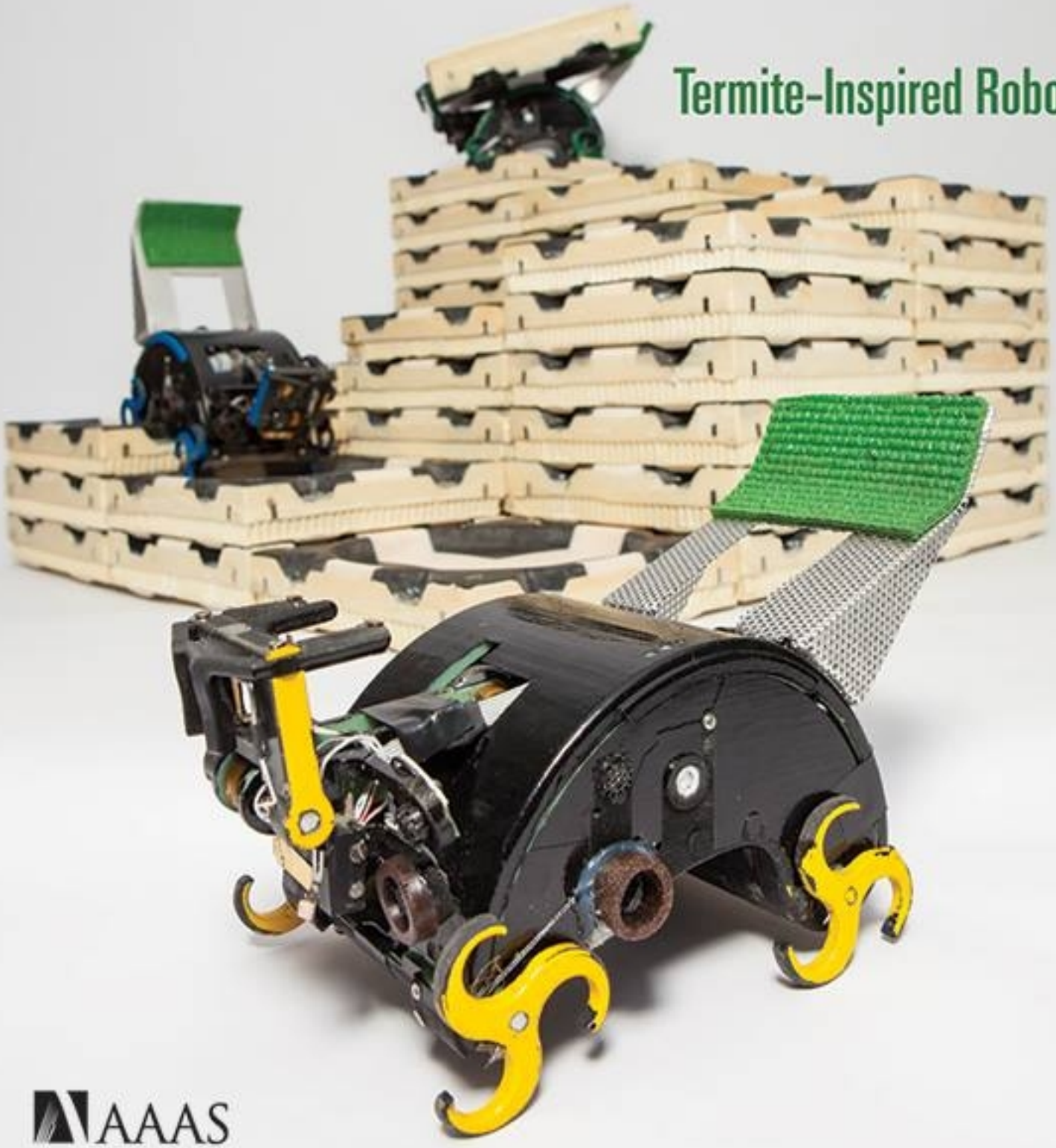
14 February 2014 | $10

Science

Termite-Inspired Robots



AAAS

EDITORIAL

NEWS OF THE WEEK

NEWS & ANALYSIS

NEWS FOCUS

LETTERS

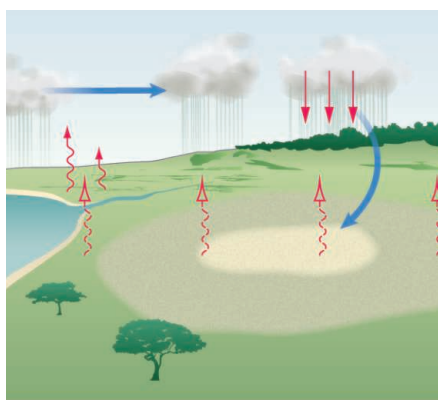
BOOKS *ET AL.*

POLICY FORUM

PERSPECTIVES

CONTENTS continued >>

page 722

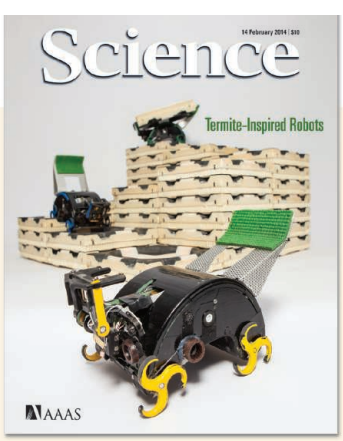


page 737

ON THE WEB THIS WEEK

>> *Science* Podcast
Listen to stories on termite-inspired robots, cells with many, many genomes, and a roundup of stories from our daily news site.

>> Find More Online
Check out the latest in a series of Perspectives on Challenges in Climate Science at **www.sciencemag.org/extra/climate**.



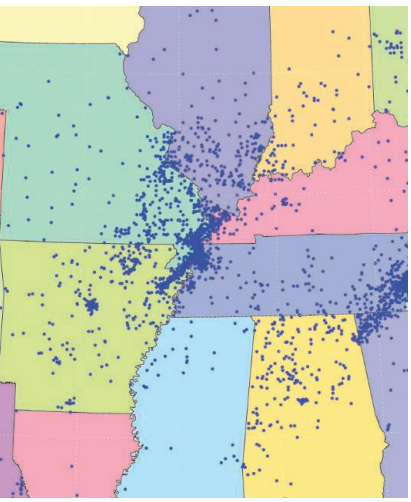
COVER

A multirobot construction system inspired by mound-building termites. Independent climbing robots with onboard sensors automatically build user-specified structures out of specialized brick-sized building material. The robots are limited to local sensing and coordinate their activity indirectly by manipulating their shared environment and reacting to what they encounter. See pages 742 and 754, as well as supplementary movies online at www.sciencemag.org/content/343/6172/754/suppl/DC1.

Photo: Eliza Grinnell, Harvard School of Engineering and Applied Sciences

DEPARTMENTS

Published by AAAS



page 762

page 788

RESEARCH ARTICLES

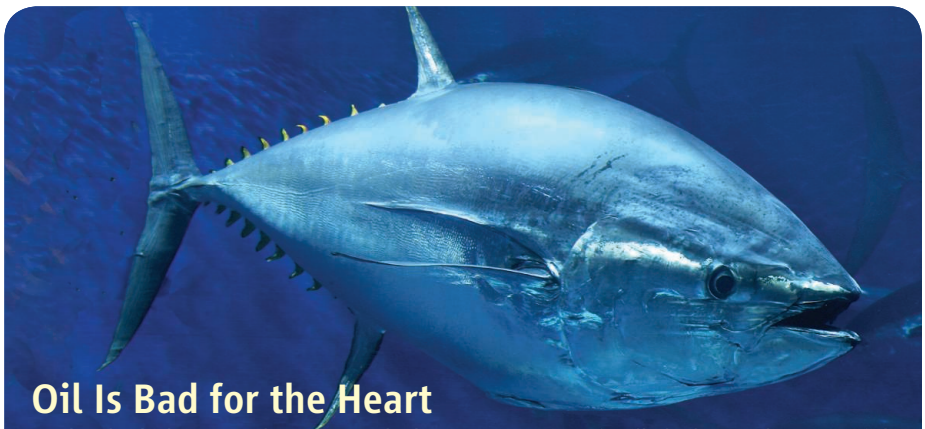
REPORTS

CREDIT: (BOTTOM) VINCENT MONCORGÉ

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): $149 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $990; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**

Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $10.00 current issue, $15.00 back issue prepaid includes surface postage; bulk rates on request. **Authorization to photocopy** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $30.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

EDITED BY STELLA HURTLEY



Oil Is Bad for the Heart

Crude oil, which commonly makes up the largest proportion of spilled oil, is cardiotoxic to fish embryos. To better understand the processes involved, **Brette *et al.*** (p. 772) exposed captive young tuna to oil samples collected from the Deep Water Horizon spill site to determine the mode of cardiotoxicity. Crude oil prolonged the action potential of cardiomyocytes and disrupted the excitation-contraction coupling in these cells, functionally disrupting cellular excitability and creating the potential for cardiac arrhythmias. Such cardiac impacts may be more broadly distributed in vertebrates exposed to crude oil.

The in-Laws Through History

Admixture, the result of previously distant populations meeting and breeding, leaves a genetic signal within the descendants' genomes. However, over time the signal decays and can be hard to trace. **Hellenthal *et al.*** (p. 747) describe a method, using a technique called chromosome painting, to follow the genetic traces of admixture back to the nearest extant population. The approach revealed details of worldwide human admixture history over the past 4000 years.

Losses and Gains

In order to better understand the process by which de novo genes originate, **Zhao *et al.*** (p. 769, published online 23 January) examined testis-based gene expression among *Drosophila melanogaster* strains and identified both fixed and polymorphic de novo genes. The results suggest that spontaneous activation of previously noncoding DNA may be an important factor in generating genetic novelty.

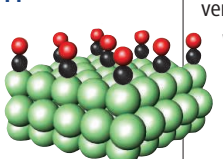
On the Fast Track

Membranes based on graphene can simultaneously block the passage of very small molecules while allowing the rapid permeation of water. **Joshi *et al.*** (p. 752; see the Perspective by **Mi**) investigated the permeation of ions and neutral molecules through a graphene oxide (GO) membrane in an aqueous solution. Small ions, with hydrated radii smaller than 0.45 nanometers, permeated through the GO membrane several orders of magnitude faster than predicted, based on diffusion theory. Molecular dynamics simulations revealed that the GO membrane can attract a high concentration of small ions into the membrane, which may explain the fast ion transport.

Robot Rules

In the case of mound-building termites, colonies comprising thousands of independently behaving insects build intricate structures, orders of magnitude larger than themselves, using indirect communication methods. In this process, known as stigmergy, local cues in the structure itself help to direct the workers. **Werfel *et al.*** (p. 754; see the Perspective by **Korb**) wanted to construct complex predetermined structures using autonomous robots. A successful system was designed so that for a given final structure, the robots followed basic rules or "structpaths" in order to complete the task.

Speeding Up Surface Diffraction



Surface diffraction methods can determine the atomic structure of the topmost layer of a crystal and also subsurface structures. However, many surface diffraction methods either require ultrahigh vacuum conditions, which limits the reaction conditions that can be studied, or require long data acquisition times, which limits temporal resolution. Using high x-ray energies, **Gustafson *et al.*** (p. 758, published online 30 January; see the Perspective by **Nicklin**) were able to measure the intensities of surface-diffracted beams to follow the surface oxidation that accompanies the changes in a palladium surface during the catalytic oxidation of CO with O_2.

Rolling Under New Madrid

During 1811–1812, the New Madrid Seismic Zone experienced a sequence of three large intraplate earthquakes and at least one comparably sized aftershock. There have been no earthquakes of similar magnitudes since then. Using a combination of historical data dating back to the original large events and an epidemic-type aftershock sequence model, **Page and Hough** (p. 762, published online 23 January) found that the current low seismicity is not part of an aftershock sequence. Instead, despite low observable deformation rates, there is ongoing accumulation of strain, leaving the potential for large earthquakes in the region.

Keeping Alphaviruses Under Wraps

Viruses mutate to avoid detection, and the host responds in kind. For example, 2′-O methylation of the 5′ cap of viral RNA allows viruses to escape detection by the interferon-stimulated host defense protein, IFIT1. Alphaviruses, however, lack this modification but are able to remain undetected in the presence of IFIT1. How? Using a combination of viral mutants and biochemical analysis, **Hyde *et al.*** (p. 783, published online 30 January) found that alphaviruses contain secondary structural motifs in the 5′ untranslated region of their genomic RNA that allow them to avoid detection by IFIT1. When these regions were rendered nonfunctional, IFIT1 was able to keep the virus under control.

Limb Regeneration

Flatworms possess pluripotent stem cells that can regenerate any cell type in the body, whereas vertebrates mobilize committed progenitor cells whose fate is predetermined. Investigating limb regeneration in a crustacean, **Konstantinides and Averof** (p. 788, published online 2 January) found that arthropods use committed progenitor cells to regenerate missing tissues, including satellite-like cells to regenerate muscle. The study reveals similarities between arthropod and vertebrate muscle regeneration, pointing to a common basis for muscle regeneration that may date back to the common ancestors of all bilaterian animals.

Published by AAAS

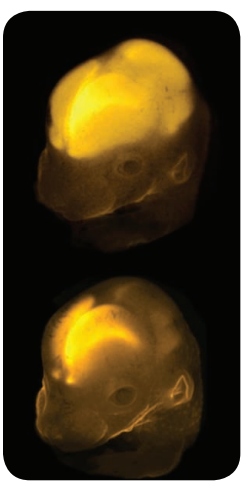
Additional summaries

Toddler Welcome

It has been assumed that most, if not all, major signals that control vertebrate embryogenesis have been identified. Using genomics, **Pauli *et al.*** (p. 746, published online 9 January) have now identified several new candidate signals expressed during early zebrafish development. One of these signals, Toddler, is a short, conserved, and secreted peptide that promotes the movement of cells during zebrafish gastrulation. Toddler signals through G protein–coupled receptors to drive internalization of the Apelin receptor, and activation of Apelin signaling can rescue *toddler* mutants.

Fine-Tuning Brain Gyrations

A handful of patients who suffer from seizures and mild intellectual disability have now led the way to insights about how one piece of regulatory DNA controls development of a section of the human cortex. Imaging the brains of these patients, **Bae *et al.*** (p. 764; see the Perspective by **Rash and Rakic**) observed malformations on the surface folds in a brain region that includes "Broca's area," the main region underlying language. The three affected families shared a 15–base pair deletion in the regulatory region of a gene, *GPR56*, which encodes a G protein–coupled receptor required for normal cortical development that is expressed in cortical progenitor cells.



CREDIT: BYOUNG-IL BAE

Developmental Complexity

Although related, the plants *Arabidopsis thaliana* and *Cardamine hirsuta* have different sorts of leaves—one, a rather plain oval and the other, a complicated multipart construction. Comparing the development of the two leaf types, **Vlad *et al.*** (p. 780) uncovered a gene that regulates developmental growth. The *C. hirsuta* gene encoding the REDUCED COMPLEXITY (RCO) homeodomain protein arose through gene duplication and neofunctionalization, but was lost in the *A. thaliana* lineage. In *C. hirsuta*, RCO suppresses growth in domains around the perimeter of the developing leaf, yielding complex-shaped leaves. *A. thaliana*, lacking RCO, produces simple leaves. When RCO was expressed in *A. thaliana*, the leaves became more complex. Thus, the capacity to produce complex leaves remains, despite loss of the initiator.

Introducing MARS-Seq

Immune cells are typically differentiated by surface markers; however, this designation is somewhat crude and does not allow for fine distinctions that might be characterized by their RNA transcripts. **Jaitin *et al.*** (p. 776) used massively parallel single-cell RNA-sequencing (MARS-Seq) analysis to explore cellular heterogeneity within the immune system by assembling an automated experimental platform that enables RNA profiling of cells sorted from tissues using flow cytometry. More than 1000 cells could be sequenced, and unsupervised clustering analysis of the RNA profiles revealed distinct cellular groupings that corresponded to B cells, macrophages, and dendritic cells. This approach provides the ability to perform a bottom-up characterization of in vivo cell-type landscapes independent of cell markers or prior knowledge.

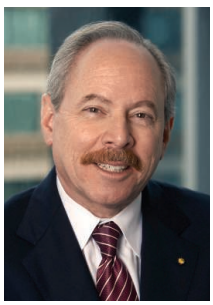
Cell-Cell Interactions in Development

In vertebrate embryos, the number, size, and positional identity of mesodermal segments (somites) located bilaterally along the anterior-posterior axis is widely believed to be controlled by a molecular clock of oscillating gene expression interacting with a traveling wave of signals to determine how many cells make up a somite. **Dias *et al.*** (p. 791, published online 9 January; see the Perspective by **Kondo**) reveal that it is possible to generate somites of normal size, shape, and identity without either a clock or a wavefront. Instead, the findings suggest that somite size and shape are regulated by local cell-cell interactions.

Folding When Wet

Most globular proteins release water as they fold to form a dry hydrophobic core. In contrast, **Sun *et al.*** (p. 795; see the Perspective by **Sharp**) report a high-resolution structure showing that the antifreeze protein Maxi retains about 400 water molecules in its core. Maxi is a dimer in which two helical monomers each bend in the middle to form a four-helix bundle. The helices are spaced slightly apart to accommodate two intersecting polypentagonal monolayers of water. The pentagons form cages around inward pointing side chains to stabilize the structure. The ordered waters extend to the protein surface where they are likely to be involved in ice binding.

Marcia McNutt is Editor-in-Chief of *Science*.

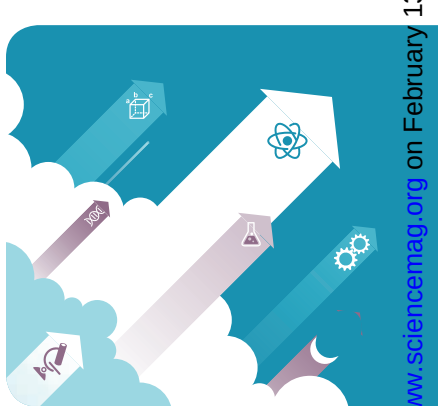


Alan I. Leshner is Chief Executive Officer of AAAS and Executive Publisher of *Science*.

Science Advances

THE MISSION OF THE NONPROFIT AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE (AAAS), the publisher of *Science*, is to advance science for the benefit of all humankind. *Science* contributes to that mission by communicating the very best research across the full range of scientific fields to an extremely broad international audience. The research enterprise has grown dramatically in the past few decades in the number of high-quality practitioners and results, but the capacity for *Science* to accommodate those works in our journal has not kept pace. Its editors turn away papers that are potentially important, well written, of broad interest, and technically well executed. Although other journals provide publishing venues for more papers, many authors still desire to be published in *Science*, a journal known for its selectivity, high standards, rapid publication, and high visibility.

To help meet this need, as well as expand the current content of *Science* so as to include even more diverse topics in science, engineering, technology, mathematics, and the social sciences, AAAS will be launching, in early 2015, a digital-only journal, *Science Advances*. Like *Science*, this new publication is designed to encourage transformative research and serve a wide readership. Our view at AAAS is that science is becoming more integrated and interdisciplinary, and therefore we prefer to provide one additional broad journal rather than a number of disciplinary titles, each with more limited scope, that would all have to be searched to find the papers of most interest. Also like *Science*, this new journal will aim for rapid publication. To contain costs for *Science Advances*, the journal will publish original research and review articles only, although select papers in *Science Advances* may be highlighted in *Science* through News and Commentary coverage.



To ensure the greatest accessibility for authors and readers, the new journal will be open access, with publication funded through author processing charges. With this publishing model, the number of papers that can be published is limited only by the quality of submissions. With digital-only publication, all papers will be posted as soon as they are ready for publication. Although *Science* has long participated in various global efforts to give researchers in the world's poorest countries free access to peer-reviewed results, *Science Advances* will allow AAAS to serve a much larger community of Internet-connected scientists who desire to keep current with the latest scientific results.

The editorial model for *Science Advances* will be similar to what is used for many society journals. A lead editor will be supported by a large number of associate editors who will be eminent active scientists. Administrative work will be handled at *Science* headquarters to efficiently process reminders to reviewers and keep manuscripts moving briskly through to a decision. Papers favorably reviewed at *Science*, *Science Translational Medicine,* or *Science Signaling* but declined for lack of space can be considered automatically for publication in *Science Advances*. This can occur through a cascading process without further review or further effort on the part of the authors. The goal is to speed publication, alleviate the burden on the reviewer community, and reduce the risk to authors of having to resubmit elsewhere.

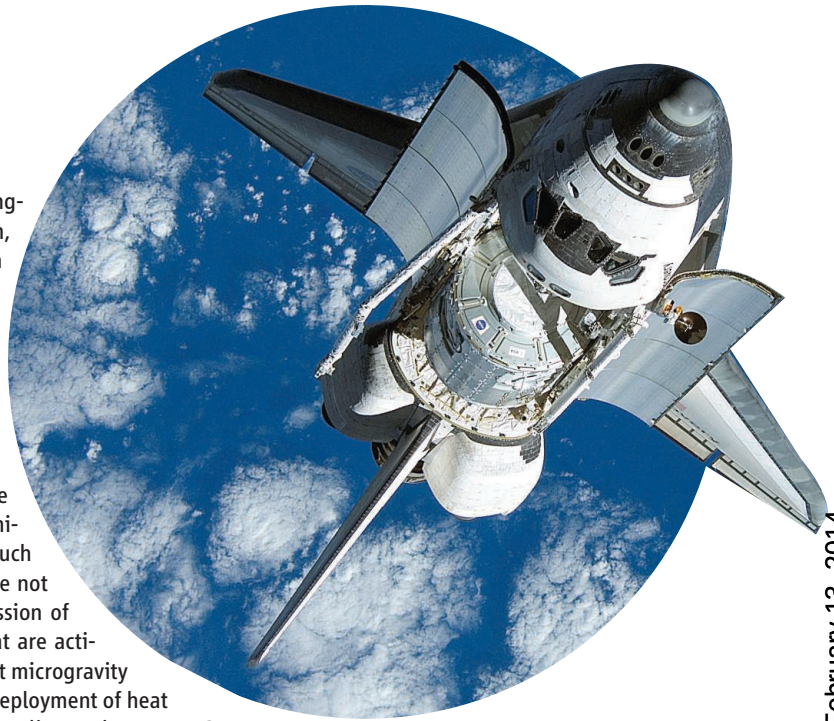
Science Advances will distinguish itself from *Science* by its editorial model, the immediate access to papers for all readers, and the fact that acceptance for publication is limited only by the quality of the paper. In the coming months, we look forward to recruiting the lead editor, associate editors, and inaugural papers that will launch a new resource of high-quality research for the scientific community.

