive relationship between kinship and reproductive success seems counterintuitive from an evolutionary perspective. We did find some evidence of a reproductive cost borne by offspring of parents related at the degree of second cousins or closer. Strikingly, however, our results show that couples related at the degree of third to fourth cousins exhibited the greatest reproductive success.

The formation of densely populated urban regions that offer a large selection of distantly related potential spouses is a new situation for humans in evolutionary terms. We note that if the relationship between kinship and fertility has a basis in human reproductive biology, then it follows that the kind of demographic transition recently experienced by the Icelandic population could directly contribute to the slowing of population growth elsewhere through the relative increase of distantly related couples.

References and Notes

1. C. Ober, T. Hyslop, W. W. Hauck, *Am. J. Hum. Genet.* **64**, 225 (1999).
2. W. J. Schull, J. V. Neel, *The Effects of Inbreeding on Japanese Children* (Harper and Row, New York, 1965).

3. A. Bittles, *Clin. Genet.* **60**, 89 (2001).
4. A. H. Bittles, J. C. Grant, S. G. Sullivan, R. Hussain, *Ann. Hum. Biol.* **29**, 111 (2002).
5. P. Philippe, *Hum. Biol.* **46**, 405 (1974).
6. S. Edmands, *Mol. Ecol.* **16**, 463 (2007).
7. M. al Husain, M. al Bunyan, *Ann. Trop. Paediatr.* **17**, 155 (1997).
8. A. H. Bittles, J. C. Grant, S. A. Shami, *Int. J. Epidemiol.* **22**, 463 (1993).
9. R. Hussain, A. H. Bittles, *J. Health Popul. Nutr.* **22**, 1 (2004).
10. Materials and methods are available as supporting material on *Science* Online.
11. K. Watkins *et al*., *United Nations Human Development Report. Beyond Scarcity: Power, Poverty and the Global Water Crisis* (Palgrave Macmillan, New York, 2006).
12. G. B. Eydal, S. Olafsson, "Demographic trends in Iceland. First report for the project Welfare Policy and Employment in the Context of Family Change" (2003); www.york.ac.uk/inst/spru/research/summs/welempfc.htm.
13. A. Helgason, B. Yngvadottir, B. Hrafnkelsson, J. Gulcher, K. Stefansson, *Nat. Genet.* **37**, 90 (2005).
14. G. A. Gunnlaugsson, L. Guttormsson, *J. Fam. Hist.* **18**, 315 (1993).
15. G. A. Gunnlaugsson, *Saga og samfélag: Þættir úr félagssögu 19. og 20. aldar* (Sagnfræðistofnun Háskóla Íslands, Reykjavik, 1997).
16. M. J. Blanco Villegas, V. Fuster, *Ann. Hum. Biol.* **33**, 330 (2006).
17. We thank A. Kong for constructive comments and suggestions. A table available in the supporting online material contains the key variables for each couple that were used in our analyses. Requests for access to more detailed data than those presented in the table should be referred to A.H. (agnar@decode.is) or K.S. (kstefans@decode.is). Owing to the sensitive nature of the underlying genealogies, access to more detailed data can only be granted at the headquarters of deCODE Genetics in Iceland.

Supporting Online Material
www.sciencemag.org/cgi/content/full/319/5864/813/DC1
Materials and Methods
Figs. S1 to S5
References

7 September 2007; accepted 14 January 2008
10.1126/science.1150232

Mutations in the Pericentrin (*PCNT*) Gene Cause Primordial Dwarfism

Anita Rauch,[1]* Christian T. Thiel,[1] Detlev Schindler,[2] Ursula Wick,[1] Yanick J. Crow,[3] Arif B. Ekici,[1] Anthonie J. van Essen,[4] Timm O. Goecke,[5] Lihadh Al-Gazali,[6] Krystyna H. Chrzanowska,[7] Christiane Zweier,[1] Han G. Brunner,[8] Kristin Becker,[9] Cynthia J. Curry,[10] Bruno Dallapiccola,[11] Koenraad Devriendt,[12] Arnd Dörfler,[13] Esther Kinning,[14] André Megarbane,[15] Peter Meinecke,[16] Robert K. Semple,[17] Stephanie Spranger,[18] Annick Toutain,[19] Richard C. Trembath,[20] Egbert Voss,[21] Louise Wilson,[22] Raoul Hennekam,[22,23,24] Francis de Zegher,[25] Helmuth-Günther Dörr,[26] André Reis[1]

Fundamental processes influencing human growth can be revealed by studying extreme short stature. Using genetic linkage analysis, we find that biallelic loss-of-function mutations in the centrosomal pericentrin (*PCNT*) gene on chromosome 21q22.3 cause microcephalic osteodysplastic primordial dwarfism type II (MOPD II) in 25 patients. Adults with this rare inherited condition have an average height of 100 centimeters and a brain size comparable to that of a 3-month-old baby, but are of near-normal intelligence. Absence of PCNT results in disorganized mitotic spindles and missegregation of chromosomes. Mutations in related genes are known to cause primary microcephaly (*MCPH1*, *CDK5RAP2*, *ASPM*, and *CENPJ*).

The growth of an individual depends on regulation of cell size and cell division. Dysfunction of these regulatory pathways not only results in somatic undergrowth but contributes to a wide variety of pathological conditions, including cancer and diabetes (*1*). To identify potential regulators of human growth, we used positional cloning to determine the underlying defect in a rare autosomal recessive disorder characterized by extreme pre- and postnatal growth retardation, namely, microcephalic osteodysplastic primordial dwarfism type Majewski II [MOPD II, Mendelian Inheritance in Man (MIM) 210720].

Individuals with MOPD II have an average birth weight of less than 1500 g at term, an adult height of about 100 cm, and a variety of associated bone and dental anomalies (Fig. 1) (*2*, *3*). Despite the small head size (average postpubertal head circumference of 40 cm), brain development appears grossly normal with only a few individuals displaying serious mental retardation, a feature that sets MOPD II apart from primary microcephaly and Seckel syndrome. Far-sightedness, irregular pigmentation, truncal obesity, and type 2 diabetes with onset at or before puberty have been noted in older individuals with MOPD II, and life expectancy is reduced because of a high risk of stroke secondary to cerebral vascular anomalies, often classified as Moyamoya disease (*2*, *4*). Although these features led investigators to hypothesize that MOPD II is a premature aging syndrome (*5*), we found no evidence of accelerated telomeric shortening as a potential cellular explanation of premature aging in lymphocyte samples of two unrelated female patients with MOPD II (P1 and P2) (fig. S1) (*6*). MOPD II patients do not show an enhanced predisposition to cancer; consistent with this, patient lymphocytes did not show an increased frequency of sister chromatid exchange (table S1), as would be indicative of a defect in DNA repair, and typical of another syndrome associated with significant short stature, namely, Bloom syndrome (MIM 210900).

Consanguinity in the respective parents of the two unrelated female patients P1 and P2 presented the possibility of locating a MOPD II locus by homozygosity mapping (*6*, *7*) (Fig. 2A). This approach allows the identification of an autosomal recessive disease locus by tracking its segregation within a common chromosomal segment that originates from a shared recent ancestor and is transmitted through both parents. Genome-wide linkage analysis using polymorphic short tandem repeat markers revealed a single disease locus on chromosome 21q22.3. When a third consanguineous family was included, a maximum lod (logarithm of the odds ratio for linkage) score of 3.7 was obtained at marker *D21S1446* (Fig. 2 and fig. S2), confirming linkage to this locus. The linked region encompasses 4.6 megabases at the distal end of chromosome 21 and contains the pericentrin (*PCNT*) gene, which we considered a suitable candidate gene because of its postulated role in chromosome segregation. Mutational analysis of the 47 exons of *PCNT* in 25 unrelated patients with a clinical diagnosis of MOPD II, including those from the three linked families, revealed homozygous and compound heterozygous

[1]Institute of Human Genetics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany. [2]Department of Human Genetics, University of Würzburg, Würzburg, Germany. [3]Leeds Institute of Molecular Medicine, St. James's University Hospital, Leeds, UK. [4]Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, Netherlands. [5]Institut für Humangenetik und Anthropologie, Heinrich-Heine-Universität, Düsseldorf, Germany. [6]Faculty of Medicine, United Arab Emirates University, Al-Ain, UAE. [7]Department of Medical Genetics, the Children's Memorial Health Institute, Warsaw, Poland. [8]Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Radboud University, Nijmegen Medical Centre, Netherlands. [9]North Wales Clinical Genetics Service, Glan Clwyd Hospital, Rhyl, and Institute of Medical Genetics, University Hospital of Wales, Cardiff, UK. [10]Genetic Medicine Central California, Fresno, and University of California, San Francisco, CA, USA. [11]IRCCS-CSS, San Giovanni Rotondo and CSS-Mendel Institute, Rome, and Department of Experimental Medicine and Pathology, University of Rome La Sapienza, Rome, Italy. [12]Centre for Human-Genetics, University of Leuven, Leuven, Belgium. [13]Department of Neuroradiology, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany. [14]Department of Clinical Genetics, Leicester Royal Infirmary, Leicester, UK. [15]Unité de Génétique Médicale, Faculté de Médecine, Université Saint-Joseph, Beirut, Lebanon. [16]Abteilung für Medizinische Genetik, Altonaer Kinderkrankenhaus, Hamburg, Germany. [17]Department of Clinical Biochemistry, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK. [18]Praxis fuer Humangenetik, Bremen, Germany. [19]Department of Genetics, Bretonneau University Hospital, Tours, France. [20]Department of Medical and Molecular Genetics, School of Medicine, King's College London, UK. [21]Cnopf's Pediatric Hospital, Nuremberg, Germany. [22]Department of Clinical Genetics, Great Ormond Street Hospital for Children, London, UK. [23]Clinical and Molecular Genetics Unit, Institute of Child Health, University College London, London, UK. [24]Department of Paediatrics, University of Amsterdam, Amsterdam, Netherlands. [25]Department of Woman and Child, University of Leuven, Leuven, Belgium. [26]Department of Pediatrics and Adolescent Medicine, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany.

*To whom correspondence should be addressed. E-mail: Anita.Rauch@humgenet.uni-erlangen.de

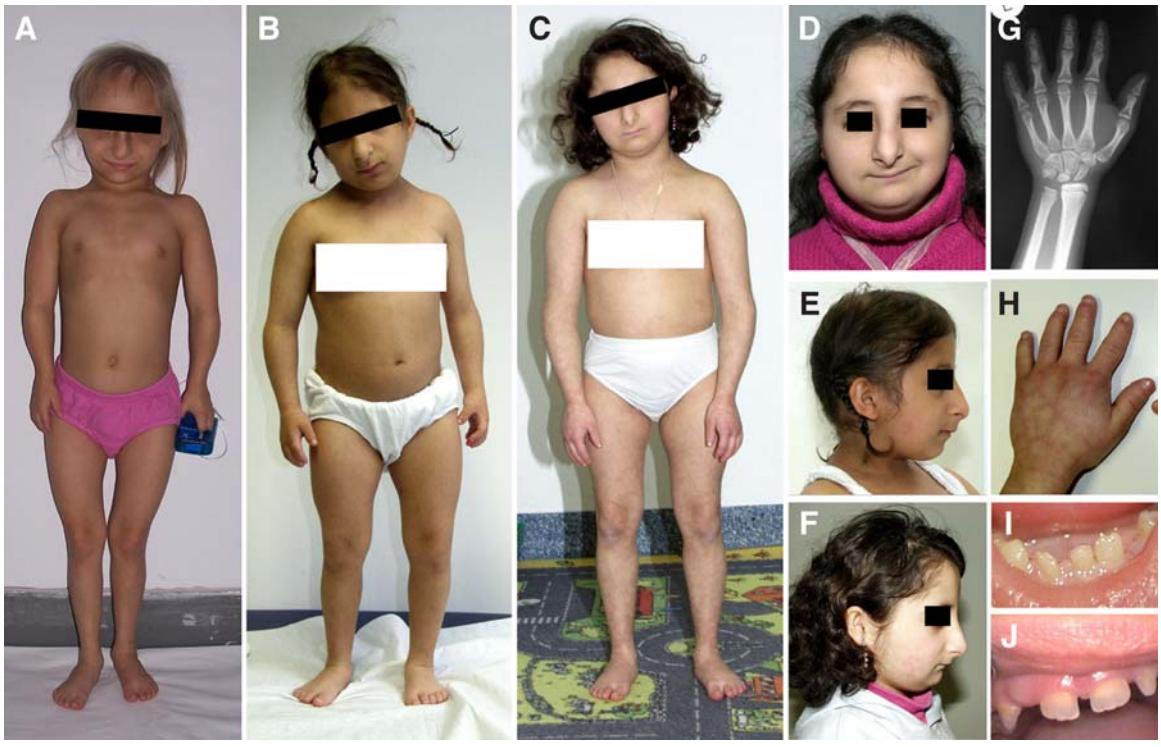
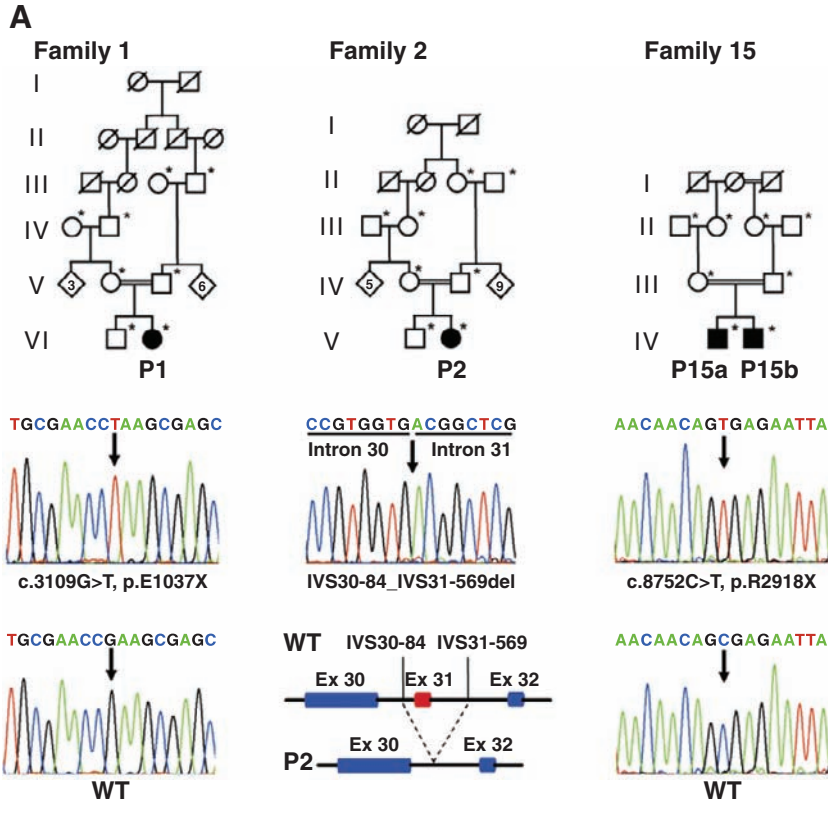


Fig. 1. Phenotype of MOPD II patients. (**A**) P18 at age 8 years 3 months with a height of 84 cm corresponding to a normal size for a female infant aged 1 year 3 months; (**B** and **E**) P1 at age 8 years 8 months with a height of 85 cm; (**C** and **F**) P2 at age 12 years 6 months with a height of 95 cm and at age 14 years with a height of 96 cm (**D**) corresponding to a normal size for a female aged 3 years. Note short lower arms especially in P18, mild truncal obesity and premature puberty in P1, significant facial asymmetry in P2 (D), and absence of a sloping forehead typical of microcephaly syndromes. All three patients demonstrate a long nose with prominent tip and hypoplastic alae and small mandible described as typical for patients with MOPD II. (**G** and **H**) X-ray and an image of the dorsum of the left hand of patient P2 showing generalized brachydactyly with diaphyseal constriction (overmodeling) of metacarpals and phalanges, as well as abnormal flat shape of the distal radius and ulna epiphyses. (**I** and **J**) Hypoplasia and partial agenesis of teeth from patient P2, enamel hypoplasia in teeth from patient P18.