– Marcia McNutt and Alan I. Leshner

10.1126/science.1251654

Published by AAAS

EDITED BY KRISTEN MUELLER AND JESSE SMITH



IMMUNOLOGY

Sick Space Flies

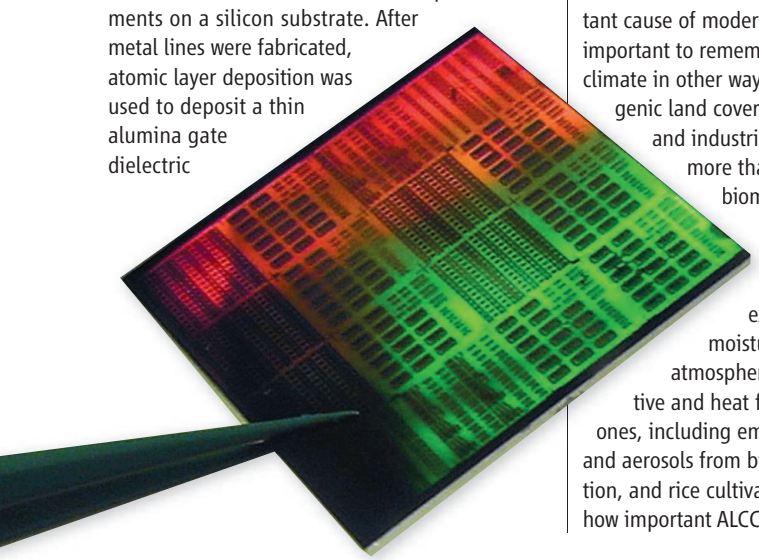
A concern regarding manned long-term space missions is that changes in gravitational force compromise the human immune system, but the underlying cellular and molecular reasons have not been clear. Taylor *et al.* studied innate immunity in *Drosophila melanogaster* that traveled aboard Space Shuttle Discovery in 2006. Flies reared in space were compared to flies that underwent development on Earth. Upon the return of the space-reared flies to Earth, both groups of flies were subjected to bacterial (*Escherichia coli*) or fungal (*Beauveria bassiana*) infections, and their gene expression profiles were examined. Genes associated with Toll receptor–mediated immune responses to fungal infection were activated only in the Earth flies. The expression of specific antimicrobial peptides also failed in the space flies. Other mechanisms, such as the Imd signaling pathway response to bacterial infection, were not affected in space flies. The space flies exhibited increased expression of heat shock response genes, a subset of stress response genes that are activated to manage aberrant protein folding. The authors suggest that microgravity may alter the folding and stability of proteins, which triggers the deployment of heat shock proteins that in turn, may interfere with the Toll receptor signaling pathway. — LC

PLOS One **9**, e86485 (2014).

ENGINEERING

Growing Graphene Receivers

One use of the extremely high conductivity of graphene is in radio-frequency (RF) applications. However, the devices they use require specialized fabrication methods to avoid damaging the active graphene channel layer. Conventional fabrication of practical RF devices starts by placing the channel material on a silicon substrate and then fabricating other passive device components on top of it, using deposition steps that can involve high temperatures that can degrade the device performance. Han *et al.* report on a fabrication scheme for a graphene RF receiver that first assembles the passive elements on a silicon substrate. After metal lines were fabricated, atomic layer deposition was used to deposit a thin alumina gate dielectric layer, and the active graphene was grown through a chemical vapor deposition process. None of the processing steps required temperatures in excess of 400°C. The final device, which contains three GFETs transistors, four inductors, two capacitors, and two resistors in a 0.6-mm-square area, operates at 4.3 gigahertz and could receive and restore digital text transmitted with that carrier frequency. — PDS



Nat. Commun. 10.1038/ncomms4086 (2014).

ATMOSPHERIC SCIENCE

Early Intervention

Anthropogenic carbon dioxide emissions from the use of fossil fuels may be the most important cause of modern global warming, but it is important to remember that humans can affect climate in other ways, such as through anthropogenic land cover change (ALCC). Agriculture and industrial activities have modified more than half of Earth's natural biomes, and ALCC has influenced global climate both through biogeophysical feedbacks, such as modification of the exchange of momentum and moisture between the land and the atmosphere and the alteration of radiative and heat fluxes; and biogeochemical ones, including emissions of greenhouse gases and aerosols from biomass burning, deforestation, and rice cultivation. He *et al.* investigate how important ALCC has been in the past, by using a climate model forced by recently compiled observational data to assess how ALCC affected climate over the preindustrial Holocene. They found that ALCC increased global temperatures by around 0.73°C in that interval, an amount comparable to the ~0.8°C warming that has occurred during industrial times. So it seems that early anthropogenic activity had a significant impact on climate thousands of years before the Industrial Revolution began, mostly as a result of the greenhouse gas emissions caused by activities related to farming, such as deforestation and rice cultivation. — HJS

Geophy. Res. Lett. **41**, 10.1002/2013GL058085 (2014).

PLANT SCIENCE

Predicting the Next Generation

Heterosis, in which the hybrid offspring perform better than the inbred parental lines, is a valuable but unpredictable aspect of maize cultivation. Traits that are useful in agricultural settings are often the outcome of complex genetic interactions, with many genes influencing each other and developmental outcomes in small ways. As a result, the genes controlling useful traits are often unknown. Nonetheless, crop breeders use what information they can find to generate more productive maize lines. Feher *et al.* have now used metabolites, the downstream output of complex gene suites, to predict heterosis. Looking at the early development of the

Published by AAAS

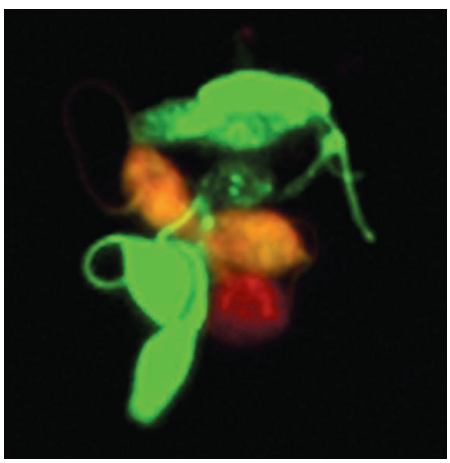
seedling's primary root (a well-functioning root gets the plant off to a good start) as their end point, the authors compared parental metabolite profiles to those of the hybrid offspring. A subset of the metabolites was identified as predictive of hybrid outcomes not only for that same metabolite but also for several other metabolites. The most effective predictions of hybrid root biomass were achieved by looking at only 5, but any 5, of the most predictive 10 to 20 metabolites. — PJH

PLOS One **9**, e85435 (2014).

MICROBIOLOGY

Ancestor Intercourse

Trypanosomes (notably including the sleeping sickness parasites) have long been thought to be



primitive protist oddities with strange biochemistries. Recent evidence from Peacock *et al.* shows that, just like the majority of eukaryotes, trypanosomes have sex. Starting from observations on the expression of meiosis-specific genes in trypanosomes within the salivary glands of the tsetse fly vector, distinctively shaped cells—putative gametes—were found. Subsequently, the cells were observed to intertwine flagella, squirm, and form intimate pairs. Labeling with different-colored fluorescent proteins revealed that membrane and cytoplasmic fusion occurred (although formal proof is still required for nuclear and kinetoplastid DNA exchange), hence confirming that even the most ancestral eukaryotes indulge in sexual reproduction. — CA

Curr. Biol. **24,** 181 (2014).

PHYSICS

A Semisynthetic Lattice

Atomic vapors at very low temperatures are useful for the quantum simulation of solid- state systems, because their properties can be finely controlled and tuned. These neutral atoms are not, however, completely analogous to the charged carriers in solids; for instance, an external magnetic field causes electrons to move in circular orbits but has no such effects on neutral atoms. Celi *et al.* propose a simple method for creating a uniform magnetic flux in a one-dimensional (1D) optical lattice that, if realized, might be used to observe exotic phenomena such as Hofstadter-butterfly–like fractal spectra or the dynamics of topological edge states. The method is based on synthetically extending the 1D lattice into the second dimension of internal atomic states (spin) by coupling those states using a pair of Raman laser beams that are directed at an angle with respect to the optical lattice; the required amount of the Raman laser light is substantially smaller than in existing schemes. The resulting band structure supports edge states in the spin variable whose dynamics should be observable through spin-sensitive density measurements. — JS

Phys. Rev. Lett. **112**, 043001 (2014).

POLICY

Cooperating on Climate

We've come to expect lack of progress at the annual United Nations climate talks. A key obstacle to agreement is the wealth inequality among the countries around the negotiating table. Such public-goods negotiations, and the exploitation of common resources, are tricky enough on their own, but addressing the gap between "haves" and "have nots" adds another level of difficulty. Building on laboratory experiments and earlier theoretical work, Vasconcelos *et al.* use Evolutionary Game Theory models to explore how wealth inequality and risk perception affect such negotiations, and address another key element, the homophily of parties; i.e., their tendency to align with others from the same wealth level. They found that if parties were willing to cooperate regardless of wealth levels, then some inequality among parties could actually lead to better cooperation, as the rich tend to contribute more and compensate for lower contributions from the poor. Contributions from the poor are still critical, though, and increased homophily, with limited cooperation across the wealth gap, can lead to collapse. Obstinately cooperative behavior, with a few poor countries cooperating with wealthier countries, can compensate for broader homophily. In addition to minimizing homophilic biases, the authors suggest that negotiations be portioned into smaller groups focused on local short-term targets for which uncertainty is relatively limited. — BW

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.1323479111 (2014).

CREDIT: L. PEACOCK *ET AL.*, *CURRENT BIOLOGY* **24**, 2 (2 JANUARY 2014) © 2014 ELSEVIER INC.

Published by AAAS

www.sciencemag.org

Science

1200 New York Avenue, NW
Washington, DC 20005
Editorial: 202-326-6550, FAX 202-289-7562
News: 202-326-6591, 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 202-326-6765, permissions@aaas.org

Member Benefits AAAS Travels: Betchart Expeditions 800-252-4910; Apple Store www.store.apple.com/us/go/eppstore/aaas; NASA Federal, 1-888-NASA-FCU (1-888-627-2328) or www.nasafcu.com; Cold Spring Harbor Laboratory Press Publications www.cshlpress.com/affiliates/aaas.htm; Hertz 800-654-2200 CDP#343457; Seabury & Smith Life Insurance 800-424-9883; Subaru VIP Program 202-326-6417; Nationwide Insurance http://nationwide.com/aaas; Other Benefits: AAAS Member Services 202-326-6417 or http://membercentral.aaas.org/discounts.

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

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

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

Information for Authors
See pages 680 and 681 of the 7 February 2014 issue or access www.sciencemag.org/about/authors

EDITOR-IN-CHIEF **Marcia McNutt**
EXECUTIVE EDITOR **Monica M. Bradford** NEWS EDITOR **Tim Appenzeller**
MANAGING EDITOR, RESEARCH JOURNALS **Katrina L. Kelner**
DEPUTY EDITORS **Barbara R. Jasny, Andrew M. Sugden, Valda J. Vinson, Jake S. Yeston**

Editorial senior editors/commentary Lisa D. Chong, Brad Wible; **senior editors** Gilbert J. Chin, Pamela J. Hines, Paula A. Kiberstis (Boston), Marc S. Lavine (Toronto), Kristen L. Mueller, Beverly A. Purnell, L. Bryan Ray, Guy Riddihough, H. Jesse Smith, Phillip D. Szuromi (Tennessee), Laura M. Zahn (San Diego); **associate editors** Melissa R. McCartney (Education Projects), Margaret M. Moerchen, Jelena Stajic, Sacha Vignieri (Oregon), Nicholas S. Wigginton (Michigan); **book review editor** Sherman J. Suter; **associate letters editor** Jennifer Sills; **editorial manager** Cara Tate; **senior copy editors** Jeffrey E. Cook, Chris Filiatreau, Cynthia Howe, Harry Jach, Lauren Kmec, Barbara P. Ordway, Trista Wagoner; **senior editorial coordinators** Carolyn Kyle, Beverly Shields; **editorial coordinators** RamatoulayeDiop, Joi S. Granger, Lisa Johnson, Anita Wynn; **publications assistants**, AneeraDobbins, Jeffrey Hearn, Dona Mathieu, Le-Toya Mayne Flood, Shannon McMahon, Scott Miller, Jerry Richardson, Brian White; **editorial assistant** Patricia M. Moore; **executive assistant** Alison Crawford; **administrative support** Maryrose Madrid

Senior web editor Sarah Crespi; **web development manager** Martyn Green **News managing editor** John Travis; **international news editor** Richard Stone; **deputy news editors** Robert Coontz, Elizabeth Culotta, David Grimm (Online), David Malakoff, Leslie Roberts; **senior correspondents** Eliot Marshall, Jeffrey Mervis, Elizabeth Pennisi **news writers** Yudhijit Bhattacharjee, Adrian Cho, Jennifer Couzin-Frankel, Carolyn Gramling, Jocelyn Kaiser, Richard A. Kerr, Robert F. Service (Pacific NW), Kelly Servick, Erik Stokstad, Emily Underwood; **web developer** Daniel Berger; **social media strategist** Meghna Sachdev; **interns** Thomas Sumner, Nadia Whitehead; **contributing correspondents** John Bohannon, Jon Cohen (San Diego, CA), Ann Gibbons, Sam Kean, Eli Kintisch, Andrew Lawler, Mitch Leslie, Charles C. Mann, Virginia Morell, Heather Pringle; **copy editors** Kara Estelle, Nora Kelly, Jennifer Levin; **administrative support** Scherraine Mack; **bureaus** San Diego, CA: 760-942-3252, FAX 760-942-4979; Pacific Northwest: 503-963-1940

Production director Wendy K. Shank; **assistant manager** Rebecca Doshi; **senior specialists** Steve Forrester, Anthony Rosen; **specialist** Jacob Hedrick **Preflight director** David M. Tompkins; **manager** Marcus Spiegler; **specialists** Jason Hillman, Tara Kelly

Art director Yael Fitzpatrick; **associate art director** Laura Creveling; **senior illustrators** Chris Bickel, Katharine Sutliff; **illustrator** Valerie Altounian; **senior art associates** Holly Bishop, Preston Huey; **art associates** Kay Engman, Garvin Grullón, Chrystal Smith; **photo editor** Leslie Blizard

***Science* International**

Europe (science@science-int.co.uk) **editorial: international managing editor** Andrew M. Sugden; **senior editor/commentary** Julia Fahrenkamp-Uppenbrink; **senior editors** Caroline Ash, Stella M. Hurtley, Ian S. Osborne, Peter Stern, Maria Cruz; **contributing editor** Helen Pickersgill; **editorial support** Rachel Roberts, Alice Whaley; **administrative support** Janet Clements, Joan Cuthbert, John Wood; **news: deputy news editor, u.k.** Daniel Clery; **contributing editor, europe** Martin Enserink; **contributing correspondents** Michael Balter (Paris), Kai Kupferschmidt (Berlin), Tania Rabesandratana (Brussels), Gretchen Vogel (Berlin)

Asia Japan Office: Asca Corporation, Tomoko Furusawa, Rustic Bldg. 7F, 77 Tenjin-cho, Shinjuku-ku, Tokyo 162-0808, Japan; +81 3 6802 4616, FAX +81 3 6802 4615, inquiry@sciencemag.jp; **contributing editor, asia** Mara Hvistendahl [China: mhvisten@aaas.org]; **contributing correspondents** Dennis Normile [Japan: dnormile@gol.com]; Christina Larson [China: christina.larson@gmail.com]; Pallava Bagla [South Asia: pallava.bagla@gmail.com]

Latin America contributing correspondent Lizzie Wade [Mexico City: lizziewade@outlook.com]

EXECUTIVE PUBLISHER **Alan I. Leshner**
PUBLISHER **Beth Rosner** CHIEF DIGITAL MEDIA OFFICER **Rob Covey**

Fulfillment Systems and Operations (membership@aaas.org); **customer service supervisor** Pat Butler; **specialists** LaToya Casteel, Michelle Ofordire, Javia Jennings, Chanta Stewart; **manager, data entry** Mickie Napoleoni; **data entry specialists** JJ Regan, Jaimee Wise, Fiona Giblin

Business Operations and Administration director Deborah Rivera-Wienhold; **business systems and financial analysis director** Randy Yi; **manager of fulfillment systems** Neal Hawkins; **systems analyst** Nicole Mehmedovich; **manager, business analysis** Eric Knott; **manager, business operations** Jessica Tierney; **business analysts** Cory Lipman, Cooper Tilton, Celeste Troxler; **financial analyst** Jeremy Clay; **rights and permissions: assistant director** Emilie David; **associate** Elizabeth Sandler; **marketing director** Ian King; **marketing managers** Alison Chandler, Julianne Wielga, Justin Sawyers; **marketing associates** Mary Ellen Crowley, Elizabeth Sattler, Rebecca Riffkin; **senior marketing executive** Jennifer Reeves; **director, site licensing** Tom Ryan; **director, corporate relations** Eileen Bernadette Moran; **senior publisher relations specialist** Kiki Forsythe; **publisher relations manager** Catherine Holland; **publisher relations, eastern region** Keith Layson; **publisher relations, western region** Ryan Rexroth; **customer relations manager** Iquo Edim; **customer relations analysts** Simon Chong, Lana Guz; **associate director, marketing** Christina Schlecht; **marketing associates** Paulina Curto, Mitchell Edmund; **electronic media: director** Lizabeth Harman; **assistant manager** Lisa Stanford; **production specialists** Antoinette Hodal, Nichele Johnston, Lori Murphy, Kimberly Oster; **web and new media: senior project manager** Trista Smith, **project leader** Luke Johnson **computer specialists** Walter Jones, Kai Zhang, **web developer** Chris Coleman; **program director, aaas member central** Peggy Mihelich

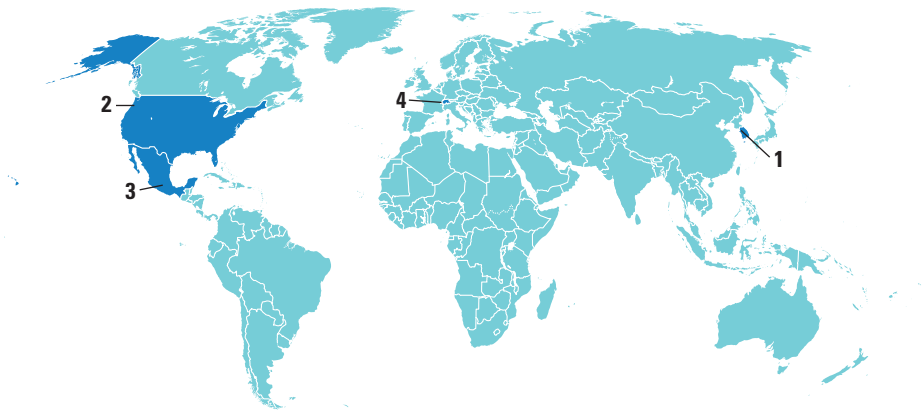
director, global collaboration, custom publications, advertising Bill Moran
editor, custom publishing Sean Sanders: 202-326-6430
assistant editor, custom publishing Tianna Hicklin 202-326-6463
sponsored content specialist Candice Nulsen 202-256-1528
associate director, collaboration, custom publications/china/taiwan/korea/singapore Ruolei Wu +86-1367-101-5294

Product (science_advertising@aaas.org); **midwest** Rick Bongiovanni: 330-405-7080, FAX 330-405-7081; **east coast/e. canada** Laurie Faraday: 508-747-9395, FAX 617-507-8189; **west coast/w. canada** Lynne Stickrod: 415-931-9782, FAX 415-520-6940; **uk europe/asia** Roger Goncalves: TEL/FAX +41 43 243 1358; **japan**, Makiko Hara: +81 (0) 3 6802 4616, FAX +81 (0) 3 6802 4615; ads@sciencemag.jp; **china/taiwan** Ruolei Wu: +86 1367 1015 294 rwu@aaas.org

worldwide associate director of *Science* Careers Tracy Holmes: +44 (0) 1223 326525, FAX +44 (0) 1223 326532

Classified (advertise@sciencecareers.org); **u.s./canada/south america** Tina Burks: 202-326-6577; **u.s. corporate** Candice Nulsen 202-256-1528; **sales administrator** Marci Gallun; **europe/row sales** Axel Gesatzki, Sarah Lelarge; **sales assistant** Kelly Grace; **japan** Yuri Kobayashi +81 (0)90-9110-1719; careerads@sciencemag.jp; **china/taiwan** Ruolei Wu: +86 1367 1015 294 rwu@aaas.org; **advertising support manager** Karen Foote: 202-326-6740; **advertising production operations manager** Deborah Tompkins; **senior production specialist/graphic designer** Amy Hardcastle; **production specialist** Yuse Lajiminmuhip; **senior traffic associate** Christine Hall; **sales coordinator** Shirley Young; **marketing manager** Allison Pritchard; **marketing associate** Aimee Aponte

AROUND THE WORLD



Seoul **1**

New Bird Flu Strain Threatens Korean Research

A dangerous new strain of bird flu in South Korea has spread nationwide despite efforts to clamp down on the virus. Authorities have culled 2.8 million domestic chickens and ducks since the outbreak began, and the strain has also killed dozens of Baikal teal and other migratory birds. As yet, there are no reports of human infections. Scientists are puzzling over where the H5N8 strain, never before seen in a highly pathogenic form, originated.