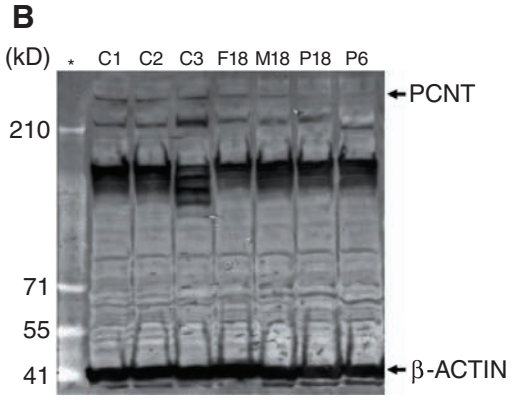


Fig. 2. Pedigrees used for linkage analysis and the respective homozygous mutations identified in *PCNT*. (**A**) Families 1 and 2 were used for the whole-genome scan; families 1, 2, and 15 were used for fine mapping. Individuals marked with asterisks were included in the linkage analysis. (**B**) Western blot analysis of lymphoblastoid cell lines from MOPD II patients P18 and P6, the parents of P18 (F18, father, and M18, mother), and normal controls (C1 to C3). Note the undetectable PCNT (370 kD) in P18 and P6 and reduction of protein level in both parents. *SeeBluePlus2 Prestained standard (Invitrogen, Carlsbad, CA).

null mutations distributed throughout the gene in all patients (table S2 and Fig. 2A). We observed a total of 29 different mutations consisting of 12 stop mutations and 17 frameshift mutations (4 splice-site mutations, 2 small insertions, 10 small deletions, 1 exon deletion). Two mutations occurred twice in unrelated patients, namely, R1923X, in patients P3 and P4, and c.841_842insG in patients P11 and P13. R1923X occurred independently twice, as the respective *PCNT* haplotypes differed in a total of 24 single-nucleotide polymorphisms, whereas c.841_842insG appears to have been transmitted through an unknown common ancestor in patients P11 and P13 (both of Turkish origin), because these patients were identical for all polymorphic sites identified within the *PCNT* genomic region. In contrast, 17 patients with a clinical diagnosis of MOPD I or III, Seckel syndrome, or unclassified growth retardation syndromes showed no *PCNT* mutations. Absence of the PCNT protein (also known as kendrin, PCNT2, or PCNTB) was confirmed by Western blot analysis of lymphoblastoid cell lines from patients P18 and P6 (Fig. 2B). It is noteworthy that both investigated heterozygous parents of patient 18 showed reduced protein levels in lymphoblasts. This might explain our finding of significant reduction of the mean height of heterozygous MOPD II parents (table S3). *PCNT* is apparently not sensitive to gene dosage alterations, because mRNA levels were normal in patients with either monosomy or trisomy of the *PCNT* locus (fig. S3A). MOPD II patients showed either normal or variably diminished mRNA levels (fig. S3B), most likely due to varying degrees of nonsense-mediated mRNA decay resulting from pretranslational mRNA surveillance mechanisms (*8*). Our findings thus characterize MOPD II as a distinct clinical entity caused by biallelic loss-of-function mutations in *PCNT*. Given that all *PCNT* mutations observed in MOPD II patients are mutations leading to a loss of functional protein, it remains to be determined whether *PCNT* missense variants are associated with incomplete or distinct phenotypes.

PCNT is a giant coiled-coil protein (~370 kD) that localizes specifically to centrosomes throughout the cell cycle (*9*). The centrosome is a cell component that organizes cytoplasmic organelles and primary cilia in interphase cells, and mitotic spindle microtubules to ensure proper chromosome segregation during cell division (*10*). PCNT and AKAP9 (A kinase anchor protein 9; formerly known as CG-NAP) share a highly related C-terminal calmodulin-binding domain and mediate, in a noncompensating manner, nucleation of microtubules by anchoring the γ-tubulin ring complex, which initiates the assembly of the mitotic spindle apparatus (*9*, *11*, *12*). Pericentrin and AKAP9 are orthologs of the yeast Spc110 protein, whose absence causes defective spindle formation and results in a lethal failure to segregate chromosomes in the budding yeast (*13*, *14*). Programmed cell death (apoptosis) after activation of mitotic checkpoints and arrest of cells in G_2 phase–to-mitosis transition was shown in some, but not all, vertebrate cell lines depleted of PCNT by small interfering RNA (*12*). It is likely that pericentrin-depleted human cells are more susceptible to death because of defective mitosis and chromosome segregation. This would result in a decrease in total cellularity of the embryo and growth restriction in the adult. In accord with this hypothesis, we observed abnormal mitotic morphology in 71% of MOPD II fibroblast cells (Fig. 3), together with low-level mosaic variegated aneuploidy (MVA) and premature sister chromatid separation (PCS) (fig. S4 and table S4). As suggested for the centrosome in general, our findings would indicate an additional role of PCNT in the spindle assembly checkpoint, in the absence of which cells do not arrest in metaphase but prematurely separate sister chromatids and then exit mitosis (*15*). PCS and MVA, at higher rates than we observed in MOPD II cells, are characteristic of individuals with MVA syndrome (MIM 257300) characterized by cancer susceptibility, growth retardation of intrauterine onset, and microcephaly because of homozygous mutations in the gene encoding BUBR1, a protein, which is known to be involved in the mitotic spindle checkpoint and the initiation of apoptosis in polyploidy cells (*16*).

Although the precise pathogenic mechanisms involved remain unclear, it is noteworthy that mutations in centrosomal and mitotic spindle–related genes have now been identified in three forms of primary microcephaly (*CDK5RAP2*: MCPH3, MIM 604804; *ASPM*: MCPH5, MIM 608716; *CENPJ*: MCPH6, MIM 608393). In addition, biallelic mutations in *MCPH1*, which functions in the regulation of chromosome condensation, have been reported in primary microcephaly with mental retardation and short stature (MIM 606858).

There is an ongoing debate as to whether the Late Pleistocene hominid fossils from the island of Flores, Indonesia, represent a diminutive, small-brained new species, *Homo floresiensis*, or pathological modern humans (*17*–*28*). We note that individuals with MOPD II have several features in common with *Homo floresiensis,* including an adult height of 100 cm, grossly normal intelligence despite severely restricted brain size, absence of a sloping microcephalic morphology, and a number of minor morphological features including facial asymmetry, small chin, abnormal teeth, and subtle

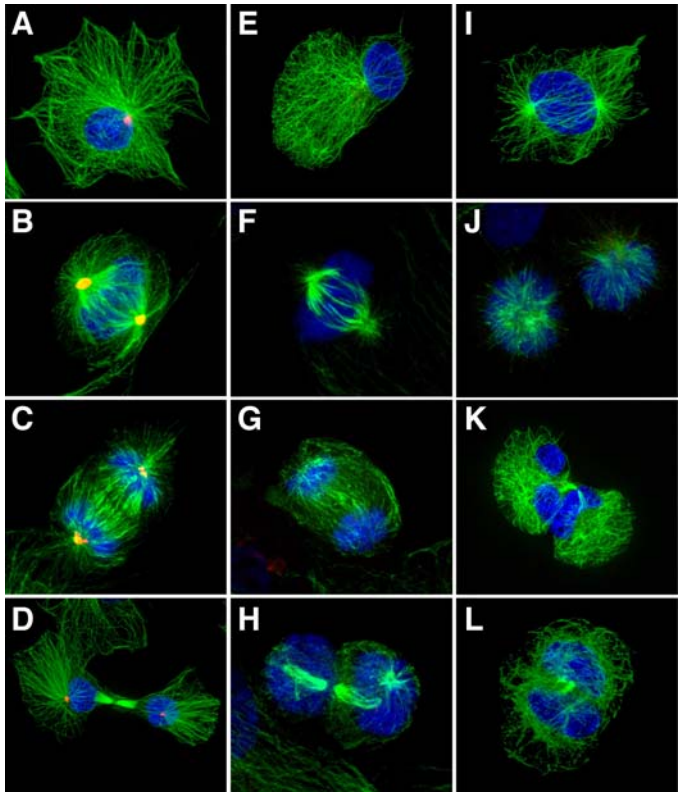


Fig. 3. Abnormal mitotic morphology of patient fibroblasts. Immunofluorescence images of fibroblast cells with antibodies against PCNT (red) and α-tubulin (green), and 4′,6′-diamidino-2-phenylindole (DAPI) staining of chromosomes (blue). (**A** to **D**) Representative morphology of fibroblasts from a healthy individual during (A) interphase, (B) metaphase, (C) anaphase, and (D) cytokinesis. (**E** to **L**) Undetectable PCNT in fibroblasts from the MOPD II patient P1 in interphase (E) and during mitosis [(F) to (L)] as well as representative examples of abnormal morphology with disorganized mitotic microtubules during prometaphase (I), metaphase [(F) and (J)] and anaphase (G); incorrect vertical orientation of metaphases (J); and disorganized cytokinesis [(H), (K), and (L)] with abnormal nuclei pattern (K). Clearly abnormal spindle pattern was observed in 71% of mitotic fibroblasts from the MOPD II patient (n = 100; control 9%, n = 100; $P < 3 \times 10^{-20}$; Fisher's exact test).

bony anomalies of the hand and wrist. Given these similarities, it is tempting to hypothesize that the Indonesian diminutive hominids were in fact humans with MOPD II. With the identification of the genetic basis of MOPD II, this hypothesis may soon be testable.

References and Notes

1. M. N. Hall, M. Raff, G. Thomas, Eds., *Cell Growth: Control of Cell Size* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004).
2. J. G. Hall, C. Flora, C. I. Scott Jr., R. M. Pauli, K. I. Tanaka, *Am. J. Med. Genet. A.* **130**, 55 (2004).
3. F. Majewski, M. Ranke, A. Schinzel, *Am. J. Med. Genet.* **12**, 23 (1982).
4. F. Brancati, M. Castori, R. Mingarelli, B. Dallapiccola, *Am. J. Med. Genet. A.* **139**, 212 (2005).
5. J. G. Hall, *Am. J. Med. Genet. A.* **140**, 1356 (2006).
6. Materials and methods are available as supporting material on *Science* Online.
7. E. S. Lander, D. Botstein, *Science* **236**, 1567 (1987).
8. O. Isken, L. E. Maquat, *Genes Dev.* **21**, 1833 (2007).
9. M. R. Flory, M. J. Moser, R. J. Monnat Jr., T. N. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5919 (2000).
10. S. Doxsey, W. Zimmerman, K. Mikule, *Trends Cell Biol.* **15**, 303 (2005).
11. M. Takahashi, A. Yamagiwa, T. Nishimura, H. Mukai, Y. Ono, *Mol. Biol. Cell* **13**, 3235 (2002).
12. W. C. Zimmerman, J. Sillibourne, J. Rosa, S. J. Doxsey, *Mol. Biol. Cell* **15**, 3642 (2004).
13. D. A. Stirling, T. F. Rayner, A. R. Prescott, M. J. Stark, *J. Cell Sci.* **109**, 1297 (1996).
14. D. A. Stirling, M. J. Stark, *Biochim. Biophys. Acta* **1499**, 85 (2000).
15. H. Muller, M. L. Fogeron, V. Lehmann, H. Lehrach, B. M. Lange, *Science* **314**, 654 (2006).
16. S. Hanks *et al.*, *Nat. Genet.* **36**, 1159 (2004).
17. D. Argue, D. Donlon, C. Groves, R. Wright, *J. Hum. Evol.* **51**, 360 (2006).
18. P. Brown *et al.*, *Nature* **431**, 1055 (2004).
19. D. Falk *et al.*, *Science* **308**, 242 (2005).
20. D. Falk *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2513 (2007).
21. R. D. Martin, A. M. Maclarnon, J. L. Phillips, W. B. Dobyns, *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* **288**, 1123 (2006).
22. R. D. Martin *et al.*, *Science* **312**, 999 (2006); www.sciencemag.org/cgi/content/full/312/5776/999b.
23. D. Falk *et al.*, *Science* **312**, 999 (2006); www.sciencemag.org/cgi/content/full/312/5776/999c.
24. G. D. Richards, *J. Evol. Biol.* **19**, 1744 (2006).
25. J. Weber, A. Czarnetzki, C. M. Pusch, *Science* **310**, 236 (2005); www.sciencemag.org/cgi/content/full/310/5746/236b.
26. D. Falk *et al.*, *Science* **310**, 236 (2005); www.sciencemag.org/cgi/content/short/310/5746/236c.
27. E. Culotta, *Science* **317**, 740 (2007).
28. M. W. Tocheri *et al.*, *Science* **317**, 1743 (2007).