Outbreak. Korean health officials remove eggs from a duck farm suspected of carrying bird flu.

Researchers are now scrambling to keep the virus out of the country's premier poultry research center. A wild goose infected with the virus was found dead on 1 February just 10 kilometers from the Suwon campus of the National Institute of Animal Science (NIAS), near Seoul. The facility houses more than 13,000 hens and nearly 5000 ducks for research on breed improvement and animal husbandry. "If the virus infects the facility, we would cull all of the poultry," says NIAS's Yong-sup Song. That would put a serious dent in the center's genetic resources and set back ongoing research programs. **http://scim.ag/avianflu**

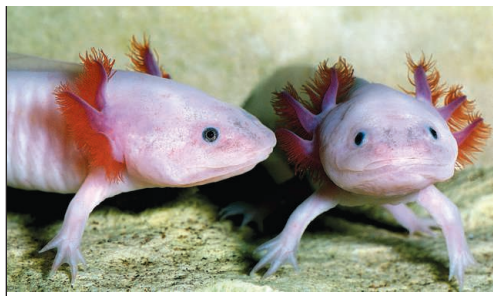
Portland, Oregon **2**

First Fish Ready to Swim Off Endangered Species List

A tiny minnow has bounced back from near-extinction. The U.S. Fish and Wildlife Service (FWS) says populations of the Oregon chub (*Oregonichthys crameri*) are healthy enough to remove the 9-cm-long fish from its list of threatened and endangered wildlife. Last week's announcement marks the first time an endangered fish has recovered enough to be delisted.

The chub lives in beaver ponds, oxbows, and calm streams of the Willamette River Valley of western Oregon. After the 1940s, populations plummeted from habitat damage by logging, pollution, and dams. When the fish was listed by FWS in 1993, only nine known populations remained. Predation by largemouth bass and other non-native fishes was the largest threat to the remaining chub.

Since then, the Oregon Department of Fish and Wildlife and other groups started 20 new populations of the chub in predator-free ponds. And dozens of other surviving populations have been discovered in the wild. Changes to dam management have lowered the threat to remaining habitat. FWS will accept expert and public comment on its proposal to delist until 7 April.



Mexico City **3**

Salamander Sightings Prove Reports of Extinction Premature

The axolotl salamander's only known home in the wild, the Xochimilco canals of Mexico City, has become increasingly polluted, but recent reports of the amphibian's extinction have been (not-so-greatly) exaggerated.

Two weeks after announcing that months of searching the canals hadn't turned up any axolotls, scientists in Mexico City have some good news: two of the unique salamanders were spotted on 4 February. "There's been an alarming reduction in population density," says Luis Zambrano, a biologist at the National Autonomous University of Mexico in Mexico City who studies the axolotl. "But I can guarantee that [the axolotl] is not yet extinct" in the wild.

Axolotls (*Ambystoma mexicanum*) have long intrigued scientists with their odd appearance, their impressive ability to regenerate limbs, and the fact that they don't undergo metamorphosis like other salamanders, instead retaining their feathery gills and other tadpolelike features into adulthood. Axolotls are popular pets and lab animals, but Zambrano says no reintroduction efforts with captive populations will be tried until scientists are positive the axolotl is "100% extinct" in the wild.

Geneva, Switzerland **4**

CERN to Study Possibility Of 100-Kilometer Atom-Smasher

Scientists at the European particle physics laboratory will study plans for a pair of circular particle colliders 80 to 100 kilometers in circumference, to be built one after the other in the same tunnel. The plan would depart from the current vision for global particle physics, in which the successor to CERN's current 27-kilometer-long Large Hadron Collider (LHC) would be a straight linear collider that would smash electrons into positrons. The LHC smashes protons

Published by AAAS

and in 2012 discovered the Higgs boson.

In the new scheme, physicists would build a somewhat lower energy circular electron-positron collider and later a proton collider able to reach energies seven times as high as the LHC. In reusing the tunnel, CERN would repeat a strategy that reduced costs during the construction of the LHC. The 5-year study is "not about digging holes in the ground now or asking governments for money," says CERN spokesman James Gillies. "It's just about considering the technology that would be available in 20 or 30 years." **http://scim.ag/atomsmasher**

NEWSMAKERS

Three Q's


Thevis

Each Olympics, there is a race between athletes seeking an artificial advantage and the anti-doping experts trying to catch them. At the Winter Olympics in Sochi, a new chemical may be in the mix. **Mario Thevis**, a forensic chemist at the German Sport University Cologne, tested a substance obtained by German journalists during an undercover investigation of a Russian scientist selling "full size MGF" for $1000 per milligram. Thevis confirmed the sample contained mechano growth factor (MGF), which can prompt muscle growth and is undetectable by current testing methods.

Q: What is the substance you found in the sample?

M.T.: The closest way to describe it is human IGF-1 isoform 4. The mRNA of isoform 4 is elevated when mechanical stress is applied to muscle tissue. We could deduce that we were dealing with a highly pure and therefore probably highly dangerous substance.

Q: What might the side effects be?

M.T.: We don't know. It could cause any of the side effects associated with IGF-1, such as cardiovascular issues. Some of the growth factors also have cancer-causing effects. We can't prove or rule out any of these.

Q: Now that you know this might be in circulation among athletes, why can't you test for it right away?

M.T.: Practically we can, but we have to demonstrate that our test is fit for the purpose. We have to evaluate whether our detection limits are in the range for physiological or therapeutic amounts, even though we have no idea how much that would be.

Extended interview at http://scim.ag/_sochi.

Science LIVE

Join us on Thursday, 20 February, for a live chat with experts on **assessing the harm of drugs for rational drug policy**. **http://scim.ag/science-live**

Random Sample

Charles Darwin Gets Busted

An empty pedestal has become the inspiration for a collection of zany sculptures of Charles Darwin. Students and staff from University College London competed to fill a void left by the relocation of an original bust of the eminent naturalist.

The competition's seven entries—on display in a 7-week-long exhibition titled *Darwin (or) Bust* at the Grant Museum of Zoology in London—are based on 3D scans of the original plaster bust, which was moved to a newly constructed building on campus. The new designs come with a twist. One entry renders Darwin's face as a collage of pages from his seminal book, *The Origin of Species*; another, a likeness molded from a transparent gel through which ants will be enticed to tunnel; and a third has a pensive-looking Darwin crocheted out of yarn. The winning entry was announced at the exhibit's grand opening on 12 February, the 205th anniversary of Darwin's birth. We wanted people to "get creative, get technical, [and] get messy" while reimagining the great man, says Grant Museum curator Mark Carnall.

FINDINGS

Invasive Fire Ants Meet Their Match

For 60 years, imported red fire ants have terrorized people, livestock, and native ants as they spread in large numbers across the southeastern United States, aggressively stinging those who got in the way and proving nearly invincible. But another foreigner, the invasive tawny crazy ant, can detoxify the fire ant's deadly venom and, for the past decade, has been moving into fire ant territories, Edward LeBrun, an ecologist at the University of Texas, Austin, and his colleagues report online this week in *Science*. In Texas, LeBrun had noticed unusual grooming-like behavior by tawny crazy ants that fire ants had swiped with their stingers laden with venom drops. He discovered that the tawny ants were wiping themselves down with their own venom to counter the chemical attack. It's unclear how or when these ants took up this defense, says Michael Kaspari, an entomologist at the University of Oklahoma, Norman, but it looks like "this 60-year dynasty of the fire ants [in the United States] is coming to a close, and it's coming to a close in a fairly unusual way." **http://scim.ag/tawnyants**


Detox. A tawny ant reaches for antivenom on its abdomen.

CREDITS (TOP TO BOTTOM): © UCL, INSTITUTE OF MAKING/ROBERT EAGLE (3); IMAGO/CAMERA 4/NEWSCOM; LAWRENCE GILBERT

Published by AAAS

Clovis cache. The child's skull was found in Montana with a host of Clovis tools.

GENOMES

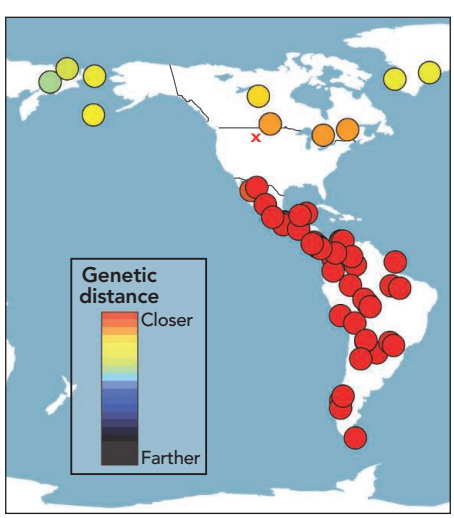
Ancient Infant Was Ancestor Of Today's Native Americans

In 1968, when Sarah Anzick was 2 years old, a construction worker discovered more than 100 stone and bone tools on her family's land near Wilsall, Montana. The artifacts were blanketed with red ochre, and with them, also covered with ochre, was the skull of a young child. In the years since, archaeologists concluded that the skull was about 12,700 years old—the oldest known burial in North America—and that the tools belonged to the Clovis culture, one of the first in the New World. Meanwhile, Sarah Anzick grew up, became a genome researcher at the National Institutes of Health (NIH), and dreamed of sequencing the rare bones.

This week, she is the second author on a paper in *Nature* that reports the complete sequence of the Anzick child's nuclear genome. The sequencing effort, led by ancient DNA experts Eske Willerslev and Morten Rasmussen of the University of Copenhagen, comes to a dramatic conclusion: The 1- to 2-year-old Clovis child, now known to be a boy, is directly ancestral to today's native peoples from Central and South America. "Their data are very convincing … that the Clovis Anzick child was part of the population that gave rise to North, Central, and Southern American groups," says geneticist Connie Mulligan of the University of Florida in Gainesville.

If correct, the findings refute the Solutrean hypothesis, which postulates that ancient migrants from Western Europe founded the Clovis culture (*Science*, 16 March 2012, p. 1289). The data also undermine contentions that today's Native Americans descend from later migrants to the Americas, rather than from the earlier Paleoindians. And that could help tribes that want to claim and rebury ancient American skeletons such as that of the 9400-year-old Kennewick Man from Washington state. "This is proof that Kennewick Man was Native American," says archaeologist Dennis Jenkins of the University of Oregon, Eugene. Sarah Anzick, whose family is in possession of the infant, says that it is likely to be reburied in May.

Researchers have long wanted to examine the DNA of the first Americans for clues to their origins. But even after scientists developed tools to get DNA from poorly preserved bones, they lacked the full cooperation of today's Native Americans. The Anzick child remained available for study in part because it was found on private land, so the U.S. Native American Graves Protection and Repatriation Act (NAGPRA)—which gives native peoples the right to claim and rebury many human remains—does not apply (*Science*, 8 October 2010, p. 166).



Common ancestor? The Montana infant (X) is closely related to today's Native Americans (circles).

Willerslev and colleagues extracted DNA from bone fragments taken from the child's skull and one of its ribs, then sequenced the genome. They compared the genome with those of 143 modern non-African populations, including 52 Native American ones, in a database compiled over several decades by geneticist David Reich of Harvard Medical School and others. The database includes 45 DNA samples from Central and South America and seven from Canada and the Arctic, but none from the lower 48 states, in part because U.S.-based Native American groups have historically resisted providing DNA samples, and because Reich felt that true informed consent was lacking for some samples.

Despite the North American data gap, the team was able to determine that the Anzick genome was much more closely related to Native Americans than to any other group worldwide (see map). The child's DNA more closely resembles that of Central and South Americans than Native Americans from the far north, although the relationship is still very close, Willerslev says. Comparing the Anzick genome with that of a 24,000-year-old Siberian boy (*Science*, 25 October 2013, p. 409) and a 4000-year-old Paleo-Eskimo from Greenland confirms that Native Americans originally come from Northeast Asia.

How to explain the north-south difference? The team proposes that an ancestral population that lived several thousand years before the Clovis period split into two groups, one staying north and one going south. Just where and when this split happened cannot be determined from the genetic data, Willerslev and Rasmussen say. The northerners then likely mated with peoples who came in later from Asia, and so became slightly more genetically distant from Anzick.

The study "is a real technical and analytical achievement," says anthropologist Theodore Schurr of the University of Pennsylvania, who was not a co-author. It "effectively puts

Published by AAAS

the Solutrean hypothesis to rest," he says. But advocates of that idea take umbrage at such a dismissal. "This is a single individual and can in no way represent all that was happening," says archaeologist Bruce Bradley of the University of Exeter in the United Kingdom.

Schurr cautions that the lack of U.S.-based Native American genomes could have biased the analysis of how closely related the Anzick boy is to today's native peoples. "The authors might want to be more cautious about making such definitive statements" about the Clovis culture's ancestral status "without having … a much broader sampling of North American Indian populations," he says.

Team members say they hope to get more U.S. data. "We hope the continued dialogue with local populations and studies like this will entice … Native peoples to participate in genetic studies," Rasmussen says. Shane Doyle, a professor of Native American studies at Montana State University, Bozeman, and a member of the Crow tribe, shares that hope. Doyle is coordinating negotiations about reburying the child with the Anzick family, the researchers, and members of 11 local tribal groups, but he sees the value of such research for today's Native Americans. "This is absolutely going to change the game about how we think about Paleoindians and their links to modern-day tribes," Doyle says.

Both Doyle and Anzick (who notes that she is acting for her family, not NIH) say they are agonizing over how, and how soon, the child should be reburied. They worry that reburial will destroy data that might be retrieved years from now with better genetic techniques. Schurr agrees: "This is why scientists are fighting against NAGPRA repatriations of Paleoamerican remains, as much can be learned from these ancient samples."

But Doyle and Anzick insist that the child should be reburied out of respect for his Native American descendants. "The boy has given us an amazing gift," Doyle says. "Now we must repay that by putting him back where he belongs." **–MICHAEL BALTER**

ASTROPHYSICS

India Poised to Join Hunt for Gravitational Waves

NEW DELHI—Every so often in the universe, two neutron stars or black holes collide so violently that space and time themselves shudder. An emerging global network of detectors is watching for these ripples in spacetime, which are predicted by Albert Einstein's general theory of relativity. Now it has a new recruit: India. "India intends to host the third detector" in a U.S.-based array known as the Laser Interferometer Gravitational-Wave Observatory (LIGO), Prime Minister Manmohan Singh announced on 3 February at the Indian Science Congress in Jammu. The network's expansion should help physicists pinpoint sources of the waves—assuming they can be detected.

Indian scientists say the government is likely to commit $201 million over 15 years to its facility, LIGO-India. "LIGO will bring some of the best international and Indian astrophysicists to work on Indian soil," says Ratan Kumar Sinha, a nuclear engineer and chair of India's Atomic Energy Commission. The U.S. National Science Foundation (NSF) is jazzed as well. "I'm very excited about this because the science reward is so good," says F. Fleming Crim, NSF's assistant director for mathematical and physical sciences. However, he cautions, the Indian government must still develop a management structure for the project and commit to a schedule and budget.

To detect gravitational waves, LIGO aims to measure the stretching of space itself. Built by NSF for $294 million, LIGO comprises two exquisitely sensitive optical interferometers located 3000 kilometers apart, in Hanford, Washington, and Livingston, Louisiana. Each uses laser light to continuously compare the relative lengths of two 4-kilometer-long arms set at right angles, searching for changes as small as 1/10,000 the width of a proton.

LIGO ran from 2002 to 2010 but didn't spot a signal. Neither have two similar interferometers in Europe, GEO600 near Hannover, Germany, and VIRGO near Pisa, Italy. But so far scientists have surveyed only our cosmic neighborhood, which is unlikely to harbor a source. By 2015, however, a $205 million upgrade called Advanced LIGO will make both U.S. detectors 10 times more sensitive—able to probe a volume 650 million light-years in radius that should contain at least a few sources.

"I'm very excited about this because the science reward is so good."
—F. FLEMING CRIM, NSF

As part of the upgrade, LIGO scientists want to extend the global network. They are looking for a home for an extra set of mirrors and parts that they had planned to use in a second interferometer at Hanford that would have crosschecked the first one. LIGO's 8-year run indicated that crosschecking is not necessary, says LIGO Chief Scientist Stanley Whitcomb of the California Institute of Technology in Pasadena. So in 2010, LIGO proposed building a third station in Australia (*Science*, 27 August 2010, p. 1003). After the Australian government declined that offer in 2011, Indian researchers expressed interest.

LIGO and VIRGO have shared data since 2007, but an Indian detector would bolster the network's ability to pinpoint sources. If the two LIGO detectors and VIRGO all picked up a signal, researchers could compare its arrival times to locate the source in the sky. But such triangulation would work poorly for sources lying near the plane defined by three detectors. A fourth detector lying outside that plane would help locate sources over the entire sky to within a few degrees.

Some Indian researchers question whether the hefty investment in LIGO-India is the best use of their country's science budget; others worry that it will be hard to find a suitably quiet location for the vibration-sensitive facility. Astrophysicist Bala Iyer, chair of the governing council of IndIGO, a consortium of Indian gravitational wave researchers, dismisses those concerns. "The community is very happy," he says, noting that ongoing survey work has identified several possible locations.

Even before LIGO-India comes on, perhaps in 2020, the global detector network should receive a similar boost from a different Asian site. Tunnels should be completed next month for the 3-kilometer-long arms of the $156 million Kamioka Gravitational Wave Detector (KAGRA), which scientists aim to fire up in 2017, says Takaaki Kajita, a physicist at the University of Tokyo. Of course, he says, all five detectors together would do the best job—which is why physicists are hoping that Singh's promise soon becomes reality.

–PALLAVA BAGLA AND ADRIAN CHO

Published by AAAS

2014 SCIENCE INDICATORS

New NSF Report Shows Where U.S. Leads and Lags

It's easy to lose track of the continued dominance of the United States on many indicators of scientific prowess as national policymakers fret about the rise of China and other Asian countries. But it shows up clearly in the latest edition of the National Science Foundation's (NSF's) biennial compendium of global trends in science and engineering.

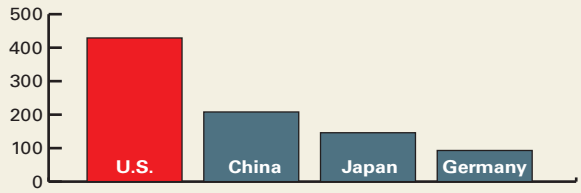
The rapid U.S. rebound from the 2008 recession is only one of many trends highlighted in *Science and Engineering Indicators 2014*, explains Ray Bowen of the National Science Board, the NSF oversight body that issued last week's report. So is the economic payoff from the continued investment in science by China and its neighbors.

The 600-page tome, accompanied by an even longer appendix of tables, offers a feast for policy wonks. The following tidbits from that banquet reveal areas where the United States remains preeminent, and other areas in which it trails its global competitors. The data are from 2012 or the most recent year.