29. We thank the families for giving their consent for this study, D. Schweitzer and K. Thoma for excellent technical assistance, and H. Regus-Leidig and J. H. Brandstätter for help with microscopic imaging. Supported by Bundesministerium für Bildung und Forschung (BMBF) network grant SKELNET GFGM01141901 to A. Rauch and A. Reis. The Wellcome Trust supported R.C.T. (grant 062346/Z/00/Z) and R.K.S. (grant 080952/Z/06/Z) and E.K. is a Medical Research Council Clinical Training Fellow. The study was approved by the Ethical Review Board of the Medical Faculty of the Friedrich-Alexander University Erlangen-Nuremberg. This paper is dedicated to the memory of the late Frank Majewski, born 14 May 1941, died 22 December 2001, Düsseldorf, in recognition of his contributions in clinical genetics.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1151174/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S4
References

1 October 2007; accepted 18 December 2007
Published online 3 January 2008;
10.1126/science.1151174
Include this information when citing this paper.

Reciprocal Binding of PARP-1 and Histone H1 at Promoters Specifies Transcriptional Outcomes

Raga Krishnakumar,[1,2]* Matthew J. Gamble,[1]* Kristine M. Frizzell,[1,2] Jhoanna G. Berrocal,[1,2] Miltiadis Kininis,[1,3] W. Lee Kraus[1,2,3,4]†

Nucleosome-binding proteins act to modulate the promoter chromatin architecture and transcription of target genes. We used genomic and gene-specific approaches to show that two such factors, histone H1 and poly(ADP-ribose) polymerase-1 (PARP-1), exhibit a reciprocal pattern of chromatin binding at many RNA polymerase II–transcribed promoters. PARP-1 was enriched and H1 was depleted at these promoters. This pattern of binding was associated with actively transcribed genes. Furthermore, we showed that PARP-1 acts to exclude H1 from a subset of PARP-1–stimulated promoters, suggesting a functional interplay between PARP-1 and H1 at the level of nucleosome binding. Thus, although H1 and PARP-1 have similar nucleosome-binding properties and effects on chromatin structure in vitro, they have distinct roles in determining gene expression outcomes in vivo.

Gene expression outcomes are determined, in part, by the composition of promoter chromatin, including the post-translational modification state of nucleosomal histones (*1*), the incorporation of histone variants (*2*), and the presence of nucleosome-binding proteins (*3*). Linker histone H1 and poly(ADP-ribose) polymerase-1 (PARP-1) are examples of nucleosome-binding proteins that modulate the chromatin architecture and transcription of target genes (*4*, *5*). H1 and PARP-1 bind to overlapping sites on nucleosomes at or near the dyad axis where the DNA exits the nucleosome (*6*, *7*). Unlike H1, PARP-1 has an intrinsic nicotinamide adenine dinucleotide (NAD^+)–dependent enzymatic activity that regulates its association with chromatin (*7*). Previous work from our laboratory has shown that H1 and PARP-1 bind in a competitive and mutually exclusive manner to nucleosomes in vitro and localize to distinct nucleosomal fractions in vivo (*7*), suggesting distinct roles for these factors in the regulation of gene expression. However, little is known about how H1 and PARP-1 are distributed across the mammalian genome and how they interact to regulate global patterns of gene expression in vivo.

To determine the patterns of H1 and PARP-1 localization across selected regions of the human genome, we performed chromatin immunoprecipitation (ChIP) in MCF-7 breast cancer cells using antibodies specific to PARP-1 and H1 (*7*, *8*), coupled with hybridization of the enriched genomic DNA to custom microarrays (i.e., ChIP-chip) (*9*). Each array represented 57 Mb of genomic DNA, including all 44 of the ENCODE regions (*10*), as well as an additional 1117 promoter regions selected from genes regulated by enzymes in the nuclear NAD^+ signaling pathway (*5*) [approximately −25 to +5 kb relative to the transcription start site (TSS)]. The raw ChIP-chip signal to input ratios were processed (*11*) and aligned to the TSSs for all 1517 RNA polymerase II (Pol II)–transcribed promoters on the array (i.e., ENCODE + selected). We observed an enrichment of PARP-1 and a depletion of H1 in the region surrounding the TSSs (Fig. 1A and fig. S1). Significant peaks of PARP-1 and troughs of H1 [$P < 0.01$, Wilcoxon signed-rank test (*12*, *13*)] were clustered around the TSSs, but were also found in upstream and intergenic regions (Fig. 1, B and C, and figs. S2 and S3). This pattern of PARP-1 and H1 localization was also revealed by averaging the ChIP-chip data over the 30-kb tiled region for all promoters on the array or in a 20-kb region centered around significant PARP-1 peaks or H1 troughs ($P < 0.01$, Wilcoxon signed-rank test) (fig. S4). Collectively, our ChIP-chip data identify a reciprocal relation for chromatin binding by PARP-1 and H1 across the genome.

Although eukaryotic promoters generally show reduced nucleosome occupancy (*14*, *15*),

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA. [2]Graduate Field of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA. [3]Graduate Field of Genetics and Development, Cornell University, Ithaca, NY 14853, USA. [4]Department of Pharmacology, Weill Medical College of Cornell University, New York, NY 10021, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed at Department of Molecular Biology and Genetics, Cornell University, 465 Biotechnology Building, Ithaca, NY 14853, USA. E-mail: wlk5@cornell.edu

this was not an important determinant for the reciprocal pattern of PARP-1 and H1 binding. For example, whereas PARP-1 peaks and H1 troughs are strongly correlated at promoters (Spearman rank correlation: −0.495, $P = 3.7 \times 10^{-94}$), they show little correlation with the presence of H3 (Fig. 2A; see also SOM Text). In addition, the pattern of PARP-1 and H1 binding at promoters (e.g., low versus high PARP-1/H1 ratios) is independent of the pattern of H3 occupancy at promoters (Fig. 2B). Finally, the reciprocal pattern of PARP-1 and H1 binding is observed in intergenic regions where H3 is not depleted (fig. S4B). Despite the reduced H3 occupancy at promoters, well-positioned nucleosomes are present at PARP-1–bound promoters that likely serve as targets for the binding of PARP-1 (fig. S5).

In a previous study (*7*), we concluded that PARP-1 may act to repress Pol II transcription based on the observations that (i) PARP-1 represses in vitro transcription by Pol II with chromatin templates in the absence of NAD^+ and (ii) PARP-1 does not colocalize with active Pol II (Ser^5-P) on *Drosophila* polytene chromosomes. Our current ChIP-chip results suggest that the latter observation may simply be a consequence of the localization of PARP-1 and active RNA Pol II to distinct regions of a gene (i.e., upstream versus downstream of the TSS; see SOM text). To explore the relations between PARP-1, H1, and gene expression in more detail and under physiological NAD^+ concentrations, we coupled our ChIP-chip analyses with gene expression microarray analyses for MCF-7 cells grown under the same conditions. PARP-1 peaks showed a significant positive correlation with gene expression (Spearman rank correlation, $P = 7.1 \times 10^{-49}$), whereas H1 showed a significant negative correlation with gene expression (Spearman rank correlation, $P = 7.85 \times 10^{-39}$) (Fig. 2A). In addition, PARP-1 was enriched and H1 was depleted near the TSSs of expressed genes relative to unexpressed genes (Fig. 2B) (*16*). We then grouped all genes containing both a significant PARP-1 peak and a significant H1 trough ($P < 0.01$, Wilcoxon signed-rank test) and compared them to a group that lacked both a PARP-1 peak and an H1 trough (*17*). More than 90% of the genes containing both a PARP-1 peak and an H1 trough at the promoter were expressed, whereas less than 45% of the genes lacking both a PARP-1 peak and an H1 trough at the promoter were expressed (Fig. 2C). This correlation was also observed when looking broadly across ENCODE regions enriched in expressed or unexpressed genes (Fig. 1C and fig. S2; see asterisks). Together, these results indicate that the pattern of PARP-1 and H1 promoter localization is indicative of gene expression outcomes.

Finally, to explore further the functional relations between PARP-1, H1, and gene expression, we identified subsets of PARP-1–bound genes either down-regulated or up-regulated in MCF-7 cells by stable short hairpin RNA (shRNA)-mediated knockdown of PARP-1 (Fig. 3A) (*18*). For each gene, we assayed (i) promoter binding by PARP-1 and H1 using ChIP-qPCR (quantitative polymerase chain reaction) and (ii) expression by reverse transcription (RT)-qPCR, with or without PARP-1 knockdown. The subset of genes positively regulated by PARP-1 (i.e., genes whose expression decreased upon PARP-1 knockdown) showed a three- to fivefold increase in H1 binding at the promoter in response to PARP-1 knockdown without changes in H3 occupancy (Fig. 3B and figs. S6 and S7). These results provide a functional link between the chromatin binding and gene-regulatory actions of PARP-1 and H1 at this subset of target promoters. Specifically, they suggest that PARP-1 acts to exclude H1 from these promoters and that upon PARP-1 knockdown, H1 is able to rebind and inhibit transcription. In contrast, the subset of genes negatively regulated or not regulated by PARP-1 (i.e., genes whose expression decreased or was unchanged upon PARP-1 knockdown) showed little or no change in H1 binding at the promoter in response to PARP-1 knockdown (Fig. 3C and fig. S8). These genes, some of which show a reciprocal pattern of PARP-1 and H1 localization at their promoters (Fig. 3C and fig. S8), may be subject to other PARP-1–related transcriptional regulatory mechanisms (*5*, *19*) or indirect regulatory effects.

Collectively, our data reveal the genomic localization patterns of H1 and PARP-1, highlighting the reciprocal relation for their binding at promoters and other genomic locations. In addition, our results provide a functional link between chromatin binding by PARP-1 and H1 at a subset of target promoters and the corresponding gene expression outcomes. Finally, our results suggest that PARP-1 acts to exclude H1 from a subset of PARP-1–regulated promoters in vivo. Our data fit well with and extend the results of previous biochemical and cell-based assays showing a role for PARP-1 in the transcription-related regulation of chromatin structure (*7*, *20*, *21*) and functional interplay between H1 and PARP-1 (*7*, *22*, *23*). Further, our results show that although H1 and PARP-1 have similar nucleosome-binding properties and effects on chromatin structure in vitro (*7*, *20*), they have distinct roles in regulating gene expression outcomes in vivo. Future studies will examine the determinants that direct the specific pattern of H1 and PARP-1 binding at promoters, including the role of PARP-1's NAD^+-dependent enzymatic activity.

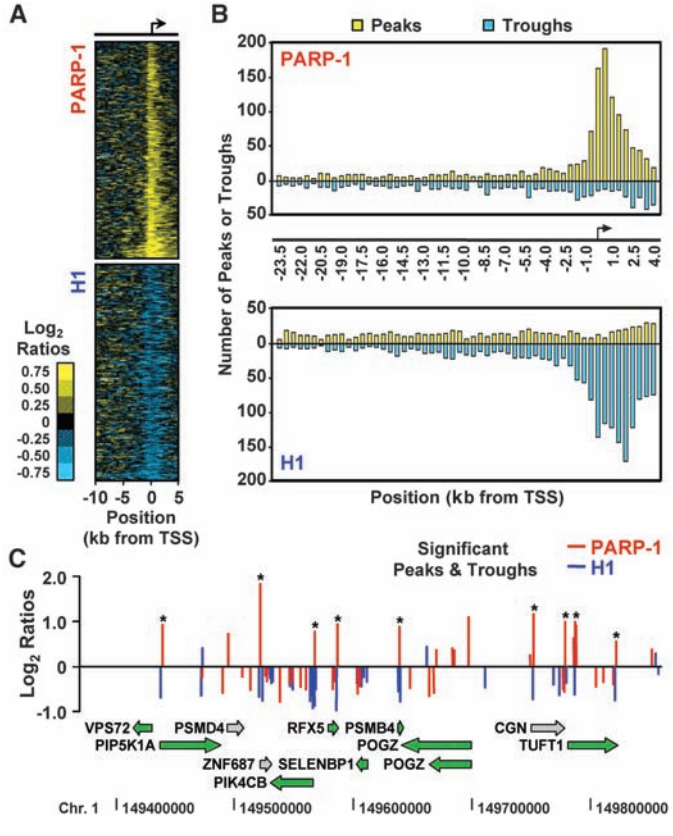


Fig. 1. Distinct patterns of genomic localization for H1 and PARP-1. (**A**) Heat maps of H1 and PARP-1 ChIP-chip data shown from −10 to +5 kb relative to the TSS. The data are limited to 758 promoters with the highest PARP-1 ChIP-chip signals and are ordered based on the intensity of the PARP-1 signal at the promoter. See fig. S1 for the full data set. (**B**) Histograms showing the number of statistically significant peaks and troughs of PARP-1 and H1 across the entire 30-kb tiled region for the 1517 promoters on the ChIP-chip array. (**C**) Diagram of statistically significant peaks and troughs of PARP-1 and H1 across an ENCODE region from chromosome 1 (Chr. 1). Annotated Reference Sequence (RefSeq) genes are represented by arrows indicating the length of the gene and direction of transcription. Green arrows: expressed in MCF-7 cells as determined by expression microarrays; gray arrows: ambiguous or no expression information available. Asterisks indicate genomic locations with a PARP-1 peak and a H1 trough.

References and Notes

1. S. L. Berger, *Curr. Opin. Genet. Dev.* **12**, 142 (2002).
2. R. T. Kamakaka, S. Biggins, *Genes Dev.* **19**, 295 (2005).