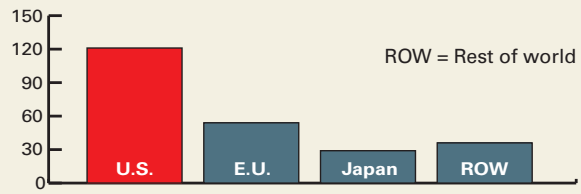
–JEFFREY MERVIS

Areas Where the U.S. Leads

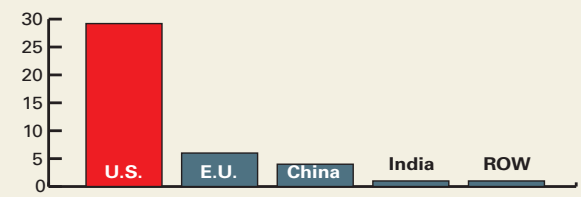
R&D Spending
(In billions of dollars)



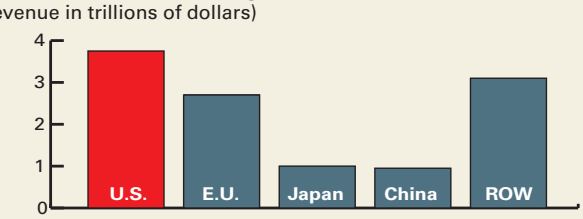
Income From IP exports
(In billions of dollars)



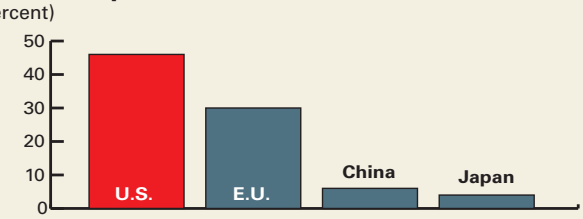
Venture Capital Investment
(In billions of dollars)



Commercial Knowledge Intensive Services
(Revenue in trillions of dollars)

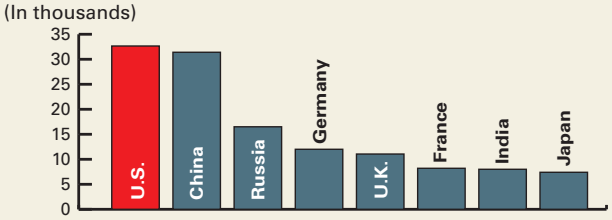


Share of Top 1% Most-Cited S&E Articles
(Percent)

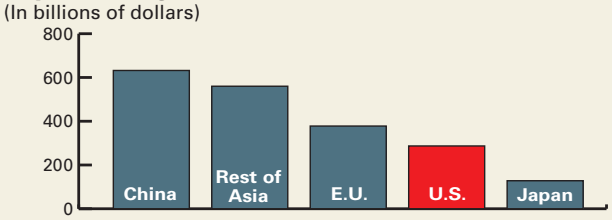


Areas Where the U.S. Is Among the Pack

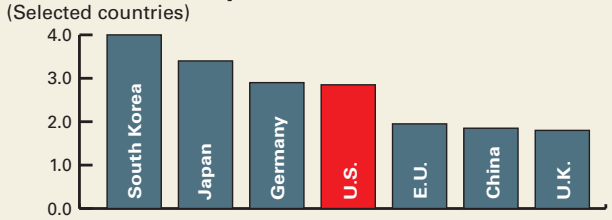
S&E Ph.D.s Awarded
(In thousands)



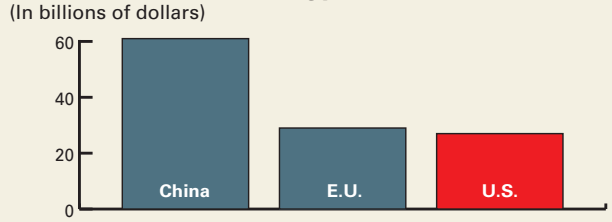
Export of High-Tech Products
(In billions of dollars)



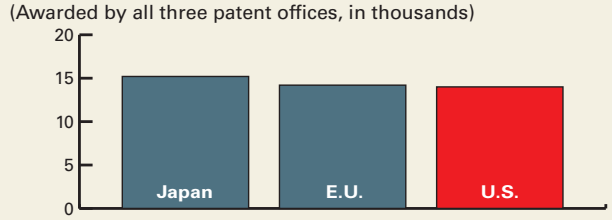
Research Intensity: Percent of GDP
(Selected countries)



Commercial Clean Energy Investment
(In billions of dollars)

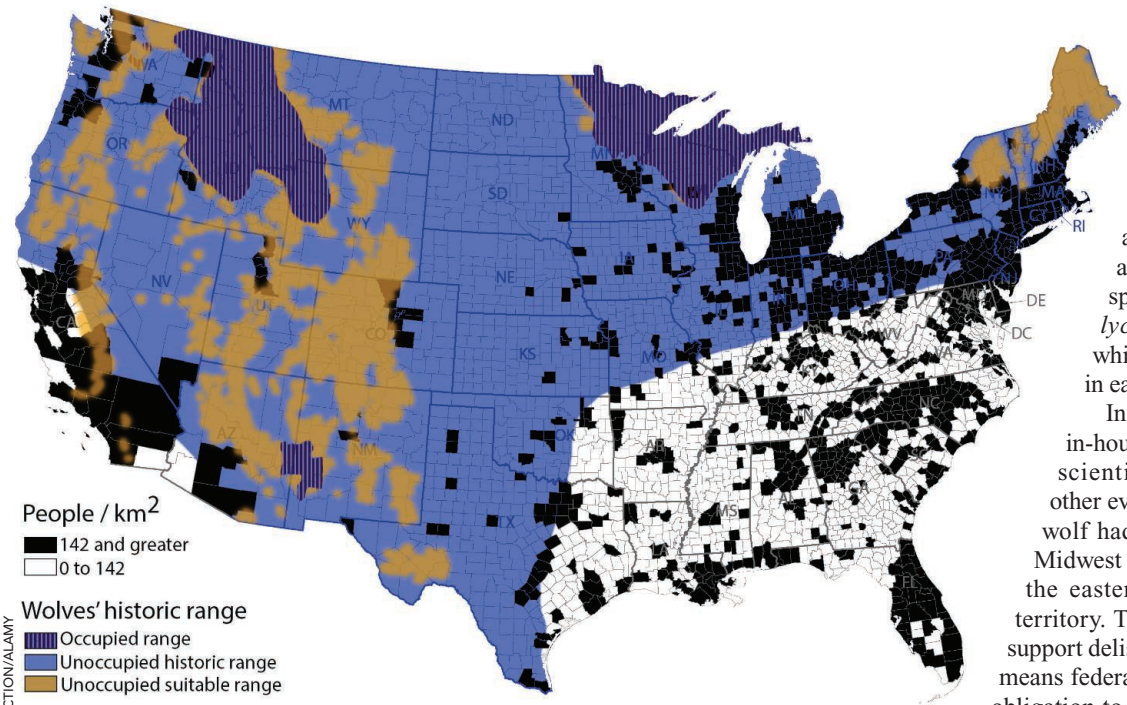


Triadic Patents
(Awarded by all three patent offices, in thousands)



CREDIT: *SCIENCE AND ENGINEERING INDICATORS 2014*

Published by AAAS



Contraction. Gray wolves once inhabited a large swath of the continental United States, but are now confined to a few areas.

CREDITS (TOP TO BOTTOM): JEREMY BRUSKOTTER/OHIO STATE UNIVERSITY AND COURTESY OF THE CENTER FOR BIOLOGICAL DIVERSITY; © NATIONAL GEOGRAPHIC IMAGE COLLECTION/ALAMY

ENDANGERED SPECIES

Science Behind Plan to Ease Wolf Protection Is Flawed, Panel Says

A controversial proposal to lift U.S. government protections for most gray wolves living in the lower 48 states suffered a major blow last week, when an independent review panel nixed the underlying science. The U.S. Fish and Wildlife Service (FWS) had argued that the gray wolf, which has rebounded in some parts of the West, does not need continued protection to recover in the East because it never lived there. But the four-member panel unanimously dismissed that argument, saying it "does not currently represent the 'best available science.' "

The verdict is widely seen as a game-changer. "The service will have to take this into account," says Steven Courtney, an ecologist at the National Center for Ecological Analysis and Synthesis (NCEAS) at the University of California (UC), Santa Barbara. Courtney led the NCEAS review, which the agency requested.

The panel's 7 February report is the latest twist in a messy and emotionally fraught saga. Wolf researchers estimate that some 2 million wolves lived in the continental United States 600 years ago (see map). After being hunted to near extinction, the gray wolf (*Canis lupus*) was placed on the federal endangered species list in 1975. It was later reintroduced in the Rocky Mountains over vehement objections from ranchers and others. The population ultimately recovered to some 6000 animals in some western and upper midwestern states, and in 2011 the federal government lifted wolf protections in six of those states, all of which now have legal hunts. In June 2013, FWS released its proposal to totally remove the gray wolf from the endangered species list in every state, while adding protections for the Mexican gray wolf, a subspecies in the Southwest. It also proposed to recognize a new—and controversial—species of wolf, *Canis lupus lycaon*, or the eastern wolf, which some argue is found today in eastern Canada.

Legal hunt. Hunters in Idaho and five other states can now kill wolves.

In a 2012 monograph published in-house without review, four FWS scientists drew on genetic and other evidence to argue that the gray wolf had never inhabited the upper Midwest and Northeast; instead, only the eastern wolf had occupied that territory. That scenario, if true, would support delisting the gray wolf because it means federal officials don't have a legal obligation to try to restore the species to 22 eastern states.

But the NCEAS panel, which included specialists on wolf genetics, said the idea of an eastern wolf is "not universally accepted and … 'not settled,' " and rejected the idea that the two species had never mixed in the East. There's no question that the gray wolf used to be present in the East, too, says panelist Paul Wilson, a conservation geneticist at Trent University, Peterborough, in Canada, who believes that the eastern wolf is a separate species.

It appears the agency was trying to use "some kind of taxonomic sleight of hand" to support delisting, says panelist Robert Wayne, a conservation geneticist at UC Los Angeles. That would set a "dangerous precedent," he adds: It would be the first time that the federal government has removed a species from the list as a result of a taxonomic redefinition, and not a population recovery.

It's not yet clear how FWS will respond to the setback. Agency Director Dan Ashe called the NCEAS report "an important step" in a statement, but didn't tip his hand. The agency has reopened public comment on the delisting proposal until 27 March; so far, more than 1 million people have sent in responses, the most in the agency's history.

In the meantime, wolves continue to be heavily hunted in the states where delisting has occurred, with at least 1000 killed in the current season. Those states are allowed to reduce the number of wolves within their borders to 100 animals, or 10 packs with 10 individuals each. But some scientists fear that is too few to maintain healthy populations over the long term.

–VIRGINIA MORELL

Published by AAAS

Claiming credit. Marthe Gautier's talk at a recent genetics meeting in Bordeaux was canceled.

HISTORY OF SCIENCE

After More Than 50 Years, a Dispute Over Down Syndrome Discovery

It would have been a personal triumph for Marthe Gautier, an 88-year-old pediatric cardiologist and scientist living in Paris. On 31 January, during a meeting in Bordeaux, Gautier was to receive a medal for her role in the discovery of the cause of Down syndrome in the late 1950s. In a speech, she planned to tell an audience of younger French geneticists her story about the discovery—and how she felt the credit she deserved went to a male colleague, Jérôme Lejeune.

But Gautier's talk was canceled just hours in advance, and she received the medal a day later in a small, private ceremony. The French Federation of Human Genetics (FFGH), which organized the meeting, decided to scrap the event after two bailiffs showed up with a court order granting them permission to tape Gautier's speech. They were sent by the Jérôme Lejeune Foundation, which wanted to have a record of the talk. The foundation, which supports research and care for patients with genetic intellectual disabilities and campaigns against abortion, said it had reason to believe Gautier would "tarnish" the memory of Lejeune, who died in 1994.

A brilliant cytogeneticist with a storied career, Lejeune has become widely known as the scientist who discovered that Down syndrome is caused by an extra copy of chromosome 21. He received many awards, including one from former U.S. President John F. Kennedy. But in recent years, Gautier has claimed that she did most of the experimental work for the discovery. In the French newspaper *Le Monde*, Alain Bernheim, the president of the French Society of Human Genetics, last week compared her case to that of Rosalind Franklin, whose contribution to the discovery of the double helix structure of DNA in the early 1950s was long overlooked.

In an e-mail to *Science*, Gautier referred to an interview published on the Web for her version of events more than half a century ago. In it, she explained that she worked on Down syndrome in the pediatric unit led by Raymond Turpin at the Armand-Trousseau Hospital in Paris, which she joined in 1956 after a year at Harvard Medical School in Boston.

First author. Jérôme Lejeune, who passed away in 1994.

Human cytogenetics was just coming of age. In 1956, a Swedish team showed that humans have 46 chromosomes in every cell, not 48, as was widely believed. In the United States, Gautier had learned to grow heart cell cultures, so she proposed to set up an advanced cell culture lab and study Down syndrome. She says she received her first patient sample in May 1958; examing slides, she soon noticed an extra chromosome, but she was unable to identify it or take pictures with her low-power microscope. In June 1958, she "naively" accepted an offer from Lejeune, who Gautier says was studying Down syndrome using other techniques, to take her slides and get them photographed.

Gautier claims she was "shocked" when, after more than 6 months of silence, she learned that the discovery was about to be published in the journal of the French Academy of Sciences, with Lejeune as the first author and Turpin the last; Gautier was in the middle, her last name misspelled as Gauthier. Gautier doesn't dispute that Lejeune identified the 47th chromosome as an extra copy of chromosome 21, but she maintains that she was the first to notice the abnormal count.

While ackowledging that Gautier played a role, the Jérôme Lejeune Foundation claims that Lejeune himself made the discovery. "In July 1958, during a study of chromosomes of a so-called 'mongoloid' child, [Lejeune] discovered the existence of an extra chromosome on the 21st pair," according to the foundation's website. The foundation has denied that Lejeune appropriated Gautier's discovery; in a press statement, it says a letter Turpin sent in October 1958 suggests Gautier still hadn't seen the 47 chromosomes.

Things came to a head at the meeting in Bordeaux. After calling off Gautier's talk and the award ceremony, FFGH issued a statement saying it would have been "unacceptable" to hold the ceremony under the threat of a legal suit. But the federation also said it "bitterly regretted" the cancellation and condemned the use of legal power to put pressure on a scientific meeting.

Simone Gilgenkrantz, a professor emeritus of human genetics at the University of Lorraine in France and a friend of Gautier's, says the presentation, which she has seen, was "completely innocuous." Gautier writes in an e-mail to *Science* that she accepted the decision and that she felt unprepared to deal with what she calls "an aggression." "To talk under the pressure of justice is not tolerable for me or anyone else," she writes.

Ideology is fueling some of the rancor. Lejeune, a staunch Catholic, was horrified by the advent of prenatal diagnostics, which made it possible to screen fetuses for Down syndrome and other abnormalities, and abort those afflicted. He set out to find a

Published by AAAS

therapy for genetic intellectual disabilities like Down syndrome, but also campaigned tirelessly against abortion—which made him a lightning rod among the left wing in France. (Lejeune was friends with Pope John Paul II and the Vatican is now considering a request to beatify him.) In its statement, the foundation lashed out at Gautier's supporters for trying to discredit an ideological opponent. It said Gautier, at her age, can't be blamed for her "confusion," but called stories backing her version of events in *Le Monde* and *Libération*—both left-wing papers—"ideological terrorism."

Gilgenkrantz, who convinced Gautier to tell her story in 2009, says it should be told regardless of the politics involved. To her, it's one more tale of a female scientist wronged at a time when French science was still very sexist. "This is a story that must be known," she says, "in the name of women."

But Bernard Dutrillaux, who worked in Lejeune's lab from the mid-1960s until the early 1980s, believes that some score-settling may be going on. "Lejeune made a lot of enemies" among his peers, he says. Still, he condemns the foundation's legal maneuvers. Both sides, Dutrillaux says, should know better than to fight such "petty rear-guard battles." **–ELISABETH PAIN**

ENERGY

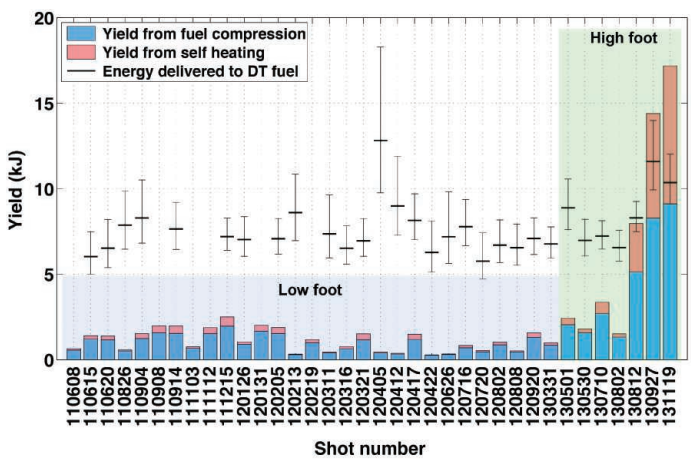
Laser Fusion Shots Take Step Toward Ignition

As it approaches its fifth birthday, the National Ignition Facility (NIF), a troubled laser fusion lab in California, has finally produced some results that outsiders can get enthusiastic about. In a series of experiments last year, NIF researchers produced yields of energy 10 times greater than achieved before and demonstrated the phenomenon of self-heating that will be crucial if fusion is to reach its ultimate goal of "ignition"—a self-sustaining reaction that produces more energy than it consumes. "This is a very significant achievement, and it's a very good place to start for going to higher yield," says Steven Rose of the Centre for Inertial Fusion Studies at Imperial College London.

NIF, at Lawrence Livermore National Laboratory in California, aims to release enormous amounts of energy by fusing together nuclei of two isotopes of hydrogen: deuterium and tritium. It heats the nuclei to enormous temperatures and pressures so that they smash together with enough force to overcome their mutual repulsion.

NIF has struggled to get its process to work, however (*Science*, 21 September 2012, p. 1444). Last year, instead of going flat-out for ignition, researchers there adopted a more exploratory approach to try to identify the problems. The data from last year's efforts, published this week in *Nature* and last week in *Physical Review Letters*, are the first sign that this approach is working. "It's a nice result," says Robert McCrory, director of the Laboratory for Laser Energetics at the University of Rochester in New York, who quickly cautions that NIF is still a long way from ignition. "People expecting a breakthrough soon will be disappointed."

To reach the extreme conditions necessary for fusion, NIF relies on a laser the size of a football stadium. It produces 192 ultraviolet beams in a pulse lasting just 15 nanoseconds that can deliver 1.9 megajoules (MJ) of energy, roughly the same as the kinetic energy of a 2-ton truck traveling at 160 kilometers per hour. The ultraviolet beams are converted into x-rays, which then compress a fuel capsule, a hollow plastic sphere smaller than a peppercorn, containing 0.17 milligrams of frozen deuterium and tritium. The aim is to produce a ball of fuel with a temperature of 50 million kelvin and 100 times the density of lead, conditions that can spark fusion.



On the up. A chart of NIF's fusion shots since mid-2011 shows that a "high-foot" laser pulse has boosted yields. Dates are given as YYMMDD.

Researchers realized last year that during this implosion, the plastic capsule was breaking up and mixing with the fuel, making it harder to achieve fusion. So they adjusted the timing of the laser pulse. Traditionally, it ran at a low power for most of its 20 nanoseconds to get the implosion moving without heating up the fuel and then finished with a burst of high power for the final spark. The rationale for this "low foot" approach was that the still-cool fuel would compress to a higher density at the end. The downside was that the slower speed allowed the capsule time to break up.

So they decided to try a pulse that started off with a higher power to implode faster and ended the pulse sooner, after 15 ns. Although such a "high foot" pulse wouldn't produce such high density at the end, the researchers hoped it would help control the mixing. A laser shot carried out on 13 August last year proved them right, with a huge jump in energy output. Another two shots, on 27 September and 19 November, did even better, actually producing more energy (14.4 and 17.3 kJ) than was deposited in the fusion fuel at the start (11 and 9 kJ), the first time that has been achieved in a laser fusion experiment. "We took a step back from what had been tried before and that gave us a leap forward," says NIF team leader Omar Hurricane.

Importantly, the team also saw a self-heating phenomenon that will be vital for increased yield. Fusion reactions produce alpha particles (helium nuclei); when reactions start in the core of the fuel, the alphas help by heating the surrounding cooler fuel up to reaction temperature. "This is the first strong indication of that bootstrap process," Hurricane says.

However, NIF is far from real gain (more energy out than the total input) because so much energy is lost converting the laser beams to x-rays and training them on the capsule. The team's best shot had a gain of less than 0.01. But there is general optimism following the past year's progress. "These are the right experiments to do," says Michael Campbell, a former NIF director now at Sandia National Laboratories. "Who knows how far they can take this?" **–DANIEL CLERY**

Published by AAAS

CREDIT: YANG ZHIPING/COLOR CHINA PHOTO/AP IMAGES

The Coming Copper Peak

Production of the vital metal will top out and decline within decades, according to a new model that may hold lessons for other resources

IF ELECTRONS ARE THE LIFEBLOOD OF A modern economy, copper makes up its blood vessels. In cables, wires, and contacts, copper is at the core of the electrical distribution system, from power stations to the delicate electronics that can display this page of *Science*. A small car has 20 kilograms of copper in everything from its starter motor to the radiator; hybrid cars have twice that. But even in the face of exponentially rising consumption—reaching 17 million metric tons in 2012—miners have for 10,000 years met the world's demand for copper.

But perhaps not for much longer. A group of resource specialists has taken the first shot at projecting how much more copper miners will wring from the planet. In their model runs, described this month in the journal *Resources, Conservation and Recycling*, production peaks by about midcentury even if copper is more abundant than most geologists believe. That would drive prices sky-high, trigger increased recycling, and force inferior substitutes for copper on the marketplace.

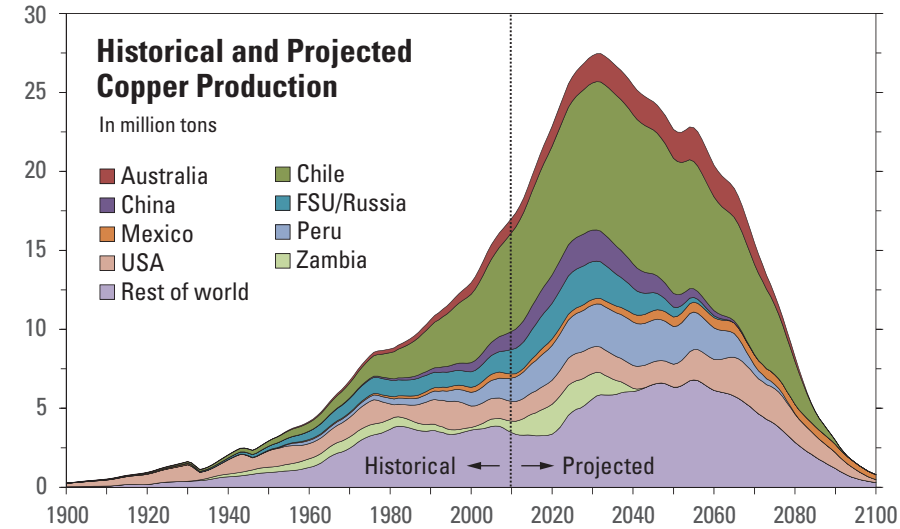
Predicting when production of any natural resource will peak is fraught with uncertainty. Witness the running debate over when world oil production will peak (*Science*, 3 February 2012, p. 522). And the early reception of the copper forecast is mixed. The work gives "a pretty good idea that likely we'll get a peak somewhere around midcentury," says industrial ecologist Thomas Graedel of Yale University. Technological optimists disagree. "Not that it couldn't happen, but I don't think it's likely to happen," says resource economist John Tilton, research professor emeritus at the Colorado School of Mines in Golden. New and better technology for extracting copper from the earth has always come to the rescue before, he notes, so he expects a much-delayed peak that businesses and consumers will comfortably accommodate by recycling more copper and using copper substitutes.

The copper debate could foreshadow others. The team is applying its depletion model to other mineral resources, from oil to lithium, that also face exponentially escalating demands on a depleting resource.

So far, so good

The techno-optimists were right about copper in the past. From nearly nothing in the mid-18th century, copper production soared along an exponential curve notched only by world wars and economic crises. That's all the more impressive considering the accompanying decline in the richness, or grade, of the ore being mined. Anyone extracting a

Published by AAAS



Heading up, until … The world has been producing ever larger amounts of copper (*left*, in a purifying electrolytic bath), but given the planet's finite endowment, production must peak. A model projection has production peaking by 2040.

Headed down. Miners must move megatons of low-grade copper ore at the Zijinshan mine in South China.

mineral resource goes for the richest, most easily mined deposits first, so ore grades ran from 10% to 20% copper until late in the 19th century and then plummeted to 2% to 3% in the early 20th century. Since the mid-1990s, the world copper grade has been just below 1% and in slow decline, according to data compiled by resource geologist Gavin Mudd of Monash University, Clayton, in Australia, in a 2013 paper in the *International Journal of Sustainable Development*.

Even though the available ores have become poorer, forcing miners to claw ever greater volumes of rock from kilometer-deep open pit mines, the price of copper has trended downward since 1900 (with a notable spike since 2005 driven by China's hunger for raw materials). Multiple factors have driven the price decline. Geologists found a new type of copper deposit—the porphyry ores of buried magma formations—that is now the source of most of the world's copper. And lately they have been finding extractable porphyry copper faster than it is produced, according to Richard Schodde of MinEx Consulting in Melbourne, Australia. Equipment manufacturers have built humongous shovels and dump trucks to move huge volumes of porphyry ore. And chemical engineers have developed processes such as heap leaching—trickling weak sulfuric acid through piled crushed ore—to get copper out of low-grade ore.

Even with the technology that is now in hand—not what might be developed someday—the copper now within reach of miners is considerable. At this past October's Geological Society of America meeting, U.S. Geological Survey researchers led by geologist Jane Hammarstrom of USGS headquarters in Reston, Virginia, reported their new assessment of the porphyry copper yet to be discovered that could be economically mined with current technology. By inferring how much copper might be beneath geologically likely terrains around the world, the group estimated that 2.2 billion metric tons of economically extractable metal remain to be found. At current rates of production, that's a 125-year supply for the world.

Not so fast

The world's copper future is not as rosy as a minimum "125-year supply" might suggest, however. For one thing, any future world will have more people in it, perhaps a third more by 2050. And the hope, at least, is that a larger proportion of those people will enjoy a higher standard of living, which today means a higher consumption of copper per person. Sooner or later, world copper production will increase until demand cannot be met from much-depleted deposits. At that point, production will peak and eventually go into decline—a pattern seen in the early 1970s with U.S. oil production.

For any resource, the timing of the peak depends on a dynamic interplay of geology, economics, and technology. But resource modeler Steve Mohr of the University of Technology, Sydney (UTS), in Australia, waded in anyway. For his 2010 dissertation, he developed a mathematical model for projecting production of mineral resources, taking account of expected demand and the amount thought to be still in the ground. In concept, it is much like the Hubbert curves drawn for peak oil production, but Mohr's model is the first to be applied to other mineral resources without the assumption that supplies are unlimited.

Now Mohr and Mudd have teamed up with resource specialists Stephen Northey of Australia's national research agency CSIRO in Clayton, Zhehan Weng of Monash Clayton, and Damien Giurco of UTS to apply Mohr's model to copper. For their study, the group drew on a database of the extractable copper at all known mine sites that was compiled by Mudd and Weng and published in 2012. The group assumed that per capita demand for mined copper would continue to rise at the historical rate of 1.6% per year and that the world's population would grow from today's 7.1 billion people to 10 billion in 2100. They taught the model how to behave realistically by fitting it to past copper production for each country and type of deposit. In the model, increasing demand elicits increased production at existing mines and the opening of new mines.

The model delivers some good news, suggesting that production can rise to meet expected demand for the next 2 to 3 decades. "It's not a story of doom and gloom, of running out tomorrow," Giurco says, "but rather of needing to be more mindful of use." But trouble comes in the longer term. With the amount of extractable copper in the Mudd and Weng compilation, the model shows production peaking just before 2040; after that, copper can't be extracted from depleted mines any faster, no matter how high the price.

Increasing the amount of accessible cop-

Published by AAAS

per in the model by 50% to account for what might yet be discovered moves the production peak back only a few years, to about 2045. It just takes a lot of copper to satisfy exponentially growing demand, Mohr says. In additional model runs performed at the request of *Science*, Mohr found that even doubling the available extractable copper pushes peak production back only to about 2050. And quadrupling it—an optimistic projection indeed—would mean the world would run short of copper by about 2075.

Copper trouble spots

So far, so bad—but technological optimists are quick to note that human ingenuity has confounded the gloom-sayers before. "As a society, we have tended to underestimate how much copper is out there, and how creative society can be about extracting it," Tilton says. He points out that in the 1970s, USGS estimated that about 1.6 billion tons of copper could be extracted with current technology. Today, the equivalent USGS figure is 3.1 billion tons. "And it's very likely to double again," Tilton says, even without including the copper on the ocean floor along midocean ridges. "We know the copper's there—it's a matter of resolving technical problems allowing extraction," he says.

Graedel doesn't go that far, saying the world has been so thoroughly explored for copper that most of the big deposits have probably already been found. Although there will be plenty of discoveries, they will likely be on the small side, he says. As for technological breakthroughs on a par with those in the past, he says, "you can't tell."

Furthermore, the models don't take into account constraints on copper mining that could make things worse. "The critical issues already constraining the copper industry are social, environmental, and economic issues," Mudd writes in an e-mail. Any process intended to extract a kilogram of metal locked in a ton of rock buried hundreds of meters down inevitably raises issues of energy and water consumption, pollution, and local community concerns. And such "environmental and societal constraints are getting stronger," Mudd says.

Mudd has a long list of copper mining trouble spots. The Reko Diq deposit in northwestern Pakistan close to both Iran and Afghanistan holds $232 billion of copper, but it is tantalizingly out of reach, with security problems and conflicts between local government and mining companies continuing to prevent development. The big Panguna mine in Bougainville, Papua New Guinea, has been closed for 25 years, ever since its social and environmental effects sparked a 10-year civil war that left about 20,000 dead.

And, looking ahead, on 15 January the U.S. Environmental Protection Agency issued a study of the potential effects of the yet-to-be-proposed Pebble Mine on Bristol Bay in southwestern Alaska. Environmental groups had already targeted the project, and the study gives them plenty of new ammunition, finding that it would destroy as much as 150 kilometers of salmon-supporting streams and wipe out more than 2000 hectares of wetlands, ponds, and lakes.

As a crude way of taking account of such social and environmental constraints on production, Northey and colleagues reduced the amount of copper available for extraction in their model by 50%. Then the peak that came in the late 2030s falls to the early 2020s, just a decade away.

After the peak

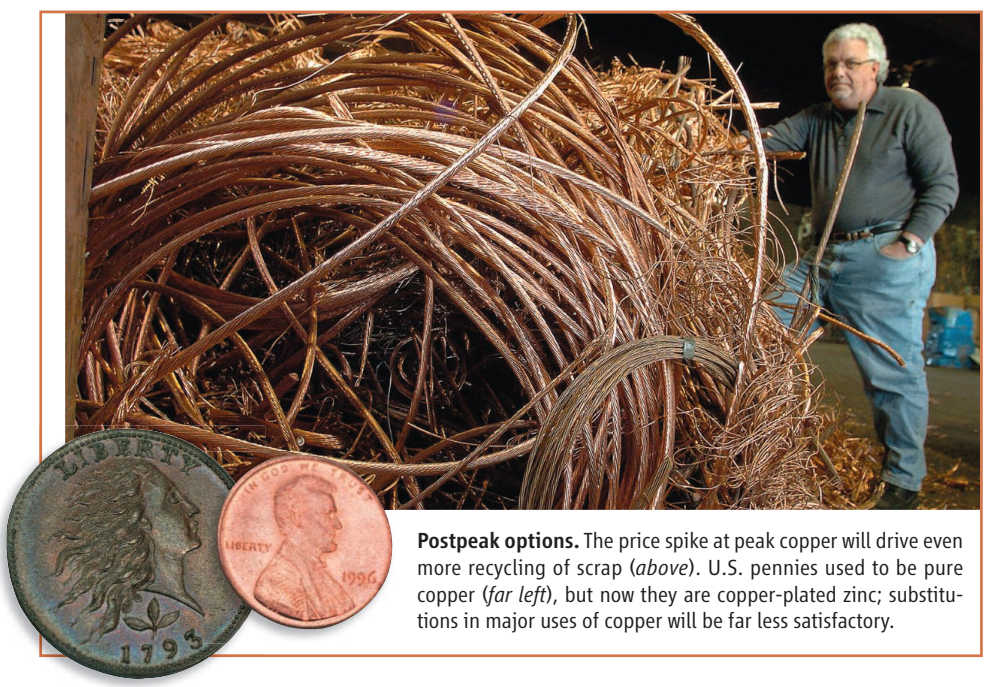
Whenever it comes, the copper peak will bring change. Alternative materials can replace copper in many uses, but substitution in some is easier than in others. In 1982, the U.S. copper penny—at least 88% copper since 1793—became 97.5% zinc and just 2.5% copper, mostly as copper plating, to discourage people from melting down the coins for their copper. But Graedel and his Yale colleagues reported in a paper published on 2 December 2013 in the *Proceedings of the National Academy of Sciences* that copper is one of four metals—chromium, manganese, and lead being the others—for which "no good substitutes are presently available for their major uses."

Recycling is more promising. Copper is already the third most recycled metal after iron and aluminum. Roughly 50% of the copper that goes out of service is returned to use, Graedel says. Governments could increase that figure by requiring product designs that, say, made recovery of copper wiring from cars easier and less expensive. Scarcity-driven price hikes will also boost recycling, Graedel notes.

Copper is far from the only mineral resource in a race between depletion—which pushes up costs—and new technology, which can increase supply and push costs down. Gold production has been flat for the past decade despite a soaring price (*Science*, 2 March 2012, p. 1038). Much crystal ball–gazing has considered the fate of world oil production. "Peakists" think the world may be at or near the peak now, pointing to the long run of $100-a-barrel oil as evidence that the squeeze is already on. Mohr's model is only slightly less pessimistic: It forecasts an oil peak in 2019, he reported in his dissertation.

Coal will begin to falter soon after, his model suggests, with production most likely peaking in 2034. The production of all fossil fuels, the bottom line of his dissertation, will peak by 2030, according to Mohr's best estimate. In the studies Mohr has had a hand in publishing, only lithium, the essential element of electric and hybrid vehicle batteries, looks to offer a sufficient supply through this century. So keep an eye on oil and gold the next few years; copper may peak close behind.

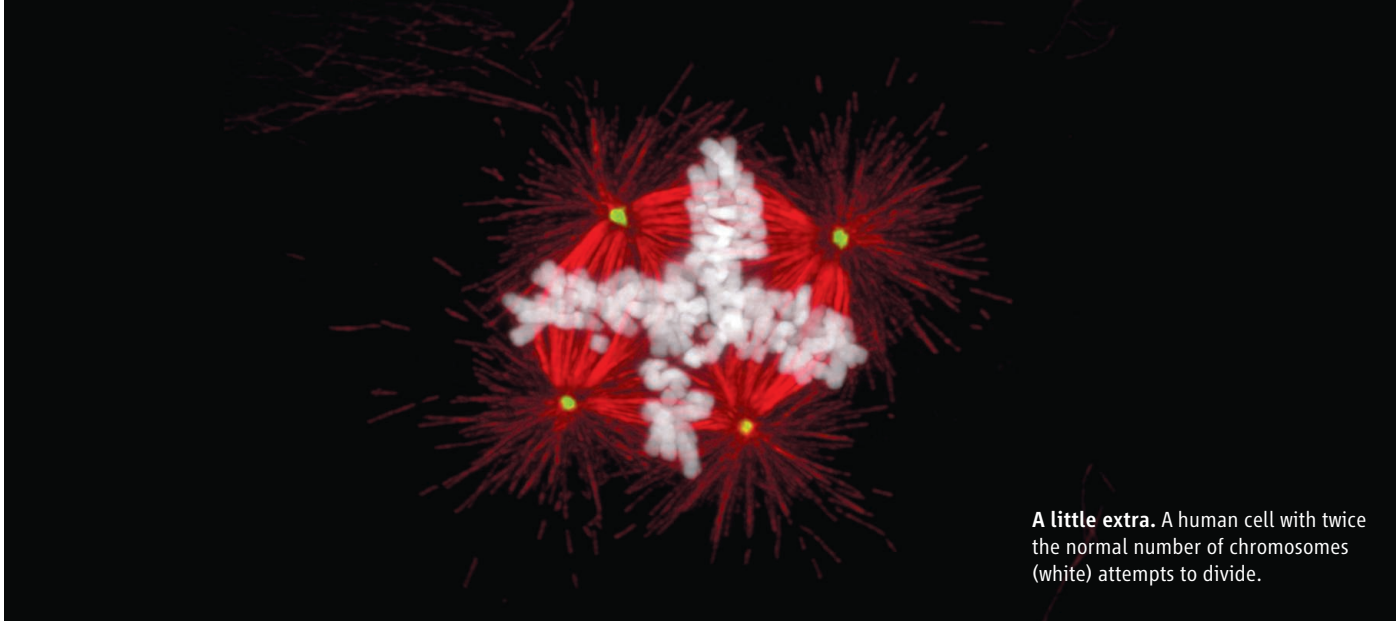
–RICHARD A. KERR



Postpeak options. The price spike at peak copper will drive even more recycling of scrap (*above*). U.S. pennies used to be pure copper (*far left*), but now they are copper-plated zinc; substitutions in major uses of copper will be far less satisfactory.

CREDITS (TOP TO BOTTOM): BRUCE KELLMAN/*THE NEWS TRIBUNE*/AP PHOTO; MARTY LEDERHANDLER/AP PHOTO

Published by AAAS



A little extra. A human cell with twice the normal number of chromosomes (white) attempts to divide.

CREDIT: NEIL J. GANEM/BOSTON UNIVERSITY

Strength in Numbers?

Some mammalian cells are loaded with extra sets of chromosomes, a state called polyploidy. What on Earth for?

A dividing cell generally follows a simple rule. After duplicating its DNA, the cell splits, yielding two daughter cells. That's why the movies of dividing mouse liver cells shot several years ago by Andrew Duncan, then a postdoc in Markus Grompe's group at the Oregon Health & Science University in Portland, flabbergasted his lab mates. "We saw a single cell giving rise to three and four daughter cells," says Duncan, who is now a tissue biologist at the University of Pittsburgh in Pennsylvania. And though chromosomes normally line up neatly across the middle of a cell before it divides, the chromosomes in many of the liver cells were arranged in unconventional formations, including multiple clusters.

The parental liver cells were forced to go through unusual maneuvers because they were polyploid, carrying extra sets of chromosomes. Polyploidy is rife among plants, insects, fish, and some other groups of organisms. But most human cells are diploid, outfitted with two sets of chromosomes that trace back to the set each provided by an egg and a sperm. Indeed, extra chromosomes usually spell trouble in mammalian cells. A few normal cells in people and other mammals, however, brim with extra genome copies—sometimes as many as a thousand. The contortions of the liver cells were surprising, but they had long been known to have a surfeit of chromosomes—as do cells in the heart and bone marrow.

For decades, researchers have speculated about whether polyploidy offers any advantages to mammalian cells, such as ramping up protein synthesis, but haven't been able to test their ideas. That has changed with the identification of several proteins that help regulate polyploidy. By cranking cells' allotment of chromosomes up or down, scientists recently have begun to explore the possible function of the odd cellular state. Do the extra chromosomes simply add bulk to cells that need it? Do they give cells reserve capacity that enables them to respond to stress and damage? "The real unanswered question is why any cell type is polyploid," says developmental geneticist Robert Duronio of the University of North Carolina (UNC), Chapel Hill. "We are poised to begin answering that question."

Online
sciencemag.org
Podcast interview with editor John Travis (http://scim.ag/pod_6172).

And even though the mystery of polyploidy's benefits remains unsolved, some researchers already hope to exploit the phenomenon. They are trying to turn polyploidy against certain cancers, compelling cells to cease their out-of-control division.

Risky excess

Polyploidy can seem like "a dangerous escapade," as Duronio and his colleagues put it in a 2009 paper. For cells that usually get along just fine with two sets of chromosomes, even one additional chromosome can be disastrous. An extra copy of chromosome 21 during development produces the disabilities of Down syndrome, for instance.

There's another potential drawback to polyploidy. "It can drive cancer," says David Pellman, a cell biologist and pediatric oncologist at the Dana-Farber Cancer Institute in Boston. He points to a 2013 *Nature Genetics* paper by Rameen Beroukhim, also of Dana-Farber, and colleagues that reported duplicated genomes in 37% of cancers. Polyploidy doesn't lead to cancer in every case, Pellman says, but it's a big enough risk that many cells go to great lengths to thwart it. p53, the watchful protein dubbed the guardian of the genome, often prompts cells with abnormal amounts of DNA to commit suicide or to curtail division. To become polyploid, therefore, cells have to disable it and other safeguards that protect against genome damage, notes biologist Gustavo Leone of Ohio State University, Columbus.

Researchers have gradually acquired a good grasp of the molecules and mechanisms that make cells polyploid, thanks mainly to their work on the cell cycle. A cell's life cycle includes milestones such as DNA duplication and division. An intricate network of proteins controls the cell's progress through the cycle, pushing it forward or holding it back. Under the right circumstances, researchers have found, some of these proteins steer cells toward polyploidy.

To tweak the chromosome content of cells, several research teams have recently genetically engineered mice to make more or less of these polyploidy promoters.

Published by AAAS

For example, biochemist Katya Ravid of Boston University School of Medicine and colleagues enhanced polyploidy to test its role in megakaryocytes, hefty immune cells that dwell in the bone marrow and generate the platelets that help stanch bleeding. Megakaryocytes often harbor more than 100 copies of their genome, and researchers conjectured that the extra genes help the cells crank out platelets.

In 2010, Ravid's team engineered mice to manufacture excess amounts of a polyploidy-promoting protein. Although the alteration boosted the number of chromosome sets the cells contained, it didn't cause a corresponding rise in platelet numbers, the team revealed in *The Journal of Biological Chemistry*. Ravid suggests that polyploidy instead benefits megakaryocytes by boosting production of proteins that the cells need for structural support and sticking to their neighbors.

Bulking up

Biophysical engineer Dennis Discher of the University of Pennsylvania School of Medicine offers another explanation. He suspects that polyploidy helps a megakaryocyte in the same way a high-calorie diet helps a sumo wrestler—by increasing bulk. A membrane perforated by small pores separates the bone marrow from the bloodstream, and a megakaryocyte has to stay on the bone marrow side. Discher and his colleagues recently examined what size pores different types of bone marrow cells could slip through, and they found that megakaryocytes had trouble squeezing through even the largest openings, probably because of their chromosome-packed nuclei. "If you ask me why this cell is polyploid, I'd say it helps anchor the body of the cell in the marrow," says Discher, whose team reported its findings in the 19 November 2013 issue of the *Proceedings of the National Academy of Sciences*.

Researchers already have evidence from other species that extra heft is a benefit of polyploidy. In a 2012 study of fruit flies, Terry Orr-Weaver of the Massachusetts Institute of Technology and her colleague Yingdee Unhavaithaya found that when they reduced the levels of a polyploidy-stimulating protein in cells forming the blood-brain barrier in flies, the cells shrank and the barrier became leaky. The pair also showed that enlarging the undersized cells restored a tight seal. Boosting the size of existing cells might cause less disruption than producing more cells through division, which requires that a cell disengage from its neighbors, notes cell biologist Brian Calvi of Indiana University, Bloomington.

Yet for one mammalian cell type that takes polyploidy to the extreme, work by Leone's team downplays the size connection. Cells in the outer layer of embryos, known as trophoblast giant cells, are polyploidy champions—in mice they pack up to 1000 genome copies. The cells help the embryo implant in its mother's womb, and researchers have suggested that adding chromosomes allows the cells to quickly enlarge, enabling the embryo to infiltrate the uterine lining.

Leone and his colleagues deleted genes for polyploidy-promoting proteins from trophoblast giant cells in mice, anticipating that embryos would die because implantation would suffer. "We were expecting that polyploidy is really significant," Leone says. Although the giant trophoblast cells were smaller than normal and carried fewer chromosomes, the mouse embryos lived and grew up into seemingly healthy adults, the researchers reported in *Nature Cell Biology* in 2012.

Deep reserves

For the heart and the liver, two hard-working organs that also teem with polyploid cells, researchers are exploring a different explanation for polyploidy: The extra chromosomes boost performance under trying conditions and increase overall resilience. Indeed, "polyploidy may be an important stress response or adaptation" for many cell types, says cell biologist Donald Fox of Duke University Medical Center in Durham, North Carolina.