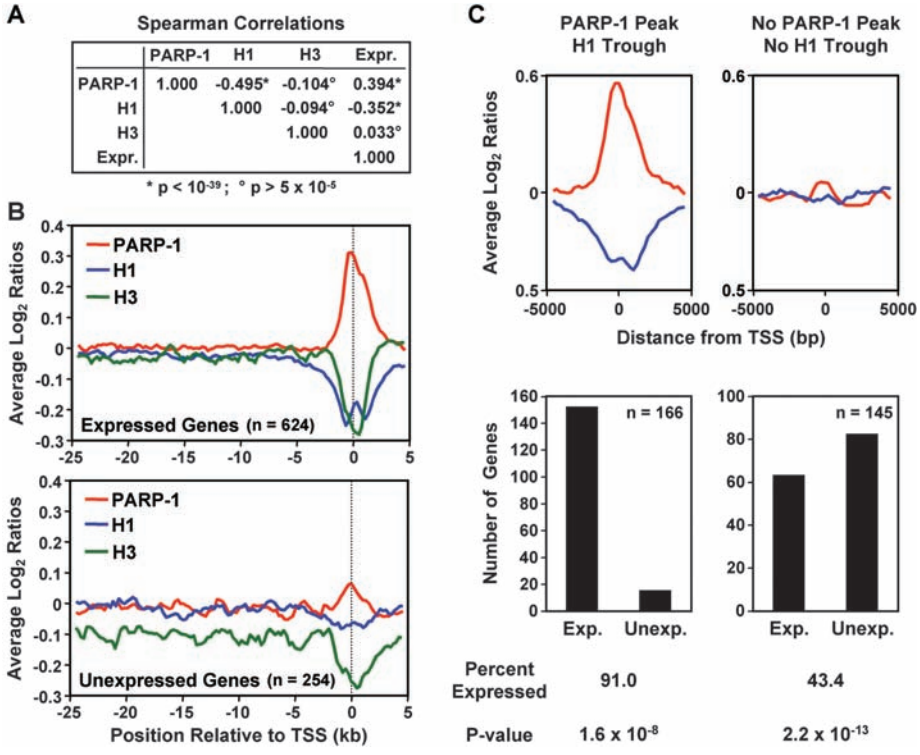


Fig. 2. A high PARP-1:H1 ratio specifies actively transcribed promoters. (**A**) Correlation analyses of PARP-1, H1, and H3 occupancy as determined by ChIP-chip (at the −250 bp-centered window) with gene expression (Expr.) as determined by microarrays. (**B**) Averaging analysis of the $\log_2$ enrichment ratios from H1 and PARP-1 ChIP-chip for unambiguously expressed (top) or unambiguously unexpressed genes (bottom). (**C**) Top: Averaging analysis of the $\log_2$ enrichment ratios from H1 and PARP-1 ChIP-chip for genes (i) having both a PARP-1 peak and an H1 trough within 1.5 kb of the TSS (left) or (ii) unambiguously lacking both a PARP-1 peak and an H1 trough within 1.5 kb of the TSS (right). Bottom: Percentage of expressed and unexpressed genes in each category. *P* values are from a Chi-squared test and indicate significant differences relative to the total gene set (n = 878; percent expressed = 71.1).

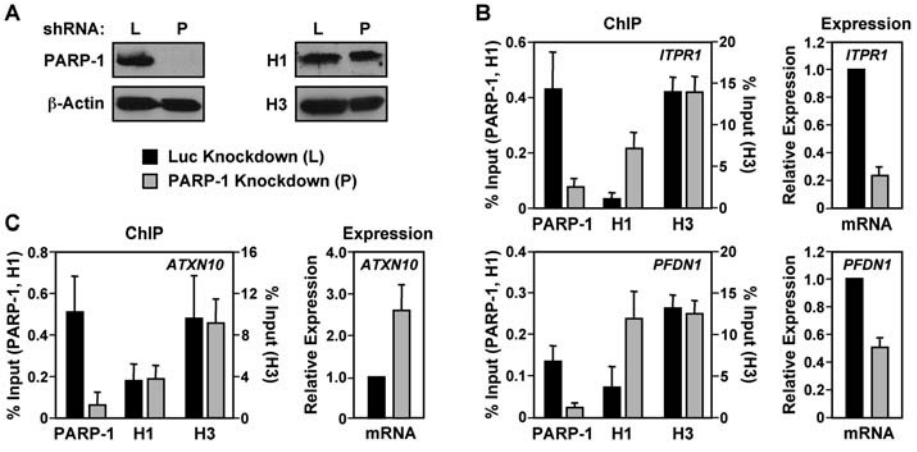


Fig. 3. PARP-1 excludes H1 from PARP-1-regulated promoters. (**A**) Western blot showing the shRNA-mediated depletion of PARP-1 in MCF-7 cells versus control luciferase (Luc) knockdown cells. (**B** and **C**) Gene-specific analysis of PARP-1, H1, and H3 promoter binding by ChIP-qPCR and mRNA expression by RT-qPCR in MCF-7 cells with or without PARP-1 knockdown. Expression data are standardized to β-actin transcripts. Bars represent the mean + SEM, $n \geq 3$.

3. S. J. McBryant, V. H. Adams, J. C. Hansen, *Chromosome Res.* **14**, 39 (2006).
4. C. L. Woodcock, A. I. Skoultchi, Y. Fan, *Chromosome Res.* **14**, 17 (2006).
5. M. Y. Kim, T. Zhang, W. L. Kraus, *Genes Dev.* **19**, 1951 (2005).
6. M. Vignali, J. L. Workman, *Nat. Struct. Biol.* **5**, 1025 (1998).
7. M. Y. Kim, S. Mauro, N. Gevry, J. T. Lis, W. L. Kraus, *Cell* **119**, 803 (2004).
8. M. Kininis *et al.*, *Mol. Cell. Biol.* **27**, 5090 (2007).
9. M. J. Buck, J. D. Lieb, *Genomics* **83**, 349 (2004).
10. The ENCODE Project Consortium, *Science* **306**, 636 (2004).
11. A single array error model was generated by the use of a 1-kb moving window with 250–base pair (bp) steps in which both the mean probe $\log_2$ ratio and *P* values from a nonparametric Wilcoxon signed-rank test were calculated for each window.
12. Significant peaks were defined as the center of three consecutive windows with positive means, the center window with a mean greater than that of either adjacent window, and all windows having *P* values less than 0.01 (Wilcoxon signed-rank test). Significant troughs were defined as the center of three consecutive windows with negative means, the center window with a mean less than that of either adjacent window, and all windows having *P* values less than 0.01 (Wilcoxon signed-rank test).
13. The use of our peak/trough selection criteria were justified by a low false-positive rate (FPR) as determined by ChIP-qPCR (PARP-1 peak FPR = 0.11; H1 trough FPR = 0.08).
14. N. D. Heintzman *et al.*, *Nat. Genet.* **39**, 311 (2007).
15. Y. Mito, J. G. Henikoff, S. Henikoff, *Science* **315**, 1408 (2007).
16. For a gene to be classified as unambiguously expressed or unexpressed, all probe sets from all three replicates corresponding to the gene must have been flagged unanimously present or absent, respectively. Any genes not meeting these criteria were marked as ambiguous and were removed from the expression-based categorization analysis.
17. For this analysis, peaks and troughs identified at *P* values between 0.01 and 0.1 were labeled as ambiguous due to high false-positive and false-negative rates.
18. The target genes used for this analysis were identified in a microarray expression screen and were then confirmed by RT-qPCR as having either a twofold reduction or a twofold increase in expression upon shRNA-mediated knockdown of PARP-1.
19. W. L. Kraus, J. T. Lis, *Cell* **113**, 677 (2003).
20. D. A. Wacker *et al.*, *Mol. Cell. Biol.* **27**, 7475 (2007).
21. A. Tulin, A. Spradling, *Science* **299**, 560 (2003).
22. B. G. Ju *et al.*, *Science* **312**, 1798 (2006).
23. A. Huletsky *et al.*, *J. Biol. Chem.* **264**, 8878 (1989).

24. We thank J. Lis, A. Clark, A. Siepel, and N. Hah for critical reading of this manuscript and A. Clark and members of the Kraus laboratory for technical advice and helpful discussions. This work was supported by grants from the NIH–National Institute of Diabetes and Digestive and Kidney Diseases (DK069710 and DK058110), the Cornell Center of Vertebrate Genomics, and the Endocrine Society (W.L.K.); a postdoctoral fellowship from the American Heart Association (M.J.G.); and predoctoral fellowships from the American Heart Association (K.M.F.), the Alfred P. Sloan Foundation (J.G.B.), and the Department of Defense Breast Cancer Research Program (M.K.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5864/819/DC1
Materials and Methods
SOM Text
Figs. S1 to S8
References

15 August 2007; accepted 21 December 2007
10.1126/science.1149250

Repression of the Transcription Factor Th-POK by Runx Complexes in Cytotoxic T Cell Development

Ruka Setoguchi,[1]*† Masashi Tachibana,[1]* Yoshinori Naoe,[1]* Sawako Muroi,[1,2] Kaori Akiyama,[1,2] Chieko Tezuka,[1] Tsukasa Okuda,[3] Ichiro Taniuchi[1,2]‡

Mouse $CD4^+CD8^+$ double-positive (DP) thymocytes differentiate into $CD4^+$ helper-lineage cells upon expression of the transcription factor Th-POK but commit to the $CD8^+$ cytotoxic lineage in its absence. We report the redirected differentiation of class I–restricted thymocytes into $CD4^+CD8^-$ helper-like T cells upon loss of Runx transcription factor complexes. A Runx-binding sequence within the *Th-POK* locus acts as a transcriptional silencer that is essential for *Th-POK* repression and for development of $CD8^+$ T cells. Thus, Th-POK expression and genetic programming for T helper cell development are actively inhibited by Runx-dependent silencer activity, allowing for cytotoxic T cell differentiation. Identification of the transcription factors network in CD4 and CD8 lineage choice provides insight into how distinct T cell subsets are developed for regulating the adaptive immune system.

The peripheral T cell repertoire is formed after developing thymocytes have undergone a series of developmental selection processes. $CD4^+CD8^+$ double-positive (DP) thymocytes undergo positive selection through T cell receptor (TCR) interaction with major histocompatibility complex (MHC) proteins. This gives rise to two functionally distinct subsets: $CD4^+CD8^-$ helper and $CD4^-CD8^+$ cytotoxic T cells. Cells expressing MHC class II–restricted TCRs differentiate into the helper lineage and cease CD8 expression, whereas cells expressing class I–restricted TCRs differentiate into the cytotoxic linage and silence CD4 expression (*1*–*3*). Recently, gain or loss of function of the BTB/POZ domain–containing zinc finger transcription factor, Th-POK, revealed that its expression is essential and sufficient for development of helper-lineage cells (*4*, *5*).

[1]Laboratory for Transcriptional Regulation, RIKEN Research Center for Allergy and Immunology, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan. [2]Precursory Research for Embryonic Science and Techonology (PRESTO), Japan Science and Technology Agency (JST), 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan. [3]Kyoto Prefectural University of Medicine, Kawaramachi-Hirokoji, Kamigyo-ku, Kyoto 602-8566, Japan.

*These authors contributed equally to this work.
†Present address: Department of Immunology, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195–7650, USA.
‡To whom correspondence should be addressed. E-mail: taniuchi@rcai.riken.jp

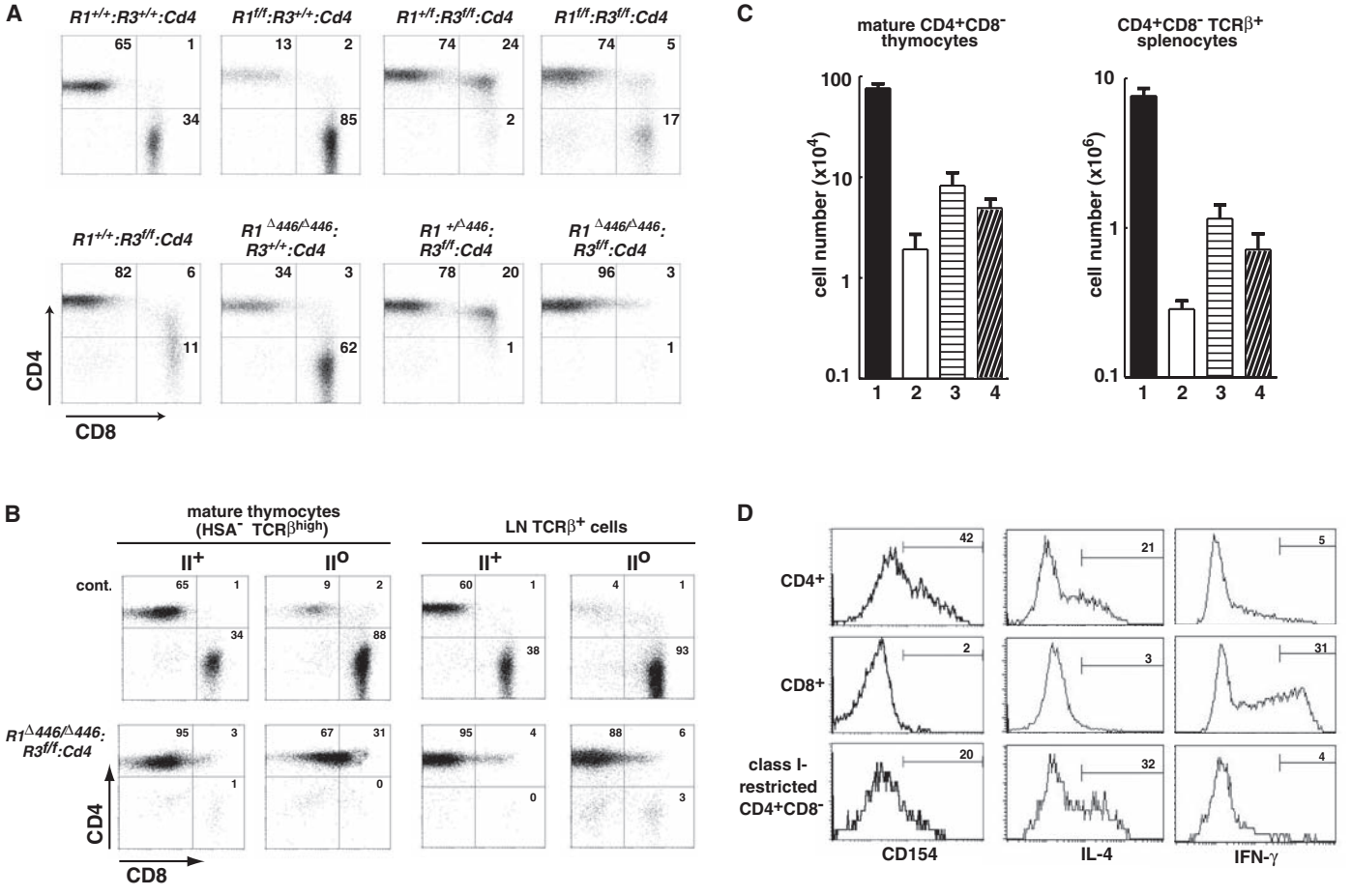


Fig. 1. Differentiation of class I–restricted cells into $CD4^+CD8^-$ helper-like cells by loss of Runx complex function. (**A**) CD4 and CD8 expression in lymph node αβT cells from mice with indicated genotypes. (**B**) CD4 and CD8 expression in mature thymocytes and LN $TCR\beta^+$ T cells either in the presence (II^+) or absence (II°) of I-A MHC class II molecules. (**C**) Cell numbers of mature thymocytes and splenocytes showing $CD4^+CD8^-$ αβT cells in class II^+ control mice (lane 1), class II° control mice (lane 2), class II^+ *Runx1*$^{\Delta446/\Delta446}$:*Runx3*$^{f/f}$:*Cd4* mice (lane 3), and class II° *Runx1*$^{\Delta446/\Delta446}$:*Runx3*$^{f/f}$:*Cd4* mice (lane 4). Error bars indicate standard deviation. (**D**) Expression of CD154 at 42 hours after in vitro TCR stimulation of control $CD4^+$, $CD8^+$, and class I–restricted $CD4^+CD8^-$ cells. Intracellular staining of IL-4 and IFN-γ analyzed at 6 hours after re-stimulation of cells that were cultured for 5 days after initial TCR stimulation. Numbers in the plots in (A), (B), and (D) indicate the percentage of cells in each quadrant or region.