Support for that notion comes from a study of the mouse heart, in which almost all the cells sport four sets of chromosomes. In 2010, stem cell biologist Thomas Braun of the Max Planck Institute for Heart and Lung Research in Bad Nauheim, Germany, and colleagues examined genetically altered mice whose muscle cells—including those in the heart—were missing a gene that spurs polyploidy. Although the gene's absence didn't make all the animals' heart cells diploid, it did reduce the number of chromosome sets they contained by about one-third.

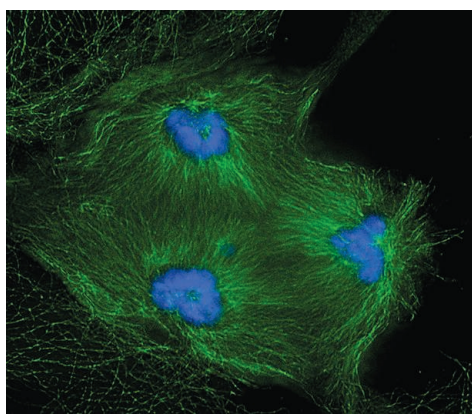
"At baseline conditions, they are pretty normal," Braun says of the mice. However, deficiencies appeared when the rodents had to cope with setbacks such as a heart attack. The hearts of animals with reduced polyploidy pumped less blood after an induced heart attack than did the hearts of control animals, the group reported in *Circulation Research*. How polyploidy enables the heart to rebound remains unclear, Braun says.

The work by Duncan and his colleagues on liver cells also backs the stress-response hypothesis. Unlike most mammalian organs, the liver has a remarkable ability to regenerate after injury. The liver is also well stocked with polyploid cells: In humans, about 50% of the liver cells called hepatocytes carry extra sets of chromosomes.

Duncan's team had originally explored

POLYPLOID CELL TYPES IN MAMMALS

CELL TYPE	LOCATION	FUNCTION	NUMBER OF GENOME COPIES
Megakaryocyte	Bone marrow	Producing blood clotting platelets	Up to 128
Hepatocyte	Liver	Detoxification, metabolism	Typically 4 to 16
Trophoblast giant cell	Embryo	Promote implantation	Up to 1000
Cardiomyocyte	Heart	Contraction	Typically 4



Bonus DNA. The polyploid cells in mammalian bodies differ in their location, function, and number of chromosome sets (table). In a liver cell (*right*), the multiple chromosome copies (blue) have sorted into three clusters in preparation for cell division.

CREDIT: (RIGHT) ANDREW W. DUNCAN AND MARKUS GROMPE

whether the polyploid cells within the liver are "terminally differentiated," meaning that they had become mature cells that don't divide and help replenish the organ. So the team transplanted polyploid hepatocyte cells into mice whose livers had been partially removed. "To our great surprise, they regenerated the liver perfectly," Duncan says.

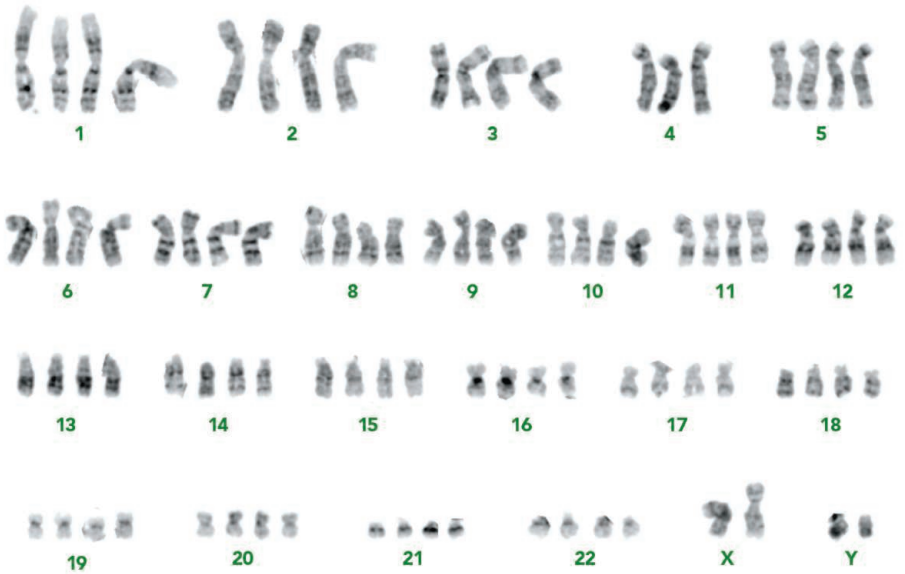
That's when Duncan put the liver cells under the microscope, turned on the camera, and noticed their unorthodox division style. The researchers discovered something else unusual about the cells. As the team revealed in *Nature* in 2010, when many of the polyploid cells divided, they spawned diploid daughter cells. But often these diploid daughters hadn't quite returned to normal—many of them had gained or lost an individual chromosome, a condition called aneuploidy that is generally considered ominous. "Most cancer folks will tell you that aneuploidy is synonymous with cancer," Duncan says.

But some researchers have proposed that aneuploidy can create useful genetic diversity in a tissue or organ, allowing cells to add a copy of a beneficial gene or throw out a copy of a detrimental one. And when Duncan and colleagues studied an example of liver regeneration in mice, they found that sites where regrowth occurred were rich in aneuploid cells. They have discovered that aneuploid cells are abundant in human livers, too.

Duncan now hypothesizes that polyploidy in the liver is a roundabout way to produce aneuploid cells that have regenerative properties. His team is now working to confirm that these cells spur regeneration in people suffering from hepatitis B, in which a virus devastates the liver. Some patients die unless they get a liver transplant, but others survive as sections of the organ regenerate. The researchers are collecting tissue samples to determine if areas of the liver that regrow are high in aneuploid cells.

But the idea that polyploidy helps tissues regenerate remains a hypothesis, as findings from Leone's group and that of Alain de Bruin, a pathologist and veterinarian at Utrecht University in the Netherlands, emphasize. They genetically engineered mice so the animals' livers lack two polyploidy-promoting proteins. "We can generate a mouse whose liver is almost entirely [diploid] cells," De Bruin says. Both teams expected that the animals would suffer ill effects. Instead, the mice were vigorous, each group reported in *Nature Cell Biology* in 2012, and their livers were no less able to regrow after injury. The mice De Bruin and colleagues studied, for example, could restore their livers after surgical removal of two-thirds of the organ. "This polyploidization does not have an effect on regeneration or on proliferation rate," De Bruin says.

The work from both teams also undermines another older polyploidy hypothesis. The liver, De Bruin notes, "is all the time exposed to toxins." Hepatocytes work hard to detoxify all those noxious substances, and some researchers had speculated that their extra genetic material could boost the output of proteins crucial to this. Yet the mice whose livers had reduced polyploidy had no problems breaking down toxins, De Bruin's group found.



Double down. The chromosome copies from a polyploid liver cell arranged by size, showing that the cell carries four copies of almost every one.

> "The real unanswered question is why any cell type is polyploid. We are poised to begin answering that question."
>
> —Robert Duronio, University of North Carolina, Chapel Hill

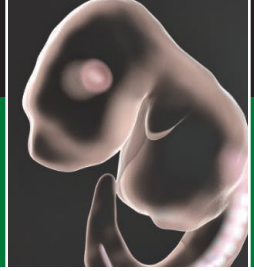
Exploiting polyploidy

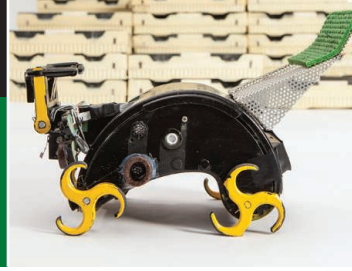
Even as they wrestle with mystery of polyploidy, researchers wonder whether they can put what they've learned to use. Leukemia biologist John Crispino of Northwestern University's Feinberg School of Medicine in Chicago, Illinois, and his colleagues have trained their sights on a type of acute myeloid leukemia, triggered by megakaryocytes, that kills most adults who develop it. Mature megakaryocytes don't divide, but in this form of cancer, the cells remain immature, don't become polyploid, and replicate prodigiously, causing the leukemia. Crispino and colleagues propose that forcing the cells to become polyploid and mature might treat the cancer.

The team revealed in *Cell* in 2012 that it had identified more than 200 compounds that, in lab dishes, spur polyploidy in human megakaryocytes. One of these molecules, alisertib, is already under-going clinical trials for several other types of cancer—though not because of its ability to stimulate polyploidy. Crispino's group is now trying to organize an initial safety trial of the drug in people with acute myeloid leukemia.

Although polyploidy research has recorded some progress in recent years, the field still hasn't nailed down the benefits polyploidy provides to different mammal cell types. To move forward, Leone says, researchers should take a cue from plant biologists, who have tested polyploidy's advantages in specific environmental conditions, showing that it boosts tolerance for salinity (*Science*, 9 August 2013, p. 658). Scientists could perform similar studies on liver cells, for example, by gauging whether polyploidy helps them deal with different diets. Delving further into polyploidy's cellular roles will probably produce some surprises, UNC Chapel Hill's Duronio predicts. "There are going to be many uses for polyploidy, and we are just scratching the surface."

–MITCH LESLIE





LETTERS

edited by Jennifer Sills

Global Warming and Winter Weather

IN MID-JANUARY, A LOBE OF THE POLAR VORTEX SAGGED SOUTHWARD OVER THE CENTRAL and eastern United States. All-time low temperature records for the calendar date were set at O'Hare Airport in Chicago [–16°F (–8°C), 6 January], at Central Park in New York [4°F (–15.6°C), 7 January], and at many other stations (*1*). Since that event, several substantial snow storms have blanketed the East Coast. Some have been touting such stretches of extreme cold as evidence that global warming is a hoax, while others have been citing them as evidence that global warming is causing a "global weirding" of the weather. In our view, it is neither.

As climate scientists, we share the prevailing view in our community that human-induced global warming is happening and that, without mitigating measures, the Earth will continue to warm over the next century with serious consequences. But we consider it unlikely that those consequences will include more frigid winters.

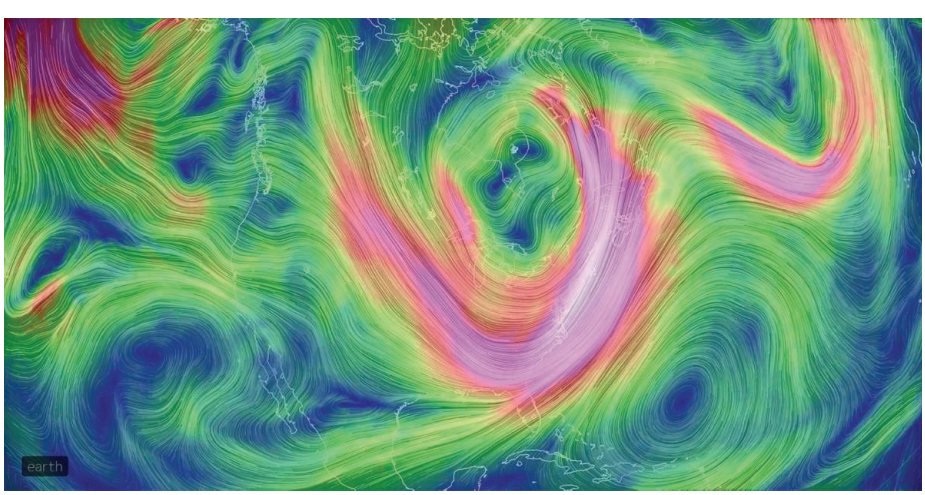
Distinguishing between different kinds of extreme weather events is important because the risks of different kinds of events are affected by climate change in different ways. For example, a rise in global mean temperature will almost certainly lead to an increase in the incidence of record high temperatures. Global warming also leads to increases in atmospheric water vapor, which increases the likelihood of heavier rainfall events that may cause flooding. Rising temperatures over land lead to increased evaporation, which renders crops more susceptible to drought. As the atmosphere and oceans warm, sea water expands and glaciers and ice sheets melt. In response, global sea-level rises, increasing the threat of coastal inundation during storms.

In contrast to the above examples, the notion that the demise of Arctic sea ice during summer should lead to colder winter weather over the United States seems counterintuitive. But that is exactly what an influential study has suggested (*2*). The authors hypothesize that global warming could perturb the polar vortex in a manner that renders the flow around it more wavy, leading to an increased incidence of both extreme warmth and extreme cold in temperate latitudes. It's an interesting idea, but alternative observational analyses and simulations with climate models have not confirmed the hypothesis, and we do not view the theoretical arguments underlying it as compelling [see (*3*–*6*)].

Other studies have suggested that the loss of Arctic sea ice may influence the atmospheric circulation in mid-latitudes during summer [e.g., (*7*)]. Sea-ice losses during late summer may indeed lead to regional changes in Arctic climate [e.g., (*5*, *8*)]. But tremendous natural variability occurs in the large-scale atmospheric circulation during all seasons, and even in summer, the links between Arctic warming and mid-latitude weather are not supported by other observational studies (*6*). The lag between decreases in sea-ice extent during late summer, and changes in the mid-latitude atmospheric circulation during other seasons (when the recent loss of sea ice is much smaller) needs to be reconciled with theory.

Summertime sea-ice extent in the Arctic has been remarkably low since 2007, and the ensuing years have been marked by some notable cold air outbreaks. It was this coincidence that prompted Francis and Vavrus (*2*) to link the cold air outbreaks to global warming. But coincidence does not in itself constitute a strong case for causality. Cold air outbreaks even more severe than occurred this winter affected the United States in the early 1960s, the late 1970s (most notably 1977), and in 1983, back when the Arctic sea ice was thicker and more extensive than it is today [e.g., (*9*)]. Over the longer time span of 50 to 100 years, it is well established that there has been a decrease in the rate at which low temperature records are being set relative to all-time high temperature records at stations across the United States (*10*). For the present at least, we believe that statistics based on the longer record are more indicative of what the future is likely to bring.

The research linking summertime Arctic sea ice with wintertime climate over temperate latitudes deserves a fair hearing. But to make it the centerpiece of the public discourse



CREDIT: FIGURE GENERATED BY CAMERON BECCARIO (EARTH.NULLSCHOOL.NET); RESULTS SOURCED FROM THE NCEP/NOAA GLOBAL FORECAST SYSTEM

Icy blast. Arctic winds flowed down to North America in January, causing record-breaking cold temperatures. Image shows streamlines of wind at the 500 mbar level at 1:00 a.m. Eastern Standard Time on 7 January 2014. Red indicates faster speeds.

Published by AAAS

on global warming is inappropriate and a distraction. Even in a warming climate, we could experience an extraordinary run of cold winters, but harsher winters in future decades are not among the most likely nor the most serious consequences of global warming.

JOHN M. WALLACE,[1]* ISAAC M. HELD,[2] DAVID W. J. THOMPSON,[3] KEVIN E. TRENBERTH,[4] JOHN E. WALSH[5]

[1]Department of Atmospheric Sciences, University of Washington, Seattle, WA 98195, USA. [2]NOAA Geophysical Fluid Dynamics Laboratory, Princeton University Forrestal Campus, Princeton, NJ 08540–6649, USA. [3]Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523, USA. [4]Climate Analysis Section, National Center for Atmospheric Research, Boulder, CO 80307, USA. [5]International Arctic Research Center, University of Alaska Fairbanks, Fairbanks, AK 99775–7340, USA.

*Corresponding author. E-mail: wallace@atmos.washington.edu

References

1. National Weather Service (www.weather.gov).
2. J. A. Francis, S. J. Vavrus, *Geophys. Res. Lett*. **39**, L06801 (2012).
3. E. A. Barnes, *Geophys. Res. Lett*. **40**, 4734 (2013).
4. J. Screen, I. Simmonds, *Geophys. Res. Lett*. **40**, 1 (2013).
5. J. A. Screen, C. Deser, I. Simmonds, R. Tomas, *Clim. Dyn*. 10.1007/s00382-013-1830-9 (2013).
6. E. A. Barnes, E. Dunn-Sigouin, G. Masato, T. Woollings, *Geophys. Res. Lett*. **41**, 10.1002/2013GL058745 (2014).
7. Q. Tang, X. Zhang, J. A. Francis, *Nat. Clim. Change* **4**, 45 (2014).
8. M. A. Balmaseda, L. Ferranti, F. Molteni, T. N. Palmer, *Q. J. R. Meteorol. Soc.* **136**, 1655 (2010).
9. M. P. Cellitti, J. E. Walsh, R. M. Rauber, D. H. Portis, *J. Geophys. Res.* **111**, D02114 (2006).
10. T. C. Peterson *et al*., *Bull. Am. Met. Soc*. **94**, 821 (2013).

The Big Picture for Big Data: Visualization

TO CONVERT INFORMATION FROM MASSIVE data sets into insights, data centers will need to support the humans who are trying to make sense of it all. Fortunately, innovations in information visualization are demonstrating that a good user interface is worth a thousand petabytes (2013 Visualization Challenge, News, 7 February, p. 600).

When GE Healthcare researcher Nick Thomas studied a visualization of the critical RBP1 protein—a genomic carrier of vitamin A necessary for reproduction and vision—he was surprised by what he saw. Thomas scanned the mosaic grid of thousands of red and green dots, as well as the linked scattergram and color-coded plate view. He confirmed expected patterns, but one unexpected bright red dot revealed RBP1's marked influence in cellular development. This clue gave Thomas an insight that, with statistical confirmation, led to an important scientific contribution (*1*).

Like a growing number of researchers, policy-makers, and interested citizens, Thomas was exploring increasingly complex data sets by adjusting filters, changing color palettes, and choosing novel visualizations to search for relationships, clusters, gaps, and outliers. Powerful information visualization tools are realizing famed statistician John Tukey's 50-year-old prediction: "The graphical potentialities of the computer...are going to be the data analyst's greatest single resource" (*2*).

Some Big Data advocates seem to promise automatic results with little human participation [e.g., (*3*, *4*)]. A more effective approach will be to put human users in control, since they can often identify patterns that machines cannot. Statistically and algorithmically oriented researchers are increasingly recognizing that visual strategies for exploring complex data lead to more potent and meaningful insights. Automated analyses can work for well-understood data, but visualizations increase the efficacy of experts in frontier topics, where big breakthroughs happen.

BEN SHNEIDERMAN

Department of Computer Science, University of Maryland, College Park, MD 20742, USA. E-mail: ben@cs.umd.edu

References

1. S. Stubbs, N. Thomas, *Methods Enzymol*. **414**, 1 (2006).
2. J. W. Tukey, *Am. Statistician* **19**, 23 (1965).
3. C. Anderson, *Wired Mag.* **16**, 7 (2008).
4. U. M. Feyyad, *IEEE Expert*. **11**, 20 (1996).

Innovation Goes Global

BY TAKING A U.S.-NATIONAL APPROACH TO innovative capabilities and comparing the present with the postwar period, W. B. Bonvillian ("Advanced manufacturing policies and paradigms for innovation," Policy Forum, 6 December 2013, p. 1173) ignores the truly transformational change that has occurred over the past several decades: the growth of the global science system. The critical knowledge needed to innovate into the next generation of production is increasingly distributed across the globe, and it is just as likely to be located in India or China as in Ohio. The Organization for Economic Co-Operation and Development reports that the growth in the number of triadic patents demonstrates the worldwide spread of innovative activities (*1*).

U.S. researchers are actively tapping this global resource by collaborating with researchers from many other countries. The global network of international links (drawn from coauthorships on publications) has tripled in density over the past 20 years (*2*), with many new members joining the global network from developing countries, particularly China. Chinese addresses now appear more frequently than any other country in publications coauthored with U.S. researchers.

Scientific globalization does not threaten an end to U.S. excellence in innovation; quite the opposite. The diffusion and rooting of scientific capacity to new places provides opportunity for greater efficiency in research activities, particularly by removing redundancy. Creative problem-solving can be enhanced by having new entrants grapple with technological challenges, as many U.S. companies are finding as they invest in foreign research.

Culturally tied knowledge is often important to market access in foreign countries. These goods require a deliberate policy shift on the part of U.S. agencies from pushing knowledge creation to fomenting knowledge scanning and integration. Scanning the globe for the best new knowledge and ensuring local uptake is the more promising approach to closing the gaps in U.S. know-how than building a U.S.-only R&D effort, as Bonvillian suggests.

CAROLINE S. WAGNER

Battelle Center for Science & Technology Policy, John Glenn School of Public Affairs, The Ohio State University, Columbus, OH 43210, USA. E-mail: wagner.911@osu.edu

References

1. OECD, *OECD Science, Technology and Industry Scoreboard 2013* (OECD Publishing, Paris, 2013).
2. L. Leydesdorff, C. Wagner, *J. Informetrics* **2**, 317 (2008).

CORRECTIONS AND CLARIFICATIONS

Perspectives: "Hiding in plain view—An ancient dog in the modern world" by H. G. Parker and E. A. Ostrander (24 January, p. 376). In the figure, the red branch should have been labeled "CTVT." The HTML and PDF versions online have been corrected.

Reports: "Transmissible dog cancer genome reveals the origin and history of an ancient cell lineage" by E. P. Murchison *et al.* (24 January, p. 437). In the title, "Transmissable" should have been "Transmissible." The HTML and PDF versions online have been corrected.

Reports: "Identification of a plant receptor for extracellular ATP" by J. Choi *et al.* (17 January, p. 290). The doi should be 10.1126/science.343.6168.290. It is correct in the HTML and PDF versions online.

Research Article: "The hidden geometry of complex, network-driven contagion phenomena" by D. Brockmann and D. Helbing (13 December 2013, p. 1337). In Fig. 2D, the label "Zamonia" should have read "Latvia." The HTML and PDF versions online have been corrected.