Runx transcription factor complexes are composed of heterodimers for one of three Runx proteins and their obligatory non–DNA-binding partner, Cbfβ protein (*6*). Because of the embryonic or neonatal lethality of mice deficient for any of Runx family genes, we used the Cre/loxP-mediated conditional gene inactivation (*7*) to clarify Runx complex function in silencing of the *Cd4* gene (*8*) and recently reported that the combined inactivation of *Runx1* and *Runx3* in DP thymocytes resulted in a dramatic loss of $CD8^+$ T cells (*9*). Runx proteins possess a conserved Val-Trp-Arg-Pro-Tyr (VWRPY) motif at the C-terminal end, allowing the recruitment of the Groucho/TLE co-repressor proteins to their target genes (*10*, *11*). To test whether VWRPY-dependent repression might be involved in the loss of $CD8^+$ T cells, we introduced the $Runx1^{\Delta 446}$ allele (*12*) that generates a mutant Runx1 protein lacking the VWRPY motif on a Runx3-deficient background ($Runx3^{f/f}$:*Cd4* mice) (*13*). A marked reduction of splenic $CD8^+$ T cells in $Runx1^{\Delta 446/\Delta 446}$:$Runx3^{f/f}$:*Cd4* mice (Fig. 1A and fig. S1) indicated that VWRPY-dependent repres-

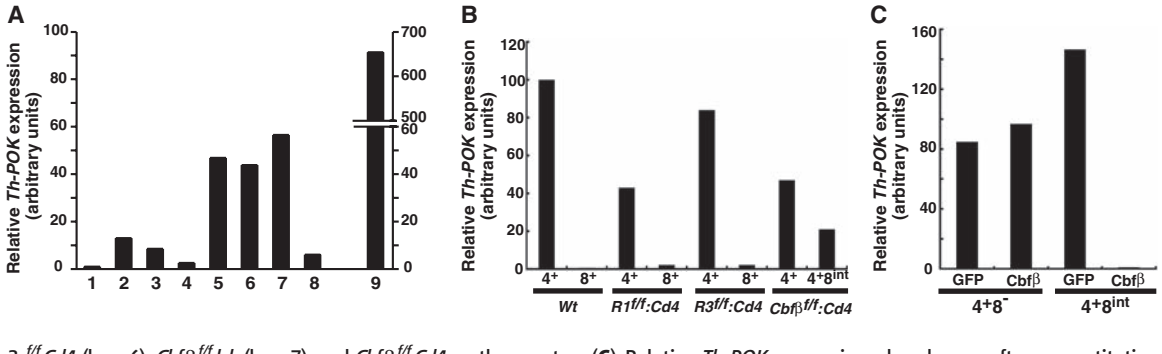
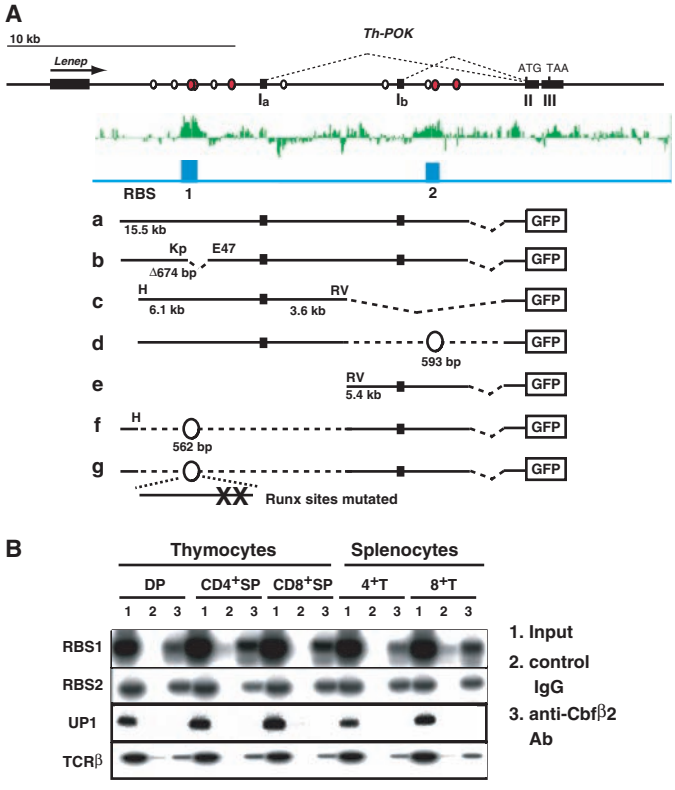


Fig. 2. De-repression of *Th-POK* by loss of Runx complex function. (**A** and **B**) Relative *Th-POK* expression abundances (normalized to *hprt*) in sorted $CD69^-$ DP thymocytes (A) from wild-type (lane 1), $Runx1^{f/f}$:*Cd4* (lane 2), $Runx1^{\Delta 446/\Delta 446}$ (lane 3), $Runx3^{f/f}$:*Cd4* (lane 4), $Runx1^{f/f}$:$Runx3^{f/f}$:*Cd4* (lane 5), $Runx1^{\Delta 446/\Delta 446}$:$Runx3^{f/f}$:*Cd4* (lane 6), $Cbf\beta^{f/f}$:*lck* (lane 7), and $Cbf\beta^{f/f}$:*Cd4* (lane 8) mice and in $CD4^+$ and $CD8^+$ peripheral T cells in mice of the indicted genotype (B). One representative result out of three experiments is shown. Lane 9 in (A) indicates *Th-POK* expression in control $CD4^+CD8^-$ SP thymocytes. (**C**) Relative *Th-POK* expression abundances after reconstitution of Runx complex function. Purified $CD4^+CD8^-$ and $CD4^+CD8^{int}$ cells from $Cbf\beta^{f/f}$:*Cd4* mice were transduced with control retroviral vector (GFP) or vector encoding Cbfβ (Cbfβ).



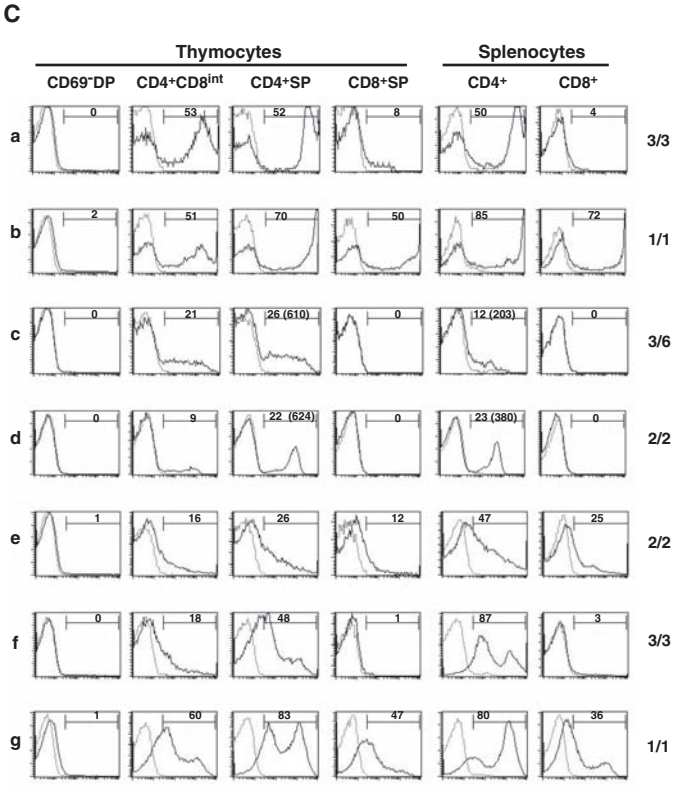


Fig. 3. Identification and characterization of RBSs at the *Th-POK* locus. (**A**) The structure of the murine *Th-POK* locus is shown at the top. Circles represent putative Runx motifs, with those in red indicating evolutionarily conserved Runx motifs. Black boxes represent exons, and each green bar represents the signal intensity of an individual oligonucleotide probe in a ChIP-on-chip experiment. Blue boxes represent RBSs. The maps for each reporter transgene construct (Tg-a to Tg-g) are indicated. The restriction sites shown are Eco47III (E47), EcoRV (RV), HindIII (H), KpnI (Kp), and XhoI (X). (**B**) ChIP experiment showing binding of Runx complexes to RBS-1 and RBS-2 in the indicated cell subsets. The regions at 1 kb upstream of exon Ia (UP1) and the TCRβ enhancer (TCRβ) were used as negative and positive controls, respectively. (**C**) Histograms showing the GFP expression in the indicated T cell subsets from representative transgenic founder for each construct. The dashed line indicates nontransgenic littermate control. Numbers in the histogram indicate the percentage of GFP^+ cells, and numbers in parenthesis indicate mean fluorescent intensity of GFP in GFP^+ cells. The numbers of transgenic founders expressing GFP among the total transgenic founders are indicated at right.

sion by Runx1 was involved in the generation of $CD8^+$ T cells. Because the leaky $CD4^-CD8^+$ subset that escaped Cre-mediated recombination (*9*) was less apparent in *Runx1*$^{\Delta 446/\Delta 446}$:*Runx3*$^{f/f}$:*Cd4* mice (Fig. 1A), we used these mice for further analyses.

Potentially, the loss of $CD8^+$ T cells could occur either by a developmental block of class I–restricted cells or by a redirection of class I–restricted cells toward the $CD4^+CD8^-$ lineage. To determine whether $CD4^+CD8^-$ cells that emerge in Runx mutant mice are class II–restricted or redirected class I–restricted cells, we crossed *Runx1*$^{\Delta 446/\Delta 446}$:*Runx3*$^{f/f}$:*Cd4* mice onto a MHC class II–deficient background (*14*). Although there was a marked decrease in $CD4^+CD8^-$ T cell numbers in control class II–deficient mice, the predominance of $CD4^+CD8^-$ T cells persisted in class II–deficient *Runx1*$^{\Delta 446/\Delta 446}$:*Runx3*$^{f/f}$:*Cd4* mice in both the thymus and the periphery (Fig. 1, B and C). These results indicated that the absence of Runx complexes forced the majority of class I–restricted cells to differentiate into $CD4^+CD8^-$ T cells.

We next examined the functional properties of these $CD4^+CD8^-$ cells. One of the characteristic features of $CD4^+$ helper-lineage T cells is the early induction of CD154, the ligand for CD40, after TCR stimulation (*15*) and the production of interleukin-4 (IL-4). These were observed in control $CD4^+$ T cells as well as in class I–restricted $CD4^+CD8^-$ cells, but not in control $CD8^+$ T cells (Fig. 1D). In contrast, although high interferon-γ (IFN-γ) production was detected in control $CD8^+$ T cells, it was absent in both wild-type $CD4^+$ T cells and in class I–restricted $CD4^+CD8^-$ cells (Fig. 1D). We conclude from these results that class I–restricted $CD4^+CD8^-$ cells that develop in Runx mutant animals are functionally helper-like T cells.

Because ectopic expression of Th-POK has been shown to redirect class I–restricted cells to become $CD4^+CD8^-$ cells (*4*, *5*), we measured the expression of *Th-POK* in several Runx mutant mice, including a strain in which the *Cbf*β gene is conditionally inactivated by either a *Lck-Cre* or a *Cd4-Cre* transgene (*13*). Consistent with a previous report (*4*), *Th-POK* expression was not detected in control $CD69^-$ DP thymocytes. In contrast, a 40-fold increase in *Th-POK* transcript abundances was detected in $CD69^-$ DP thymocytes in which Runx complexes were disrupted either by combined Runx1 mutations with a Runx3 deficiency or by loss of Cbfβ protein (*Cbfβ*$^{f/f}$:*Lck* mice) (Fig. 2A). A modest *Th-POK* de-repression by inactivation of *Runx1* alone indicated a redundant function of Runx3 in the repression of *Th-POK*.

Although *Th-POK* mRNA was undetectable in control $CD8^+$ T cells and in $CD8^+$ T cells deficient for Runx1 or Runx3, it was present in Cbfβ-deficient $CD4^+CD8^{int}$ T cells (Fig. 2B) that still developed in *Cbf*β$^{f/f}$:*Cd4* mice because of the slow turnover of Cbfβ protein after inactivation of the *Cbf*β gene (*13*). We therefore next examined whether *Th-POK* repression could be restored in these $CD4^+CD8^{int}$ cells upon re-expression of Cbfβ protein. Purified $CD4^+CD8^-$ and $CD4^+CD8^{int}$ cells were transduced with a retroviral vector encoding Cbfβ or with an empty vector control. In these experiments, expression of *Th-POK* was markedly reduced upon re-expression of Cbfβ in $CD4^+CD8^{int}$ cells, with no detectable effect in $CD4^+CD8^-$ cells (Fig. 2C). These results suggest that Runx-mediated *Th-POK* repression operates in peripheral $CD8^+$ T cells.