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the past 3 months or matters of general interest. Letters are not acknowledged upon receipt. Whether published in full or in part, Letters are subject to editing for clarity and space. Letters submitted, published, or posted elsewhere, in print or online, will be disqualified. To submit a Letter, go to www.submit2science.org.

DEVELOPMENT

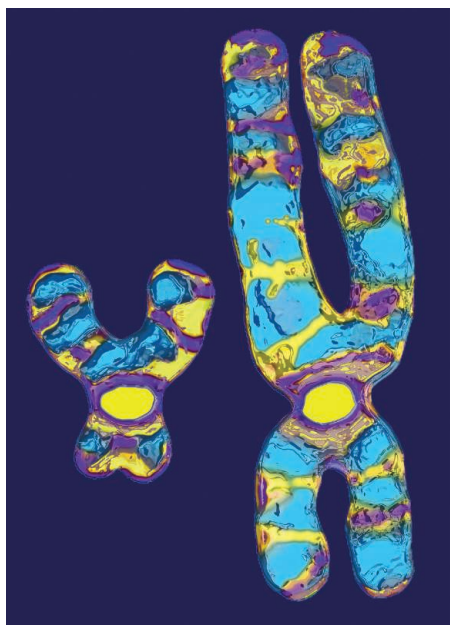
YseX Is a Matter of Concern Rather Than a Matter of Fact

Amade M'charek

The self-evidence and power of the X and Y chromosomes in science and society cannot be overestimated. As a binary couple they line up with other familiar biological categories such as eggs and sperm or estrogen and androgen, and increasingly they've come to stand for females and males. However, what if one did not take them as a matter of fact and instead asked how X and Y came to stand for female and male. What does it take to sex these chromosomes? Such questions refer not so much to the bodies in which these chromosomes are found but to the scientific practices that study them. Making the work of science visible, demonstrating how morals and values are part and parcel of the epistemology of science, means understanding the objects of science as "matters of concern" (*1*)—objects that require care and deserve density.

Historian of science Sarah S. Richardson (Harvard University) has taken this demonstration as her very task. In her erudite and well-balanced *Sex Itself*, she "examines the interaction between cultural gender norms and genetic theories of sex from the beginning of the twentieth century to the present postgenomic age." Richardson takes issue with the perpetual reductionist view on sex differences. Perplexed by the suggestion made in 2005 that genetic differences between men and women are larger than those between humans and chimpanzees, she meticulously demonstrates how the genetics of sex has been modeled on alleged and often rehearsed gender distinctions between men and women. But she does more. Richardson skillfully demonstrates how instrumental sex differences have been in the development of genetics. For example, she shows how the sex chromosomes were a key aspect in developing the chromosomal theory of inheritance and paving the way for experimental studies of gene mutations and genomic organization. The book includes several case studies that help us to understand the history of genetics in general.

Sex Itself consists of four parts. In three early chapters, Richardson examines the recognition of sex chromosomes in the early 20th century and their emergence in the context of chromosome mapping and the Mendelian theory of inheritance. She surveys the terminology (such as heterochromosomes, accessory chromosome, and sex chromosome) and makes visible the guardedness of geneticists to reducing sex determination to these chromosomes. Thomas Hunt Morgan, from whom she derives her title, preferred to say "sex factors on the chromosomes." She then discusses how the arrival of sex hormones provided sex chromosomes with a strong ally. They allowed for "a simple two-tiered model of sexual development, with genes as the initiators and sex hormones as the dominant agent of sexual differentiation." Richardson argues that hormone science thus "helped to constitute and solidify the 'sex chromosome.'"



CREDIT: CAVALLINI JAMES/© BSIP SA/ALAMY

Sex Itself
The Search for Male and Female in the Human Genome

by Sarah S. Richardson
University of Chicago Press, Chicago, 2013. 319 pp. $45, £31.50. ISBN 9780226084688.

Richardson opens her second line of argument by analyzing the gender stereotypes that contributed to the sexing of the chromosomes and the identification of the Y with males and the X with females. While intelligent men such as Brian Sykes and Craig Venter ascribe much power to their Y chromosomes, "the vessel of manhood" (*2*), by contrast Richardson locates that power in a deeply gendered history of genetics. To that end, she examines a classic case of scientific error, the so-called "XYY supermale syndrome" of the 1960s and 1970s. Based on the speculation that the Y chromosome might contain some male traits, XYY males were viewed as having a double dose of maleness. Flawed research characterized these men as more sexual and aggressive than men who carried a single Y. Thus while XX and XY stood for the phenotypic difference between females and males, XYY hinted at the behavior of males. This case mirrors X-mosaicism theories of female biology and behavior, which Richardson takes up next.

X-mosaicism has been cast as revealing huge differences between males and females and as explaining ascribed behavioral traits of women (e.g., complicated, inconsistent, and unpredictable). Whereas mosaic X inactivation is constantly presented as a necessary essential female trait (for example, in studies of the incidence of autoimmunity), Richardson argues that these studies fail to take the biological context into account. Doing so reveals that X inactivation does not make females more female, but more like males. It "serves to equalize X dosage between males and females, so that the cells of *both* sexes are functionally monosomic for X-linked genes."

In her third argumentative strand, Richardson raises important questions of whether and how feminist scholarship has contributed to the science of sex. The case of sex determination proves to be an excellent example. Richardson reports on the race to locate the male sex-determining gene on the Y chromosome and the subsequent growing dismay that the *SRY* gene may not be in control. Rejecting the assumption that females were the result of a passive sex-determining pathway, feminist scientists began in the 1990s to critique the notion of the "master gene" as well as Y chromosome–centered research in sex determination. Jennifer Graves and other leading researchers argued that a gendered view such as the "dominant Y" with masculine qualities had geneticists believing that *SRY* is an activator and ignoring the fact that it could be an inhibitor or "a spoiler that turns off genes" (*3*). By the early 2000s, the notion of the master gene faded away to make room for the complexity of sex determination and the view that both male and female pathways played active roles therein. As Richardson argues, these shifts cannot be fully explained by feminist interventions, but gender criti-

The reviewer is at the Department of Anthropology, University of Amsterdam, OZ Achterburgwal 185, 1012 DK Amsterdam, Netherlands. E-mail: a.a.mcharek@uva.nl

Published by AAAS

cism did actively interact with the research on sex determination.

In a contrasting case of the "vanishing Y" and the debate over whether the Y chromosome is gene-rich or almost dying out, Richardson zooms in on the ways gendered notions do not merely play a role in media debate. They are part and parcel of the research questions, design, and knowledge that comes out of laboratories. As there is no way of being gender neutral, we had better acknowledge and critically reflect on gender in genetics.

In the concluding chapters, Richardson attends to the genomic era and its effect on sex science. She offers reflections on the previous cases and prognoses focused on potential risks involved in the genomization of sex research. Like many scholars studying the social aspects of genomics, she voices concerns about the ways genomic science is reifying differences among people. She quite correctly notes that these concerns have been especially attended to in studies of the implications of genomics on race and racism, whereas little ink has been spilled on the reification of sex differences. Thus the urgent need for *Sex Itself*. Not simply an account of the effect of gender on genetics, it provides us with tools to think of the possibility of a gender-critical genetics.

References

1. B. Latour, *What Is the Style of Matters of Concern* (Van Gorcum, Assen, Netherlands, 2008).
2. S. Jones, *Y: The Descent of Man* (Little, Brown, London, 2002).
3. J. A. Graves, *Biol. Reprod.* **63**, 667 (2000).

10.1126/science.1249293

ENVIRONMENTAL LAW

The Case for a Public "Trust"

Rena Steinzor

Heaven only knows it is way past time for us to stop talking about whether climate change is the gravest threat that has ever confronted humankind while the shrill voices of a tiny minority insist nothing is wrong. Instead, the governments of all countries should be engaged in the most serious possible debate over how to mitigate and adapt to the effects humans have had on climate. The reasons that we have not reached this crucial phase are complicated. At the very least, they involve the myopic entitlement of developed nations and developing nations' desperate efforts to catch up. Our paralysis arises from the limits of our psychology and the constraints of our political systems; both seem entrenched in their commitment to preserving stability by maintaining the status quo. To the extent, though, that the United States legal system in and of itself frustrates an effective response, Mary Christina Wood's *Nature's Trust* makes a discrete contribution to the search for climate change solutions by rejecting dominant paradigms out of hand. Instead Wood (University of Oregon School of Law) urges the courts to pick up the isolated, tenuous threads of the "public trust" doctrine and use them to compel the executive and legislative branches to embrace the idea that all natural resources (including Earth's atmosphere) cannot be used in any way that exacerbates climate change.

The public trust, envisioned as the pooled ownership over natural resources possessed by all the country's citizens regardless of previous concepts of private property, would reject any land use that does not preserve the ability of nature to replenish itself. The doctrine would drive private rights into the background, supplanting them on the rationale that cutting back human production of greenhouse gases is of overriding importance and that nature does not have the capacity to safely absorb the quantities we have already emitted.

The nature's trust is the antithesis of what the author dubs "predatory capitalism." If, as Wood argues, Congress, the Executive Branch, polluting industries, and national environmental groups are hopelessly corrupt, the advancement of the doctrine would depend on a widespread grassroots rebellion and the independent thinking of the courts. The author is optimistic about the development of a mass movement because "assertions of commonwealth thinking now appear across the United States" in the form of "community gardens, inner-city farms, and urban homesteads." She acknowledges that it will take "enormous numbers of citizens to grow these seeds of change into a land revolution so strong that it displaces the market-driven system of land exploitation." But, she assures us, "those who truly cherish private property rights will find their calling in this land-as-commonwealth frame, as they will come to learn that their liberty and quiet enjoyment of land depends, first and foremost, on Earth's life-sustaining ecological endowment."

I must confess to wondering about the people who are unlikely to renounce private property: those grown wealthy and comfortable on the basis of land exploitation. Presumably, they will be overwhelmed and then reformed by everyone else, including those who have never felt secure enough to own anything. But because Wood is a zealot in the best sense of the word—possessing enormous energy, passion, and conviction that she has discovered the one true path forward—she does not dwell for long on the efforts the former will make to resist her. Nor does she acknowledge in any realistic way the privations we would endure to make the radical transition she envisions. Of course, she might respond to these critiques by asking whether the specter of unmitigated climate change is acceptable—and, of course, it is not. Yet the polarization of these alternatives undermines the book's credibility. How a mass movement would be sustainable if fed only by Wood's idealism, without preparing for the sacrifices that are inevitable, is far from clear.

Nature's Trust is hefty, running over 450 pages, and its primary value will be to lawyers who need a compendium of legal precedents to help them formulate test cases. A second audience might be composed of people who are not quite convinced about the vagaries of government regarding this issue. Wood effectively stokes rage against those in power—a white-hot phenomenon akin to road rage, at least with respect to Congress and the expansive Executive Branch. Oddly, she seems to have more confidence in the courts to adopt the nature's trust position, even though federal judges are not elected by anyone, including the enormous numbers of citizens she hopes will see the light.

Nonetheless, as jacket blurbs by Bill McKibben, James Hansen, and Ross Gelbspan express quite well, *Nature's Trust* is both ambitious and original. For anyone interested in using the legal system to prod action, Wood has made a major contribution.

10.1126/science.1249069

Nature's Trust:
Environmental Law for a New Ecological Age

by Mary Christina Wood
Cambridge University Press, New York, 2013. 460 pp. $115, £65. ISBN 9780521195133. Paper, $45, £27.99. ISBN 9780521144117.

The reviewer is at the University of Maryland Carey Law School, 500 West Baltimore Street, Baltimore, MD 21201, USA. E-mail: rstein@law.umaryland.edu

ENERGY AND ENVIRONMENT

Methane Leaks from North American Natural Gas Systems

Methane emissions from U.S. and Canadian natural gas systems appear larger than official estimates.

A. R. Brandt,[1]*‡ G. A. Heath,[2] E. A. Kort,[3] F. O'Sullivan,[4] G. Pétron,[5,6] S. M. Jordaan,[7] P. Tans,[5] J. Wilcox,[1] A. M. Gopstein,[8]† D. Arent,[2,9] S. Wofsy,[10] N. J. Brown,[11] R. Bradley,[12] G. D. Stucky,[13] D. Eardley,[13] R. Harriss[14]

Natural gas (NG) is a potential "bridge fuel" during transition to a decarbonized energy system: It emits less carbon dioxide during combustion than other fossil fuels and can be used in many industries. However, because of the high global warming potential of methane (CH_4, the major component of NG), climate benefits from NG use depend on system leakage rates. Some recent estimates of leakage have challenged the benefits of switching from coal to NG, a large near-term greenhouse gas (GHG) reduction opportunity (*1*–*3*). Also, global atmospheric CH_4 concentrations are on the rise, with the causes still poorly understood (*4*).

To improve understanding of leakage rates for policy-makers, investors, and other decision-makers, we review 20 years of technical literature on NG emissions in the United States and Canada [see supplementary materials (SM) for details]. We find (i) measurements at all scales show that official inventories consistently underestimate actual CH_4 emissions, with the NG and oil sectors as important contributors; (ii) many independent experiments suggest that a small number of "superemitters" could be responsible for a large fraction of leakage; (iii) recent regional atmospheric studies with very high emissions rates are unlikely to be representative of typical NG system leakage rates; and (iv) assessments using 100-year impact indicators show system-wide leakage is unlikely to be large enough to negate climate benefits of coal-to-NG substitution.

[1]Stanford University, Stanford, CA. [2]National Renewable Energy Laboratory, Golden, CO. [3]University of Michigan, Ann Arbor, MI. [4]Massachusetts Institute of Technology, Cambridge, MA. [5]National Oceanic and Atmospheric Administration, Boulder, CO. [6]University of Colorado, Boulder, CO. [7]University of Calgary, Calgary, Canada. [8]U.S. Department of State, Washington, DC. [9]Joint Institute for Strategic Energy Analysis, Golden, CO. [10]Harvard University, Cambridge, MA. [11]Lawrence Berkeley National Laboratory, Berkeley, CA. [12]Independent consultant, Gaithersburg, MD. [13]University of California, Santa Barbara, Santa Barbara, CA. [14]Environmental Defense Fund, Boulder, CO.

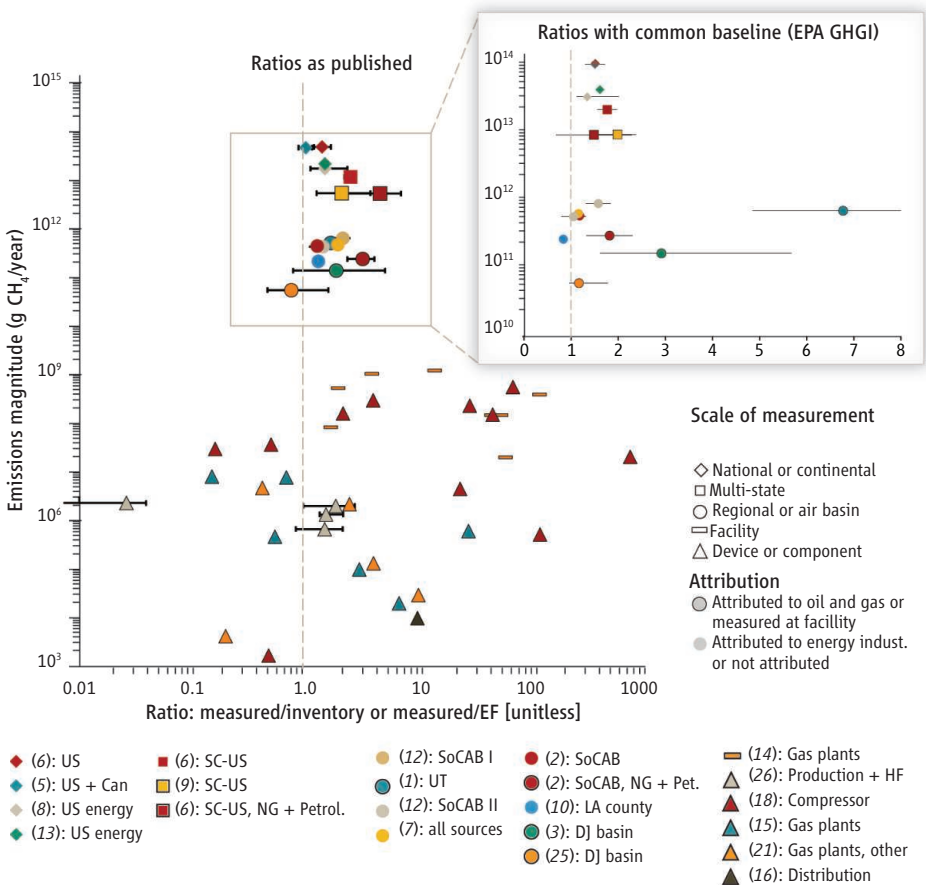
*Full affiliations for all authors are included in the supplementary materials. †Views expressed in this article are those of the author and do not necessarily reflect those of the U.S. Department of State or the U.S. Government. ‡Corresponding author. abrandt@stanford.edu

Underestimation—Device to Continent

This study presents a first effort to systematically compare published CH_4 emissions estimates at scales ranging from device-level (>10^3 g/year) to continental-scale atmospheric studies (>10^{13} g/year). Studies known to us that (i) report measurement-based emissions estimates and (ii) compare those estimates with inventories or established emission factors (EFs) are shown in the first chart.

Studies that measure emissions directly from devices or facilities ("bottom-up" studies) typically compare results to emissions factors (EFs; e.g., emissions per device). Large-scale inventories are created by multiplying EFs by activity factors (e.g., number of devices).

Studies that estimate emissions after atmospheric mixing occurs ("atmospheric" studies) typically compare measurements to emissions inventories, such as the U.S. Envi-



Inventories and emissions factors consistently underestimate actual measured CH_4 emissions across scales. Ratios >1 indicate measured emissions are larger than expected from EFs or inventory. Main graph compares results to the EF or inventory estimate chosen by each study author. Inset compares results to regionally scaled common denominator (*17*), scaled to region of study and (in some cases) the sector under examination. Multiple points for each study correspond to different device classes or different cases measured in a single study. Definitions of error bar bounds vary between studies. (US, United States; Can, Canada; SC, South Central; Petrol. and Pet., petroleum; SoCAB, South Coast Air Basin; LA, Los Angeles; DJ, Denver-Julesberg; UT, Utah; HF, hydraulic fracturing). See SM for figure construction details.

Published by AAAS

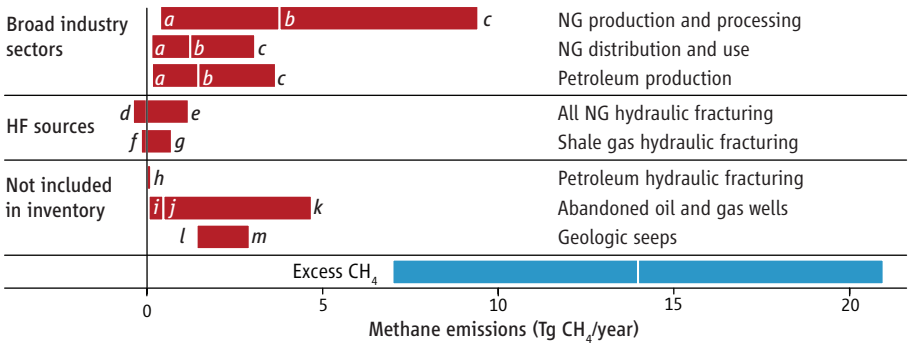
ronmental Protection Agency (EPA) national GHG inventory (GHGI). Atmospheric studies use aircraft (*1*, *5*–*8*), tower (*3*, *6*), and ground (*3*, *7*–*10*) sampling, as well as remote sensing (*7*, *11*, *12*). All such studies observe atmospheric concentrations and must infer fluxes by accounting for atmospheric transport. The various inference methods have strengths and weaknesses (see SM). The greatest challenge for atmospheric studies is attributing observed CH_4 concentrations to multiple potential sources (both anthropogenic and natural).

Results from bottom-up studies (generally $<10^9$ g CH_4/year) and atmospheric CH_4 studies at regional scale and larger (above 10^{10} g CH_4/year) are shown in the first chart. We also include studies that do not focus on NG systems, in order to place NG emissions in context with other CH_4 sources. Across years, scales, and methods, atmospheric studies systematically find larger CH_4 emissions than predicted by inventories. EFs were also found to underestimate bottom-up measured emissions, yet emissions ratios for bottom-up studies are more scattered than those observed in atmospheric studies (*13*–*16*).

Regional and multistate studies focusing on NG-producing (*1*–*3*, *9*) and NG-consuming regions (*2*, *7*, *10*–*12*) find larger excess CH_4 emissions than national-scale studies. This may be due to averaging effects of continental-scale atmospheric processes, to regional atmospheric studies focusing on areas with other air quality problems (*1*, *3*), or simply to methodological variation. Atmospheric measurements are constrained in spatial and temporal density: Regional studies cover 0.5 to 5% of NG production or consumption with dense measurements, although often limited to short-duration sampling "campaigns" (*3*, *7*); national studies cover wide areas with limited sample density (*6*) (table S5).

To facilitate comparison, the inset in the first chart normalizes atmospheric studies ($>10^{10}$ g CH_4/year) to baselines computed from the most recent (2011) EPA GHGI estimates for the year and region in which study measurements were made (*17*). After normalization, the largest (e.g., national-scale) atmospheric studies ($>10^{12}$ g CH_4/year) suggest typical measured emissions ~1.5 times those in the GHGI (*5*, *6*, *8*, *9*).