To understand mechanisms underlying Runx-mediated repression of *Th-POK*, we examined whether Runx complexes directly associate with the *Th-POK* locus. Using a ChIP-on-chip (ChIP indicates chromatin immunoprecipitation) approach with an antibody against Cbfβ2, we detected two regions occupied by Runx complexes within the *Th-POK* locus. Distal and proximal Runx-binding sequences (RBS-1 and RBS-2, respectively) are located ~3.1 kb upstream and ~7.4 kb downstream of exon Ia (Fig. 3A) and contain two or one conserved Runx motifs, respectively (Fig. 3A and figs. S2 and S3). By using ChIP analysis in T cell subsets, we confirmed an association between Runx complexes and these two regions (Fig. 3B). However, binding of Runx complexes to RBS-1 and RBS-2 was detected in both *Th-POK*–expressing and nonexpressing cells, revealing that the binding of Runx complexes to these regions did not correspond with *Th-POK* repression.

To better understand the functional activities of RBS-1 and RBS-2 in light of these results, we performed transgenic reporter assays. A 15.5-kb genomic fragment encompassing the RBSs and exons Ia and Ib was linked to a green fluorescent protein (GFP) reporter transgene cassette (Tg-a in Fig. 3A). In all transgenic mouse founders obtained with Tg-a, GFP expression was first detected in post-selection $CD4^+CD8^{int}$ thymocytes and was up-regulated in $CD4^+$ SP thymocytes, remaining

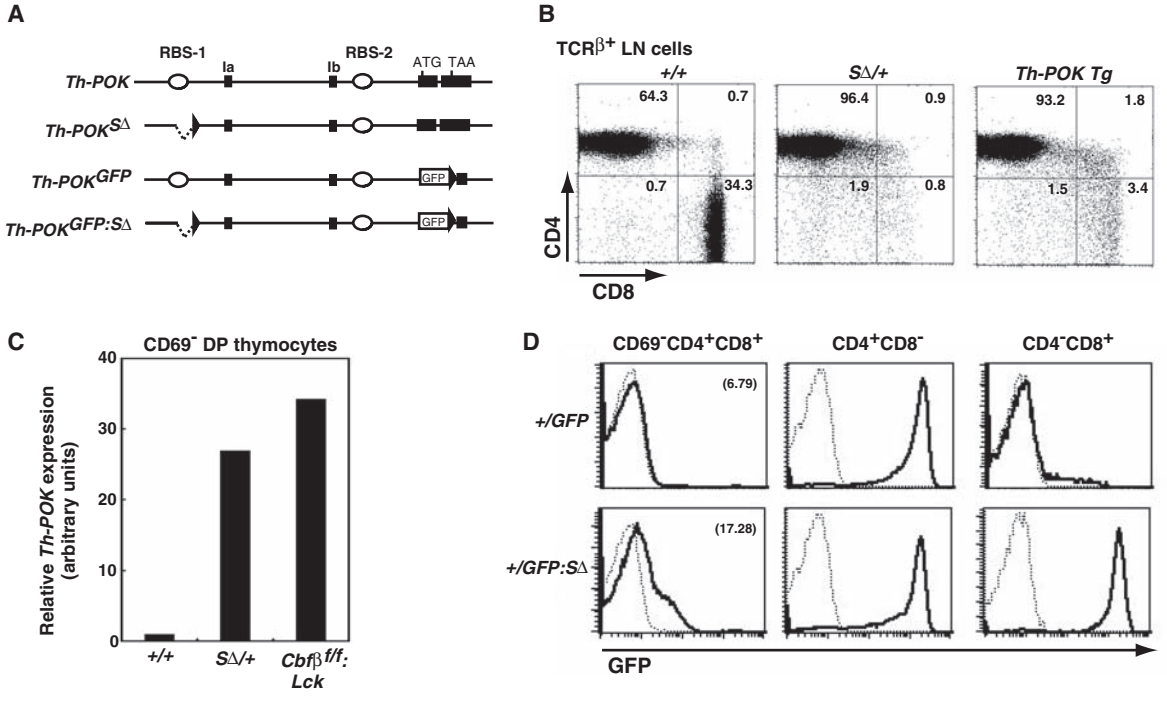


Fig. 4. Essential requirement of the *Th-POK* silencer for development of $CD8^+$ T cells. (**A**) Schematic structure of the *Th-POK* locus and targeted alleles *Th-POK*$^{S\Delta}$, *Th-POK*^{GFP}, and *Th-POK*$^{GFP:S\Delta}$. Exons and loxP sequences are indicated as black boxes and black triangles, respectively. (**B**) CD4 and CD8 expression in lymph node αβT cells from wild-type (+/+), *Th-POK*$^{S\Delta}$ heterozygous (*S*Δ/+), and *Th-POK* hemizygous transgenic (*Th-POK Tg*) mice. (**C**) Relative *Th-POK* expression abundances in sorted $CD69^-$ DP thymocytes showing de-repression of *Th-POK* upon deletion of the *Th-POK* silencer. (**D**) GFP expression from the *Th-POK*^{GFP} and *Th-POK*$^{GFP:S\Delta}$ alleles in indicated thymocyte subsets. Dashed and bold lines indicate GFP expression in control mice and *Th-POK*$^{+/GFP}$ (+/*GFP*) or *Th-POK*$^{+/GFP:S\Delta}$ (+/*GFP:S*Δ) mice, respectively. The numbers in parenthesis indicate mean fluorescent intensity of GFP in total $CD69^-$ DP thymocytes.

high in splenic $CD4^+$ T cells, whereas it was almost undetectable in splenic $CD8^+$ T cells (Fig. 3C). The 15.5-kb fragment thus contains the major *cis*-regulatory regions that direct expression of *Th-POK* in the helper lineage.

To further narrow down the critical *Th-POK* regulatory regions, we deleted either 5′ or 3′ sequences as well as RBS-1 from the 15.5-kb fragment. Whereas RBS-2 (fig. S3) was found to be required for positive transcriptional regulatory activity (as in the Tg-c and Tg-d constructs), deletion of a 674-bp fragment of RBS-1 (Tg-b) resulted in GFP expression both in $CD4^+$ helper-lineage and in $CD8^+$ cytotoxic-lineage cells, indicating that RBS-1 is a transcriptional silencer required to repress the reporter gene in CD8 lineage cells. Efficient repression of GFP in $CD8^+$ T cells by a 562-bp fragment of RBS-1 (fig. S2) in the context of Tg-e construct required Runx motifs (Tg-f and Tg-g) (Fig. 3C), consistent with Runx-dependent activity of RBS-1 silencer.

To examine the physiological function of the RBS-1 silencer, we deleted the 674-bp KpnI-Eco47III sequences from the *Th-POK* locus by homologous recombination in embryonic stem (ES) cells (Fig. 4A and fig. S4). Deletion of RBS-1 from one *Th-POK* allele led to the loss of peripheral $CD8^+$ T cells (Fig. 4B) and to the *Th-POK* de-repression in $CD69^-$ DP thymocytes (Fig. 4C). We further investigated *Th-POK* de-repression by using mice in which the coding sequence for Th-POK was replaced with the *gfp* gene ($Th\text{-}POK^{GFP}$ locus). GFP expression in mice heterozygous for $Th\text{-}POK^{GFP}$ allows us to examine expression of *Th-POK* at the single-cell level. Although GFP expression from the $Th\text{-}POK^{GFP}$ locus was not detected in $CD69^-$ DP thymocytes, deletion of RBS-1 ($Th\text{-}POK^{GFP:S\Delta}$ locus in Fig. 4A) resulted in uniform de-repression of GFP in $CD69^-$ DP thymocytes, followed by high GFP expression in both helper- and cytotoxic-lineage mature thymocytes (Fig. 4D).

Our results reveal that helper lineage–specific expression of *Th-POK* is regulated by the RBS-1 silencer, whose activity depends on binding of Runx complexes. We therefore refer to RBS-1 as the *Th-POK* silencer (fig. S5). The association of Runx complexes with the *Th-POK* silencer in cells expressing *Th-POK* indicates that specificity of silencer activity is not regulated at the level of Runx complex binding. Additional molecules that interact with Runx factors bound to the *Th-POK* silencer may therefore have a central role in regulating *Th-POK* silencer activity.

The antagonistic interplay between primary lineage-determining factors is often observed when two opposing fates are induced in progenitor cells (*16*, *17*). Th-POK was recently described as an inhibitor of Runx-dependent *Cd4* silencer activity (*18*), consistent with an antagonistic interplay between these two factors. Identification of Th-POK and Runx complex target genes will help to further unravel the transcription factors network regulating lineage specification of DP thymocytes.

Uniform de-repression of Th-POK in $CD69^-$ DP thymocytes upon deletion of the *Th-POK* silencer indicates that silencer-mediated *Th-POK* repression operates in all pre-selection DP thymocytes. It is therefore possible that TCR signals after engagement of MHC class II result in antagonism of *Th-POK* silencer activity and thus induce *Th-POK* expression. Given that sustained class II–specific TCR signals are thought to be necessary for specification of the helper lineage (*19*–*21*), reversal of silencer-mediated *Th-POK* repression may require class II–specific TCR signals during a specified time window. Our results suggest that a mechanism regulating *Th-POK* silencer activity acts as a sensor to distinguish qualitative differences in TCR signaling. Further studies on the regulatory pathways of *Th-POK* repression will shed light on how signals initiated by external stimuli are converted into genetic programs in the cell nucleus.

References and Notes

1. A. Singer, R. Bosselut, *Adv. Immunol.* **83**, 91 (2004).
2. W. Ellmeier, S. Sawada, D. R. Littman, *Annu. Rev. Immunol.* **17**, 523 (1999).
3. T. K. Starr, S. C. Jameson, K. A. Hogquist, *Annu. Rev. Immunol.* **21**, 139 (2003).
4. X. He *et al.*, *Nature* **433**, 826 (2005).
5. G. Sun *et al.*, *Nat. Immunol.* **6**, 373 (2005).
6. Y. Ito, *Genes Cells* **4**, 685 (1999).
7. H. Gu, Y. R. Zou, K. Rajewsky, *Cell* **73**, 1155 (1993).
8. I. Taniuchi *et al.*, *Cell* **111**, 621 (2002).
9. T. Egawa, R. E. Tillman, Y. Naoe, I. Taniuchi, D. R. Littman, *J. Exp. Med.* **204**, 1945 (2007).
10. D. Levanon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11590 (1998).
11. B. D. Aronson, A. L. Fisher, K. Blechman, M. Caudy, J. P. Gergen, *Mol. Cell. Biol.* **17**, 5581 (1997).
12. M. Nishimura *et al.*, *Blood* **103**, 562 (2004).
13. Y. Naoe *et al.*, *J. Exp. Med.* **204**, 1749 (2007).
14. M. J. Grusby, R. S. Johnson, V. E. Papaioannou, L. H. Glimcher, *Science* **253**, 1417 (1991).
15. M. Roy, T. Waldschmidt, A. Aruffo, J. A. Ledbetter, R. J. Noelle, *J. Immunol.* **151**, 2497 (1993).
16. E. V. Rothenberg, *Nat. Immunol.* **8**, 441 (2007).
17. S. H. Orkin, *Nat. Rev. Genet.* **1**, 57 (2000).
18. K. F. Wildt *et al.*, *J. Immunol.* **179**, 4405 (2007).
19. K. Yasutomo, C. Doyle, L. Miele, C. Fuchs, R. N. Germain, *Nature* **404**, 506 (2000).
20. X. Liu, R. Bosselut, *Nat. Immunol.* **5**, 280 (2004).
21. S. D. Sarafova *et al.*, *Immunity* **23**, 75 (2005).

22. We are grateful to D. R. Littman and W. Ellmeier for critical reading of the manuscript. This work was supported by grants from PRESTO, JST. The accession number for mouse *Th-POK* silencer is EU371956 in GenBank.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5864/822/DC1
Material and Methods
Figs. S1 to S5
References

17 October 2007; accepted 20 December 2007
10.1126/science.1151844

A Heme Export Protein Is Required for Red Blood Cell Differentiation and Iron Homeostasis

Siobán B. Keel,[1]* Raymond T. Doty,[1]* Zhantao Yang,[1] John G. Quigley,[1]† Jing Chen,[1] Sue Knoblaugh,[2] Paul D. Kingsley,[3] Ivana De Domenico,[4] Michael B. Vaughn,[4] Jerry Kaplan,[4] James Palis,[3] Janis L. Abkowitz[1]‡

Hemoproteins are critical for the function and integrity of aerobic cells. However, free heme is toxic. Therefore, cells must balance heme synthesis with its use. We previously demonstrated that the feline leukemia virus, subgroup C, receptor (FLVCR) exports cytoplasmic heme. Here, we show that FLVCR-null mice lack definitive erythropoiesis, have craniofacial and limb deformities resembling those of patients with Diamond-Blackfan anemia, and die in midgestation. Mice with FLVCR that is deleted neonatally develop a severe macrocytic anemia with proerythroblast maturation arrest, which suggests that erythroid precursors export excess heme to ensure survival. We further demonstrate that FLVCR mediates heme export from macrophages that ingest senescent red cells and regulates hepatic iron. Thus, the trafficking of heme, and not just elemental iron, facilitates erythropoiesis and systemic iron balance.

Aerobic cells require heme, a cyclic tetrapyrole containing a centrally chelated iron. It serves as the prosthetic group for hemoglobin, cytochromes, and other hemoproteins. Heme also initiates globin transcription through inhibiting the DNA binding of the repressor, Bach1 (*1*), and globin translation through inhibiting substrate phosphorylation by the repressor, erythroid-specific eukaryotic initiation factor 2α kinase (*2*). However, the trafficking of heme and its role in iron homeostasis are poorly understood.

The feline leukemia virus, subgroup C (FeLV-C), receptor, FLVCR, is a heme export protein (*3*). Cats viremic with FeLV-C develop pure red cell aplasia (PRCA), characterized by a block in erythroid differentiation at the CFU-E (colony-forming unit–erythroid)–proerythroblast stage, reticulocytopenia, and severe anemia (*4*, *5*). Studies with chimeric retroviruses suggest that the surface unit of the FeLV-C envelope protein induces this phenotype by blocking FLVCR function (*6*, *7*). Although all bone marrow cells are infected (*8*), white cell and platelet production remain normal,

which suggests that FLVCR is uniquely important for CFU-E–proerythroblast survival or differentiation.

To prove that FLVCR is required for erythropoiesis, we generated constitutive (*Flvcr*$^{+/-}$) and inducible (*Flvcr*$^{+/flox}$;*Mx-cre*) *Flvcr* mutant mice (*9*) (fig. S1). Interbred *Flvcr*$^{+/-}$ animals yielded no null offspring (*Flvcr*$^{-/-}$) among 109 progeny (table S1). Intrauterine deaths occurred at one of two embryonic times: at or before embryonic day 7.5 (E7.5) and between E14.5 and E16.5.

Developmental expression of *Flvcr* is high in the yolk sac at E7.5, the ectoplacental cone at E8.5, and the placenta after E9.5 (Fig. 1A); all are sites of nutritional transport from mother to conceptus. These are also sites of high heme oxygenase–1 expression (*10*). As heme catabolism helps to support normal fetal development (*10*), FLVCR might complement this function at or before E7.5.

We hypothesize that the later death results from deficient red cell production, because definitive fetal erythropoiesis in the mouse begins in the liver at ~E12 (*11*), hepatic FLVCR expression is high from E12.5 onward (Fig. 1A), and FLVCR-null embryos have pale livers (Fig. 1B). Flow cytometric analyses of E14.5 fetal liver cells double-stained for Ter119 (erythroid-specific antigen) and CD71 (transferrin receptor) allow quantitative assessment of the maturational stages of differentiating erythroblasts (*12*) and confirm this concept. Normally, differentiation proceeds clockwise from population I to IV (control in Fig. 1C). In contrast, the null embryos lack Ter119^{high} cells, consistent with a block at the proerythroblast stage, before hemoglobinization (population II). Circulating yolk sac–derived erythroblasts do not express *Flvcr* by in situ hybridization and have normal morphology (fig. S3), which indicates that embryonic (primitive) erythropoiesis does not require FLVCR.

Although the null embryos appear normal at E8.5, E10.5, and E12.5, defective growth is evident at E14.5. Mutants have abnormal limb, hand, and digit maturation; flattened faces; and hypertelorism (Fig. 1B)—abnormalities that resemble human congenital PRCA, termed Diamond-Blackfan anemia (*13*, *14*). Gross and microscopic examination of the cardiac, pulmonary, and genitourinary systems shows that they are normal. Although it is theoretically possible that the observed phenotype is developmentally appropriate for a growth-retarded embryo, these specific abnormalities are not reported in other mouse models lacking definitive erythropoiesis (*11*, *15*). Thus,

[1]Division of Hematology, University of Washington, Seattle, WA 98195, USA. [2]Animal Health Shared Resources, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [3]Department of Pediatrics, University of Rochester, Rochester, NY 14642, USA. [4]Department of Pathology, University of Utah, Salt Lake City, UT 84132, USA.

*These authors contributed equally to this work.
†Present address: Division of Hematology-Oncology, University of Illinois at Chicago, Chicago, IL 60612, USA.
‡To whom correspondence should be addressed. E-mail: janabk@u.washington.edu

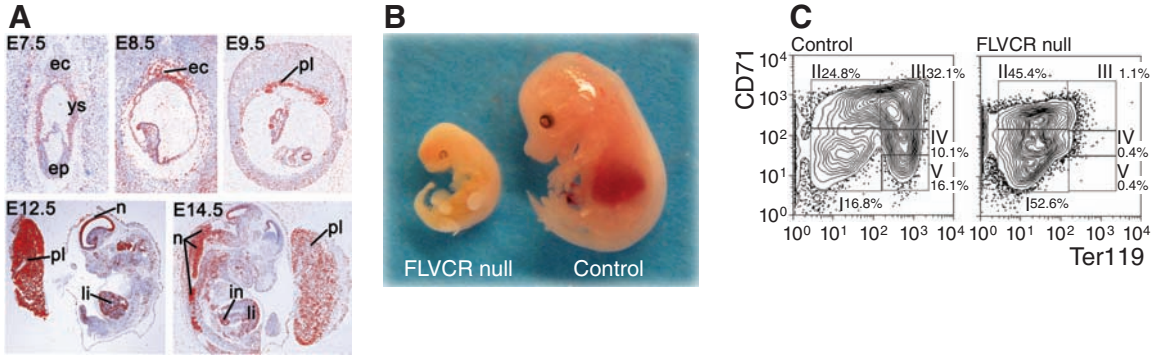


Fig. 1. Embryonic FLVCR analyses. (**A**) Wild-type mouse *Flvcr* expression (colored red) by in situ hybridization. Ectoplacental cone (ec), yolk sac (ys), embryo proper (ep), liver (li), neural tissue (n), placenta (pl) and intestine (in). Additional information is in SOM text. (**B**) E14.5 FLVCR-null embryo and a littermate control. The skeletal abnormalities are less apparent in embryos derived from interbreeding *Flvcr*$^{+/-}$ parental mice backcrossed to C57BL/6 mice for five to seven generations (SOM text). (**C**) Representative flow cytometric analyses of E14.5 liver cells from control and FLVCR-null embryos immunostained with antibodies to CD71 and Ter119. The relative percentages of the nucleated cells in each of the populations I to V are indicated.

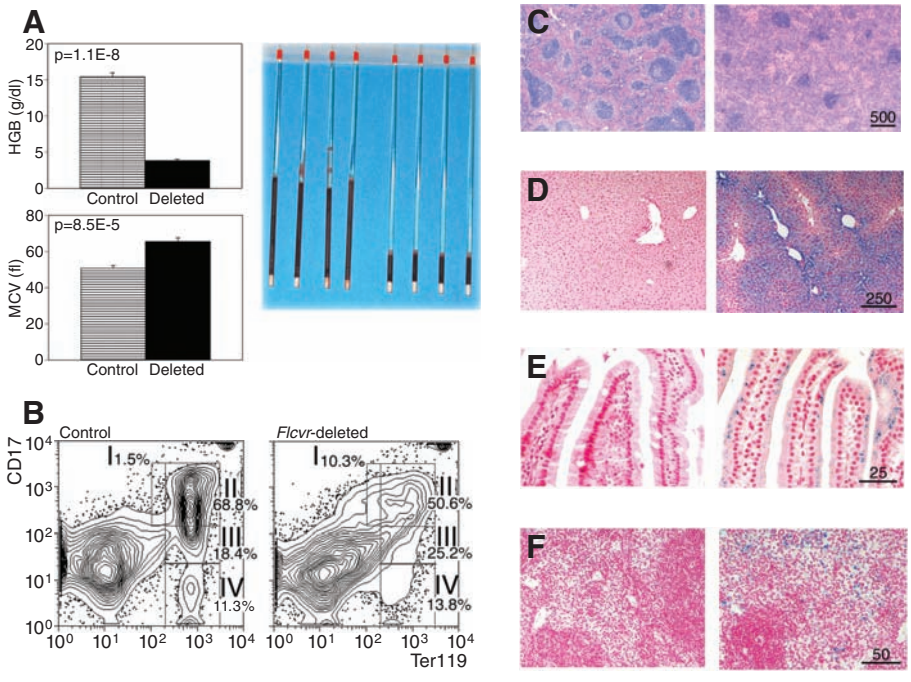


Fig. 2. Conditional deletion of *Flvcr* causes PRCA. Unless noted, data are from representative 6-week-old mice, 5 weeks post deletion, (left) controls (n = 13), (right) *Flvcr*-deleted (n = 11). (**A**) Hematologic parameters (means ± SEM, one-tailed Student's t test), hemoglobin (HGB), mean cell corpuscular volume (MCV). Duplicate spun hematocrit tubes from two control and two *Flvcr*-deleted mice. (**B**) Flow cytometric analyses of marrow from a control and *Flvcr*-deleted mouse immunostained with antibodies to CD71 and Ter119. Gating methods in Fig 1C. Ratio of the percent of cells in population IV to I and II: *Flvcr*-deleted: 49.2% ± 11.6% (n = 9) versus control: 77.1% ± 11.0% (n = 9); means ± SD, two-tailed Student's t test, $P < 10^{-4}$. The severity of block is variable between deleted animals and does not appear to correlate with the degree of anemia. (**C**) Hematoxylin-and-eosin–stained spleen sections from a control and *Flvcr*-deleted mouse. (**D** to **F**) Representative Prussian blue–stained liver sections (D) from a 6-week-old control and a *Flvcr*-deleted mouse, and duodenum (E) and spleen (F) sections from a 11-week-old (10 weeks post deletion) control and a *Flvcr*-deleted mouse. Blue staining indicates iron. Scale bars in microns.

FLVCR may serve roles during embryogenesis distinct from its critical erythropoietic function.

Although null animals die in utero, *Flvcr*$^{+/-}$ mice are clinically indistinguishable from controls (table S2); they have low mRNA expression, as anticipated, but compensate with normal FLVCR protein expression (fig. S4).

We next evaluated postnatal mice lacking FLVCR [*Flvcr*$^{flox/flox}$;*Mx-cre* (fig. S1 and Fig. 2, A to F)]. Within 4 weeks of *Flvcr* deletion, the mice are runted with pale paws. Necropsy reveals cardiomegaly and splenomegaly [*Flvcr*-deleted spleen: 326.7 mg ± 22.9 (n = 7) versus control spleen 72.9 ± 5.5 (n = 7); means ± SEM, two-tailed Student's t test, $P < 10^{-4}$], likely responses to their severe anemia.

Peripheral blood and bone marrow findings are diagnostic of PRCA. *Flvcr*-deleted mice develop a severe hyperchromic macrocytic anemia (Fig. 2A and table S3) and reticulocytopenia. Flow cytometric analyses of their bone marrow show a block in erythroid maturation at the proerythroblast stage (Fig. 2B), as do liver cells from E14.5 FLVCR-null embryos. These results are mirrored in the spleen and account for the large spleens with expanded interfollicular regions (Fig. 2C). Erythroid colony assays confirm the flow cytometry findings; CFUs-E are absent and BFUs-E (burst-forming units–erythroid) expand suboptimally [supporting online material (SOM) text], similar to results in cats viremic with FeLV-C (*5*). In addition, mice transplanted with *Flvcr*$^{flox/flox}$;*Mx-cre* bone marrow and then treated with polyinosinic-polycytidylic acid [poly(I):poly(C)] to delete *Flvcr* specifically in engrafted cells also develop PRCA (table S4). This confirms that a lack of FLVCR in hematopoietic cells (and not the microenvironment) accounts for the disease.

We then evaluated the effect of FLVCR overexpression. Pep3b (CD45.1) bone marrow was transduced with retroviral vectors, MFIG or MXIG, encoding green fluorescent protein with or without human FLVCR, respectively, and transplanted into C57BL/6 (CD45.2) mouse recipients. Twelve weeks after transplantation, the MFIG mice displayed mild hypochromic, microcytic anemia [supporting online material (SOM) text]. Because hypochromasia and microcytosis only result from heme or hemoglobin deficiency, FLVCR must export heme from differentiating erythroid cells in vivo. Because the anemia is mild, FLVCR does not outcompete globin for heme.

These observations lead us to hypothesize that FLVCR is required during definitive red cell differentiation to maintain intracellular free heme balance. In the absence of FLVCR, free heme, which

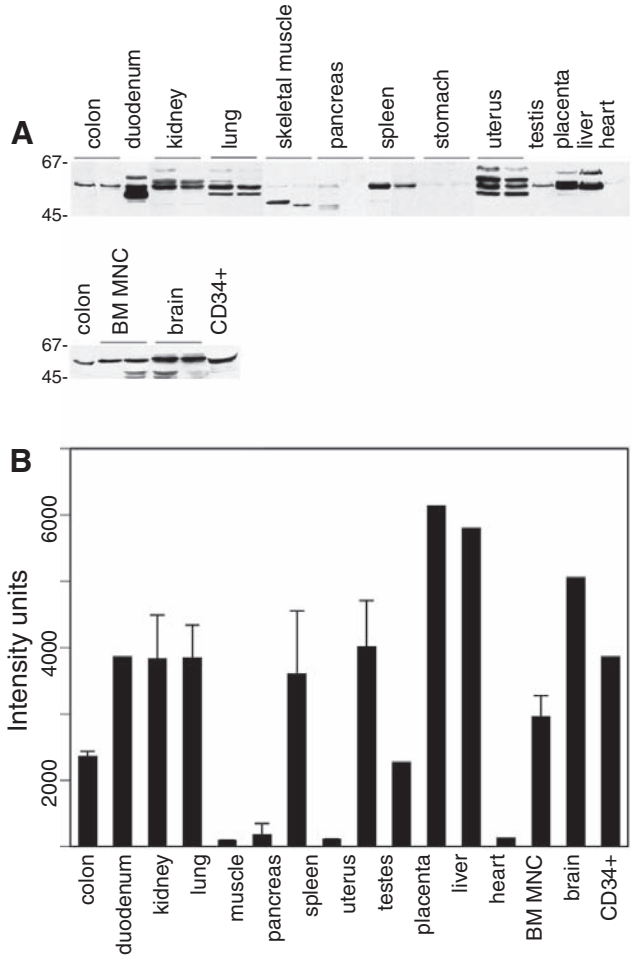


Fig. 3. FLVCR protein levels differ in human tissues. (**A**) Western blot analyses of human tissues, bone marrow mononuclear cells (BM MNC), and CD34+ stem/progenitor cells. (**B**) Densities of the 60-kB FLVCR band [shown in part (A)]. We also assayed FLVCR expression in macrophages isolated from human peripheral blood by plastic adherence for 2 hours, then cultured for 4 days with cytokines (intensity = 5214 ± 260). Quantitative RT-PCR confirmed that FLVCR expression is regulated post-transcriptionally (SOM text) (*3*, *24*).

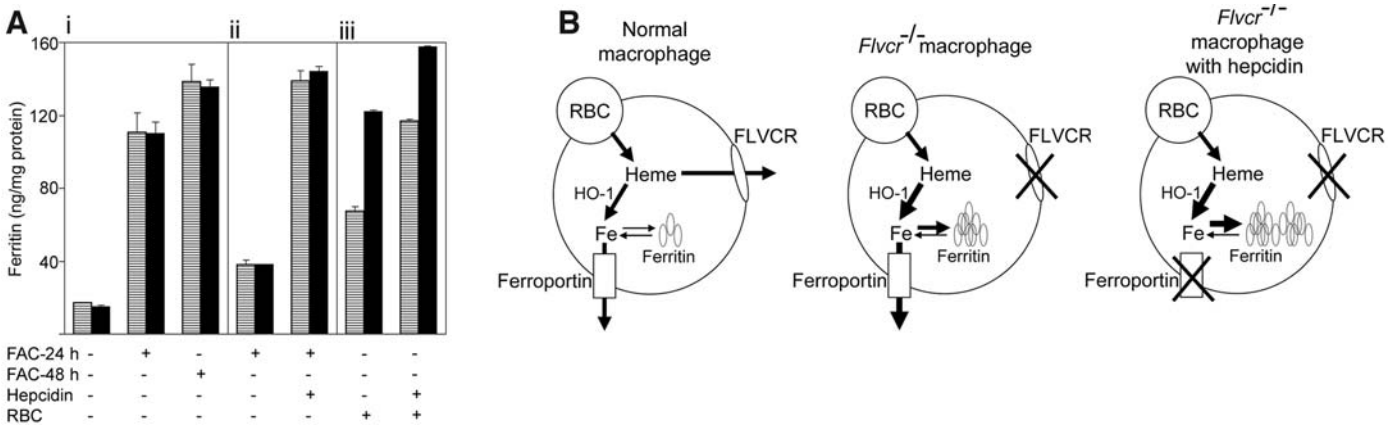


Fig. 4. FLVCR exports heme iron from macrophages. (**A**) Bone marrow–derived macrophages from control (striped) and mice in which *Flvcr* was deleted neonatally (black) were incubated in the absence or presence of FAC (10 μM Fe) for 24 or 48 hours, then washed; ferritin was measured by enzyme-linked immunosorbent assay (ELISA) (i). Cells were incubated with FAC for 24 hours (ii) or with immunoglobulin G–coated red blood cells (RBC) for 90 min (iii), washed, then incubated for an additional 24 hours with or without hepcidin (1 μg/μl) and ferritin assayed. Data represent ferritin values in macrophages derived from two control and two deleted mice ± SEM of triplicate samples per mouse. (**B**) Model of macrophage heme iron recycling. HO-1, heme oxygenase–1.

is toxic, accumulates in proerythroblasts, the stage when heme synthesis intensifies (*16*), and triggers molecular pathways that result in cell apoptosis or senescence. Although this may seem counterintuitive because red cells have high heme requirements for hemoglobin, we suspect that FLVCR functions as a safety valve to protect proerythroblasts from heme toxicity when globin expression [which is transcriptionally and translationally regulated by heme (*1*, *2*)] is insufficient. In human tissues, FLVCR is highly expressed at sites of high heme flux, including placenta, uterus, duodenum, liver, and cultured macrophages (Fig. 3), which suggests that FLVCR prevents heme toxicity or facilitates heme iron trafficking in nonerythroid cells as well.

When senescent red cells are phagocytosed and digested by macrophages, hemoglobin is degraded to heme and, subsequently, to iron, biliverdin, and carbon monoxide. Ferroportin exports iron to plasma transferrin for delivery to the marrow or liver (*17*). Hepcidin regulates this pathway by inducing the internalization and degradation of ferroportin, thereby blocking intestinal iron absorption and iron release from cellular stores and macrophages (*18*). To delineate the role of FLVCR in macrophage heme iron recycling, we exposed marrow-derived macrophages from *Flvcr*-deleted and control mice to ferric ammonium citrate (FAC) or opsonized red blood cells, in the presence or absence of hepcidin, and measured ferritin (Fig. 4A). Deleted and control macrophages exposed to FAC accumulate equivalent amounts of ferritin, which increase equivalently with hepcidin treatment. However, *Flvcr*-deleted macrophages exposed to opsonized red cells accumulate more ferritin than controls both with and without hepcidin treatment. These data support the model of macrophage heme iron recycling diagrammed in Fig. 4B; under normal physiologic conditions, heme can be exported via FLVCR or can be metabolized to iron, which is subsequently exported through ferroportin or stored as ferritin. When FLVCR is absent, the amount of iron that is generated exceeds ferroportin's export capacity, resulting in an increase in ferritin, which increases further if hepcidin is present and both heme iron and inorganic iron export is blocked. Our data confirm that not all heme in macrophages is broken down (*19*), but rather some traverses the cell intact via FLVCR. We further verified this export function by ^{55}Fe-heme and zinc mesoporphyrin studies (fig. S6).

To evaluate the role of FLVCR more broadly, we examined other tissues in *Flvcr*-deleted mice. Within 5 weeks, mice with the deletion develop pronounced iron loading in hepatocytes and subsequently within duodenal enterocytes and splenic macrophages (Fig. 2, D to F). By 7 months, there is swelling of hepatocytes lining bile canaliculi and bile stasis. In contrast, the mice in which *Flvcr* is deleted only in hematopoietic cells show no iron overload after 5 to 6 weeks (fig. S5). Liver hepcidin expression by reverse transcription polymerase chain reaction (RT-PCR) is comparably increased in mice with the deletion [1.7 ± 0.2 times control; deleted ($n = 5$), control ($n = 5$); means ± SEM; two-tailed Student's t test, $P = 0.04$] and mice lacking FLVCR only in hematopoietic cells [2.0 ± 0.3 times control; lacking FLVCR ($n = 6$), control ($n = 3$); $P = 0.03$]. These data demonstrate that hepcidin alone does not account for the iron overload and biliary pathology. One possibility consistent with our data is that FLVCR exports heme from liver into bile, thus allowing iron to exit the body.

The high hepcidin levels in *Flvcr*-deleted animals contrasts with levels in other iron-loading anemias with ineffective erythropoiesis, such as thalassemia and congenital dyserythropoietic anemia, where hepcidin is low despite high serum iron and systemic iron overload (*20*). High hepcidin levels are seen in anemic mice prevented from mounting an erythropoietic response by the use of irradiation, chemotherapy, or an antibody to erythropoietin (*21*, *22*), which indicates that erythropoietic activity is the most potent suppressor of hepcidin synthesis. Our results demonstrate that the inhibitory signal must originate from cells more differentiated than proerythroblasts and, thus, are consistent with the recent finding that growth differentiation factor GDF15 inhibits hepcidin expression (*23*).

Together, our data show that FLVCR exports heme in vivo and is required by definitive erythroid progenitors at the CFU-E–proerythroblast stage to complete terminal differentiation. We propose that heme toxicity causes PRCA in FLVCR mutant mice and FeLV-C–infected cats and may be a common pathophysiology in other models of failed erythropoiesis where heme synthesis and globin expression are dysregulated, which results in a transient excess of intracellular free heme, for example Diamond-Blackfan anemia (SOM text). Our data demonstrate that FLVCR functions in macrophage heme-iron recycling and show that systemic iron balance involves heme-iron trafficking via FLVCR, in addition to the well-described elemental iron pathways.

References and Notes

1. T. Tahara *et al.*, *J. Biol. Chem.* **279**, 5480 (2004).
2. M. Rafie-Kolpin *et al.*, *J. Biol. Chem.* **275**, 5171 (2000).
3. J. G. Quigley *et al.*, *Cell* **118**, 757 (2004).
4. N. G. Testa, D. Onions, O. Jarrett, F. Frassoni, J. F. Eliason, *Leuk. Res.* **7**, 103 (1983).
5. J. L. Abkowitz, *Blood* **77**, 1442 (1991).
6. N. Riedel, E. A. Hoover, R. E. Domsife, J. I. Mullins, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2758 (1988).
7. M. A. Rigby *et al.*, *J. Gen. Virol.* **73**, 2839 (1992).
8. J. L. Abkowitz, R. D. Holly, C. K. Grant, *J. Clin. Invest.* **80**, 1056 (1987).
9. Materials and methods are available as supporting material on *Science* Online.
10. S. Watanabe, R. Akagi, M. Mori, T. Tsuchiya, S. Sassa, *Placenta* **25**, 387 (2004).
11. H. Wu, X. Liu, R. Jaenisch, H. F. Lodish, *Cell* **83**, 59 (1995).
12. J. Zhang, M. Socolovsky, A. W. Gross, H. F. Lodish, *Blood* **102**, 3938 (2003).
13. I. A. Cathie, *Arch. Dis. Child.* **25**, 313 (1950).
14. T. N. Willig *et al.*, *Pediatr. Res.* **46**, 553 (1999).
15. V. E. Wang, T. Schmidt, J. Chen, P. A. Sharp, D. Tantin, *Mol. Cell. Biol.* **24**, 1022 (2004).
16. A. Wickrema, S. B. Krantz, J. C. Winkelmann, M. C. Bondurant, *Blood* **80**, 1940 (1992).
17. M. W. Hentze, M. U. Muckenthaler, N. C. Andrews, *Cell* **117**, 285 (2004).
18. E. Nemeth *et al.*, *Science* **306**, 2090 (2004).
19. M. D. Knutson, M. Oukka, L. M. Koss, F. Aydemir, M. Wessling-Resnick, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1324 (2005).
20. G. Papanikolaou *et al.*, *Blood* **105**, 4103 (2005).
21. M. Vokurka, J. Krijt, K. Sulc, E. Necas, *Physiol. Res.* **55**, 667 (2006).
22. M. Pak, M. A. Lopez, V. Gabayan, T. Ganz, S. Rivera, *Blood* **108**, 3730 (2006).
23. T. Tanno *et al.*, *Nat. Med.* **13**, 1096 (2007).
24. C. S. Tailor, B. J. Willett, D. Kabat, *J. Virol.* **73**, 6500 (1999).

25. We thank J. Hicks (technical support), Fred Hutchinson Cancer Research Center Experimental Histopathology and University of Washington Medical Center, Clinical Research Center. This work is supported by NIH grants and the Henigson Research Fund.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5864/825/DC1
Materials and Methods
SOM Text
Figs. S1 and S6
Tables S1 to S4
References

1 October 2007; accepted 21 December 2007
10.1126/science.1151133

New Products Focus: Imaging

Multi-Application Imaging and Analysis

The G:BOX Chemi XT16 automated chemiluminescence and fluorescence imaging and analysis system features many innovations. Inside a newly designed light-tight darkroom, the latest 16-bit camera with a new f 0.95 variable aperture lens makes it possible to quickly produce accurate images of large gels and blots. The G:BOX Chemi XT16 offers computer control of its motor-driven stage, which allows the system to memorize set positions for specific applications. The camera, with sensitive effective 6.3-megapixel resolution, is ultracooled to guarantee the instrument can separate close band and spot images with virtually no background noise, even when imaging chemiluminescence blots for long exposure times. In addition, the darkroom comes with a state-of-the-art overhead white epi-lighting module, an ultraviolet to visible light NovaGlo converter screen, and a 20-cm-by-20-cm ultraviolet transilluminator to allow rapid and accurate imaging of common DNA and protein stains such as ethidium bromide, SYBR Green, SYBR Safe, Coomassie blue, Deep Purple, Pro-Q Diamond, silver stain, and SYPRO Ruby. For additional applications, it can be fitted with other lighting options, precision filters, and a blue light converter screen.

Syngene
For information 800-686-4407
www.syngene.com

Tissue Section Scanning

The Acumen eX3 is well suited to scanning tissue sections mounted on microscope slides due to its large field of view (400 mm^2), which enables the rapid collection of multicolored fluorescence data from whole sections of tissue without the need to stitch together multiple images. The Acumen eX3 provides several modes of use. Cytometry mode rapidly scans, analyzes, and reports high-content information yielding very small postscan file sizes (>50kb). For more detailed evaluations, TIFF files can be exported and analyzed in batch mode using third-party image processing software. Cytometry mode can also be used to rapidly determine regions of interest within larger sections for visual scoring.

TTP LabTech
For information +44 1763 262626
www.ttplabtech.com

3D Chemical Imaging with AFM

The alpha500 and alpha700 microscopy series are for automated confocal Raman imaging and atomic force microscopy (AFM) on large samples. The alpha500 and alpha700 are the first instruments on the market to combine confocal Raman microscopy for three-dimensional chemical imaging and AFM for high-resolution structural imaging in an automated system. A motorized sample stage with a travel range of 150 by 100 mm for the alpha500 and 350 by 300 mm for the alpha700 allows multiarea/multipoint measurements or overview scans on a user-defined number of measurement points. Automated functions such as integrated autofocus and an automatic AFM-tip approach guarantee that standardized routine measurement procedures or manually designed sequences can be performed without any ongoing process control by an operator. The instruments significantly reduce the overall experiment duration and deliver a greater amount of data in a given time to minimize resource use in routine research and deliver high-level quality control.

WITec
For information +49 (0) 731 140 70-0
www.witec.de

High-Speed Imaging for Live Cell Research

Speed is often the deciding factor for successful imaging in high-resolution documentation of living cells, molecular processes, and rapidly fading fluorescence specimens. The monochrome digital camera Leica DFC360 FX is designed to produce brilliant images at maximum temporal resolution. Thanks to state-of-the-art charge-coupled device technology, the new camera system achieves frame rates of 20 fps for full frame and more than 100 fps in binning mode. The sensitive sensor and active Peltier cooling ensure a high dynamic range even for low-light intensities. With shutter speeds of 4 µs to 10 minutes and up to 10-fold signal amplification, the instrument offers maximum flexibility.

Leica
For information +49 (0) 6441/29-2550
www.leica-microsystems.com

Ultraviolet Area Imaging Detector

The ActiPix D100 is a miniature quantitative ultraviolet area imaging detector. It opens possibilities not available with conventional detectors, including new measurement capabilities in applications including real-time study of diffusion processes, inline quantification and sizing of biopharmaceuticals, dissolution and solubility testing, and membrane transport studies. The ActiPix D100 consists of a control box connected via a fiber optic cable and communications cable to a remote sensor head. The sensor head holds easily exchangeable, application-specific cartridges for techniques including capillary electrophoresis, nanoliquid chromatography, and imaging of lab-on-a-chip devices. The miniature detector head contains a high-resolution 1,280-by-1,024 active pixel sensor. The detector can be used as a plug-and-play accessory linked to multiple peripheral devices, such as capillary electrophoresis or nanoliquid chromatography instrumentation, with or without a mass spectrometer. Detection is performed at a selected wavelength by means of interchangeable filters. Processed data, including absorbance values covering the whole imaged area, is output in real-time to a computer using a high-speed serial data link.

Paraytec
For information +44 1904 526270
www.paraytec.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.