Why might emissions inventories be underpredicting what is observed in the atmosphere? Current inventory methods rely on key assumptions that are not generally satisfied. First, devices sampled are not likely to be representative of current technologies and practices (*18*). Production techniques



Potential contributions to total U.S. CH_4 emissions above EPA estimates. EPA estimate in blue, based on central estimate and uncertainty range from large-scale studies from the inset in the first chart. Both NG sources and possible confounding sectors are included. NG production, petroleum production, and NG distribution emissions are based on regional empirical studies (*1*, *2*, *6*), which estimate emissions rates from high-emitting sources but do not estimate prevalence. Scenarios (*a*) to (*c*) correspond to 1, 10, and 25% of gas production or consumption from such high-emitting sources. Ranges (*d*) to (*g*) correspond to estimates for flowback emissions rates during hydraulic fracturing (HF) of all gas wells and shale gas wells, relative to EPA estimates. Ranges (*h*) to (*m*) reflect sources not included in EPA CH_4 inventories but which could be mistaken for NG emissions by chemical or isotopic composition. See SM for details.

are being applied at scale (e.g., hydraulic fracturing and horizontal drilling) that were not widely used during sampling in the early 1990s, which underlies EPA EFs (*18*).

Second, measurements for generating EFs are expensive, which limits sample sizes and representativeness. Many EPA EFs have wide confidence intervals (*19*, *20*). And there are reasons to suspect sampling bias in EFs, as sampling has occurred at self-selected cooperating facilities.

Third, if emissions distributions have "heavy tails" (e.g., more high-emissions sources than would be expected in a normal distribution), small sample sizes are likely to underrepresent high-consequence emissions sources. Studies suggest that emissions are dominated by a small fraction of "superemitter" sources at well sites, gas-processing plants, coproduced liquids storage tanks, transmission compressor stations, and distribution systems (see table S6 and fig. S2). For example, one study measured ~75,000 components and found that 58% of emissions came from 0.06% of possible sources (*21*).

Last, activity and device counts used in inventories are contradictory, incomplete, and of unknown representativeness (*17*, *22*). Data should improve with increased reporting requirements enacted by EPA (*23*, *24*).

Source Attribution in Atmospheric Studies

Does evidence suggest possible sources of excess CH_4 emissions relative to official estimates within the NG sector? A key challenge is attribution of atmospheric observations to sources. Isotopic ratios (*7*, *11*) and prevalence signatures of non-CH_4 hydrocarbons (*3*, *6*–*8*) can be used to attribute emissions to fossil sources rather than biogenic sources. Evidence from regional studies suggests that CH_4 emissions with fossil signatures are larger than expected (*3*, *6*, *7*, *9*, *11*), whereas national-scale evidence suggests a mix of biogenic and fossil sources (*6*). Atmospheric studies that control for biogenic CH_4 sources (*1*, *2*, *7*) are dependent on biogenic source estimation methods that also have high uncertainties (*6*). Natural geologic seeps could confound attribution (see the second chart and SM).

Studies can attribute emissions to liquid petroleum and NG sources rather than coal by sampling in places with little coal-sector activity (*2*, *3*, *6*, *7*, *9*). Attributing leakage to the NG system, as defined by EPA industry sector classifications, is more challenging. Alkane fingerprints may allow attribution to oil-associated NG (*9*), although NG processing changes gas composition, which may complicate efforts to pinpoint leakage sources. Geographic colocation of facilities and sampling, along with geographically isolating wind directions (*2*, *3*, *7*), can allow attribution of emissions to NG subsectors. Without spatial isolation, sector attribution can require assumptions about gas composition that introduce significant uncertainty (*2*, *3*, *25*).

We plotted results of a thought experiment (see the second chart) in which we estimated emissions ranges of selected possible sources within the NG sector, as well as sources that could be mistaken for NG emissions owing to chemical and isotopic signatures. Although such an analysis is speculative given current knowledge, it illustrates ranges of possible source magnitudes.

Published by AAAS

We include in the second chart a range of excess CH_4 from all sources (7 to 21 × 10^{12} g or Tg/year) based on normalized national-scale atmospheric studies from the inset in the first chart. This excess is conservatively defined as 1.25 to 1.75 times EPA GHGI estimates. This estimate is derived from national-scale atmospheric studies and includes all sources of CH_4 emissions: It should not be expected that NG sources are responsible for all excess CH_4.

The scenarios in the second chart for NG production and/or processing, distribution, and petroleum system emissions apply observed leakage rates from the literature that are higher than EPA GHGI estimates (*1*, *2*, *7*). The frequency of such high-emitting practices is unknown, so illustrative prevalence scenarios are plotted: 1, 10, or 25% of activity is represented by high-emitters; the remaining facilities emit at EPA GHGI rates. This evidence suggests that high leakage rates found in recent studies (*1*, *2*, *7*) are unlikely to be representative of the entire NG industry; if this were the case, associated emissions would exceed observed total excess atmospheric CH_4 from all sources.

In general, the wide ranges in the second chart suggest a poor understanding of sources of excess CH_4 and point to areas where improved science would reduce uncertainty. However, hydraulic fracturing for NG is unlikely to be a dominant contributor to total emissions (*26*). Also, some sources not included in the GHGI may contribute to measured excess CH_4, e.g., abandoned oil and gas wells and geologic seeps (see SM).

Policy Challenges and Opportunities

Leakage scenarios in the second chart have implications for decision-making and policy. A key tool for environmental decision-making is life-cycle assessment (LCA), which compares impacts associated with varying methods of supplying a useful product (e.g., kWh of electricity). A key challenge in LCA studies is attribution of emissions from systems that produce two products, such as "gas" wells that also produce hydrocarbon liquids, or "oil" wells that also produce NG. This challenge is complicated by incongruence between LCA methodology and EPA sector definitions (see SM).

Recent LCAs have estimated GHG emissions from NG use in power generation and transport (see SM). LCA studies generally agree that replacing coal with NG has climate benefits (*27*). However, LCAs have relied heavily on EPA GHGI results. Updating these assessments with uncertainty ranges from the second chart (see SM) still supports robust climate benefits from NG substitution for coal in the power sector over the typical 100-year assessment period. However, climate benefits from vehicle fuel substitution are uncertain (gasoline, light-duty) or improbable (diesel, heavy-duty) (*28*). These conclusions may undercount benefits of NG, as both EPA GHGI methods and many regionally focused top-down studies attribute CH_4 emissions from coproducing NG systems to the NG sector, rather than to a mixture of oil and NG sources.

How can management and policy help address the leakage problem? Opportunities abound: Many solutions are economically profitable at moderate NG prices, with some technologies already being adopted or to be required in regulation (*23*, *26*) (e.g., reduced emissions completions). Facility studies using existing technology have found leakage detection and repair programs to be profitable (*21*).

The heavy-tailed distribution of observed emissions rates presents an opportunity for large mitigation benefits if scientists and engineers can develop reliable (possibly remote) methods to rapidly identify and fix the small fraction of high-emitting sources.

However, this heterogeneity also creates challenges in formulating statistical distributions for use in inventories. Approaches that assume "typical" emissions rates for this industry are inherently challenged. Inventories can be improved through efforts to better characterize distributions and by incorporating flexibility to adapt to new knowledge.

Improved science would aid in generating cost-effective policy responses. Given the cost of direct measurements, emissions inventories will remain useful for tracking trends, highlighting sources with large potential for reductions, and making policy decisions. However, improved inventory validation is crucial to ensure that supplied information is timely and accurate. Device-level measurements can be performed at facilities of a variety of designs, vintages, and management practices to find low-cost mitigation options. These studies must be paired with additional atmospheric science to close the gap between top-down and bottom-up studies. One such large study is under way (*29*), but more work is required.

If natural gas is to be a "bridge" to a more sustainable energy future, it is a bridge that must be traversed carefully: Diligence will be required to ensure that leakage rates are low enough to achieve sustainability goals.

References and Notes

1. A. Karion *et al.*, *Geophys. Res. Lett.* **40**, 4393 (2013).
2. J. Peischl *et al.*, *J. Geophys. Res.* **118**, 4974 (2013).
3. G. Pétron *et al.*, *J. Geophys. Res.* **117**, (D4), D04304 (2012).
4. E. G. Nisbet *et al.*, *Science* **343**, 493 (2014).
5. E. A. Kort *et al.*, *Geophys. Res. Lett.* **35**, L18808 (2008).
6. S. M. Miller *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20018 (2013).
7. P. O. Wennberg *et al.*, *Environ. Sci. Technol.* **46**, 9282 (2012).
8. Y. Xiao *et al.*, *J. Geophys. Res.* **113**, (D21), D21306 (2008).
9. A. S. Katzenstein *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11975 (2003).
10. Y.-K. Hsu *et al.*, *Atmos. Environ.* **44**, 1 (2010).
11. A. Townsend-Small *et al.*, *J. Geophys. Res.* **117**, (D7), D07308 (2012).
12. D. Wunch *et al.*, *Geophys. Res. Lett.* **36**, 10.1029/2009 GL039825 (2009).
13. J. S. Wang *et al.*, *Global Biogeochem. Cycles* **18**, 10.129/2003GB0021680 (2004).
14. A. Chambers, Optical Measurement Technology for Fugitive Emissions from Upstream Oil and Gas Facilities (Alberta Research Council, Edmonton, AB, 2004).
15. Clearstone Engineering, *Identification and Evaluation of Opportunities to Reduce Methane Losses at Four Gas Processing Plants* (Gas Technology Institute, Des Plaines, IL, 2002).
16. Innovative Environmental Solutions, *Field Measurement Program to Improve Uncertainties for Key Greenhouse Gas Emissions Factors for Distribution Sources* (Gas Technology Institute, Des Plaines, IL, 2009).
17. EPA, "Inventory of U.S. greenhouse gas emissions and sinks: 19902011" (EPA, 2013).
18. M. R. Harrison *et al.*, *Natural Gas Industry Methane Emissions Factor Improvement Study* (EPA, 2011).
19. K. E. Hummel *et al.*, *Methane Emissions from the Natural Gas Industry*, vol. 8, *Equipment Leaks* (EPA, 1996).
20. H. J. Williamson *et al.*, *Methane Emissions from the Natural Gas Industry*, vol. 4, *Statistical Methodology* (EPA, 1996).
21. National Gas Machinery Laboratory, Clearstone Engineering, Innovative Environmental Solutions, *Cost-Effective Directed Inspection and Maintenance Control Opportunities at Five Gas Processing Plants and Upstream Gathering Compressor Stations and Well Sites* (EPA, 2006).
22. Office of Inspector General, EPA, *EPA Needs to Improve Air Emissions Data for the Oil and Natural Gas Production Sector* (EPA, 2013).
23. EPA, *Fed. Regist.* 77(159), 49490 (16 August 2012); 40 Code of Federal Regulations, Parts 60 and 63.
24. EPA, *Greenhouse Gas Reporting Program*, subpart W, *Petroleum and Natural Gas Systems* (EPA, 2013).
25. M. A. Levi, *J. Geophys. Res.* **117**, (D21), 16 (2012).
26. D. T. Allen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17768 (2013).
27. C. L. Weber, C. Clavin, *Environ. Sci. Technol.* **46**, 5688 (2012).
28. R. A. Alvarez *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6435 (2012).
29. Environmental Defense Fund, Natural gas: EDF is fighting for tough rules and enforcement (EDF, 2013); www.edf.org/climate/natural-gas.

Acknowledgments: This study was funded by Novim through a grant from the Cynthia and George Mitchell Foundation. G. Mitchell pioneered hydraulic fracturing and believed that this technology should be pursued in ecologically sound ways. We acknowledge the support of the Joint Institute for Strategic Energy Analysis. C. Brown and D. Heppe of the Environmental Capital Group provided assistance. See the SM for full listing of authors' recent or ongoing consulting, honoraria, and other financial and management or advisory disclosures. We thank D. McCabe for comments.

Supplementary Materials

www.sciencemag.org/content/343/6172/733/suppl/DC1

10.1126/science.1247045

Self-Organizing Somites

Shigeru Kondo

Science **343**, 736 (2014);
DOI: 10.1126/science.1250245

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

DEVELOPMENTAL BIOLOGY

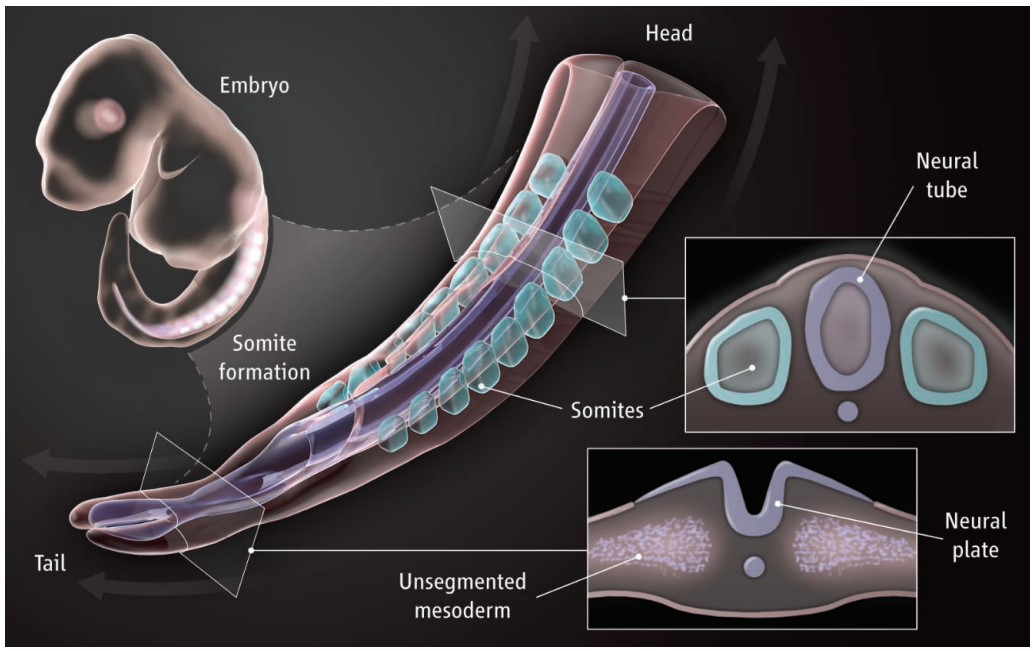
Self-Organizing Somites

The size and position of tissue during vertebrate body plan development may rely on the cooperation of mechanisms that act globally and locally.

Shigeru Kondo

During vertebrate embryogenesis, the body plan is established through the organization of blocks of tissue (somites) that form along either side of the neural tube, the precursor to the adult spinal cord (see the figure). How a regularly arranged pattern of somites arises from the patternless presomitic mesoderm has not been clear, but the "clock and wavefront" model of oscillating gene expression (the clock) and a traveling wave of signals to stop the oscillation, put forth in 1976 (*1*), has been the most widely accepted theory. On page 791 of this issue, Dias *et al*. (*2*) try to overturn this model by showing that somites can form without either oscillations or the wave.

A somite is a spherical ball with a lumen surrounding a variable amount of mesodermal cells (depending on the species); the mesodermal cells lie beneath a surface layer of epithelial cells (also of a variable thickness). According to the clock and wavefront model and the experiments supporting it (*1*, *3*–*5*), what determines the size of each somite is the combination of local oscillation in the expression of a network of genes in presomitic mesodermal cells and their exposure to a wave of extracellular signal(s) that travels from anterior to posterior in the presomitic mesoderm. The oscillation persists over the period of time required for a somite to form, which varies across species (for example, it is 30 min for zebrafish but 90 min for the chick). It stops when a cell meets a wavefront of a specific signal [such as the reduction of fibroblast growth factor (FGF)] that triggers mesodermal cells to transition and become epithelial cells. The somite boundary is thus defined by those cells in which the oscillation has stopped at a specific phase. As the wavefront continuously travels from anterior to posterior, somite boundaries are made sequentially with even spacing. The somite size is determined by the ratio of the oscillation period to the speed of the traveling wave.



Somitogenesis. (**A**) In the vertebrate embryo, blocks of tissue called somites form along the body axis. (**B**) Somites arise from presomitic mesoderm. Local cell-autonomous mechanisms may determine the size of a somite, whereas a global mechanism such as the "clock and wavefront" may determine the absolute position of each somite along the body axis. The clock and wavefront model proposes that oscillations of gene expression and waves of extracellular signal control the regularly arranged array of somites.

Dias *et al*. observed something different. The authors found that presomitic mesodermal cells removed from an early-stage chick (or quail) embryo, incubated with the molecule Noggin (for 3 hours) and then implanted into the extraembryonic region, could simultaneously form up to 14 somite-like structures that collectively resemble a "bunch of grapes." Noggin restricts the differentiation of mesodermal cells into somitic cells in the developing chick. Each somite-like structure consisted of epithelial cells surrounding a lumen and was of a size consistent with somites found in the chick embryo. They only lacked a rostral-caudal subdivision. When the somite-like structures were implanted in the position of normal somites in a host chick embryo, they developed the tissues that are normally derived from somite cells. Oscillation of segmentation gene expression was not observed; because the structures were implanted outside of the presomitic mesoderm, they were not exposed to a wavefront of signal that stops the oscillatory clock.

This finding of Dias *et al*. indicates that presomitic mesodermal cells can self-organize the somite structure without the indication of the segment position by the clock and wavefront mechanism. Rather, local interactions between cells appear to control the morphology of a somite, including its correct size.

It remains to be examined whether the presomitic mesoderm of mice or zebrafish has the same property. But the experiment of Dias *et al*. is quite simple and the result is dramatic, suggesting that the phenomenon could be universal among vertebrates. In the mouse and zebrafish (*6*, *7*), the anterior-to-posterior length of the somites decreases when the period of segment gene oscillation is shortened. This points to direct involvement of the oscillation period in somite size determination. However, the finding of Dias *et al*. does not necessarily contradict cooperation between global (clock and wavefront) and local (cell-cell interaction) mechanisms to control somitogenesis. The size of a structure

Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka, Suita, Osaka 565-0871, Japan. E-mail: shigerukondo@gmail.com

CREDIT: C. BICKEL/*SCIENCE*

Published by AAAS

that is made by a cell-autonomous mechanism is usually influenced considerably by local environmental perturbations. Because the size change of somites observed in the mouse and zebrafish studies was not so drastic and was within the flexibility of a local autonomous mechanism, the existence of a local autonomous mechanism cannot be ruled out. If the size change was more extensive (greater than 200% or less than 50%), then the possibility of a local mechanism would be quite low.

The cooperation of the two different mechanisms offers many advantages. A local cell-autonomous mechanism can determine the size of a somite but cannot determine the absolute position of each somite. To form the regularly arranged array of somites, some global mechanism like the clock and wavefront mechanism is surely required. The clock and wavefront model does have some weak points, too. For example, it does not work for the most cranial four or five somites because they arise simultaneously (*8*). Another problem concerns the precision of the positional information given by the wavefront. Temporally, the concentration gradient of FGF is thought to act as the wavefront (in the chick, mouse, and zebrafish) (*5*). However, the slope of the gradient seems too gentle to indicate the precise timing to the oscillating cells. The model also cannot specify somite size along either the dorsoventral or the lateral axis. By incorporating a local autonomous mechanism to determine the somite size, these weaknesses are removed.

Although the cooperation of a global and local mechanism is possible, it leaves the most important question unresolved: What determines the size of somites? Dias *et al*. present a mathematical model mainly based on packing constraints of cells transitioning between a mesenchymal state and a polarized epithelium. But other processes transferring the long-range signal, such as diffusion, cell projection, and mechanical stress, could be mechanisms that determine the regular size.

References

1. J. Cooke, E. C. Zeeman, *J. Theor. Biol.* **58**, 455 (1976).
2. A. S. Dias, I. de Almeida, J. M. Belmonte, J. A. Glazier, C. D. Stern, *Science* **343**, 791 (2014); 10.1126/science.1247575.
3. I. Palmeirim, D. Henrique, D. Ish-Horowicz, O. Pourquié, *Cell* **91**, 639 (1997).
4. J. Dubrulle, M. J. McGrew, O. Pourquié, *Cell* **106**, 219 (2001).
5. O. Pourquié, *Cell* **145**, 650 (2011).
6. Y. Harima, Y. Takashima, Y. Ueda, T. Ohtsuka, R. Kageyama, *Cell Rep.* **3**, 1 (2013).
7. C. Schröter, A. C. Oates, *Curr. Biol.* **20**, 1254 (2010).
8. T. M. Lim, E. R. Lunn, R. J. Keynes, C. D. Stern, *Development* **100**, 525 (1987).

10.1126/science.1250245

CLIMATE CHANGE

A Drier Future?

Global warming is likely to lead to overall drying of land surfaces.

Steven Sherwood[1] and Qiang Fu[2,3]

Global temperature increases affect the water cycle over land, but the nature of these changes remains difficult to predict. A key conceptual problem is to distinguish between droughts, which are transient regional extreme phenomena typically defined as departures from a local climatological norm that is presumed known, and the normal or background dryness itself. This background dryness depends on precipitation, but also on how fast water would evaporate. As the planet warms, global average rainfall increases, but so does evaporation. What is the likely net impact on average aridity?

Most studies of dryness focus on droughts rather than on the background aridity or changes thereto. They tend to rely on relatively simple measures that are useful for analyzing temporary anomalies but may not properly account for factors that govern the background state. Failure to explicitly account for changes in available energy, air humidity, and wind speed can cause some indices commonly used for identifying droughts to diagnose an artificial trend toward more drought in a warming climate (*1*). Recognition of this problem has undone past claims that drought is on the rise globally, and led to weaker claims about observed drought trends in the most recent Intergovernmental Panel on Climate Change report (*2*). However, that does not mean that conditions will not get drier (*3*, *4*).

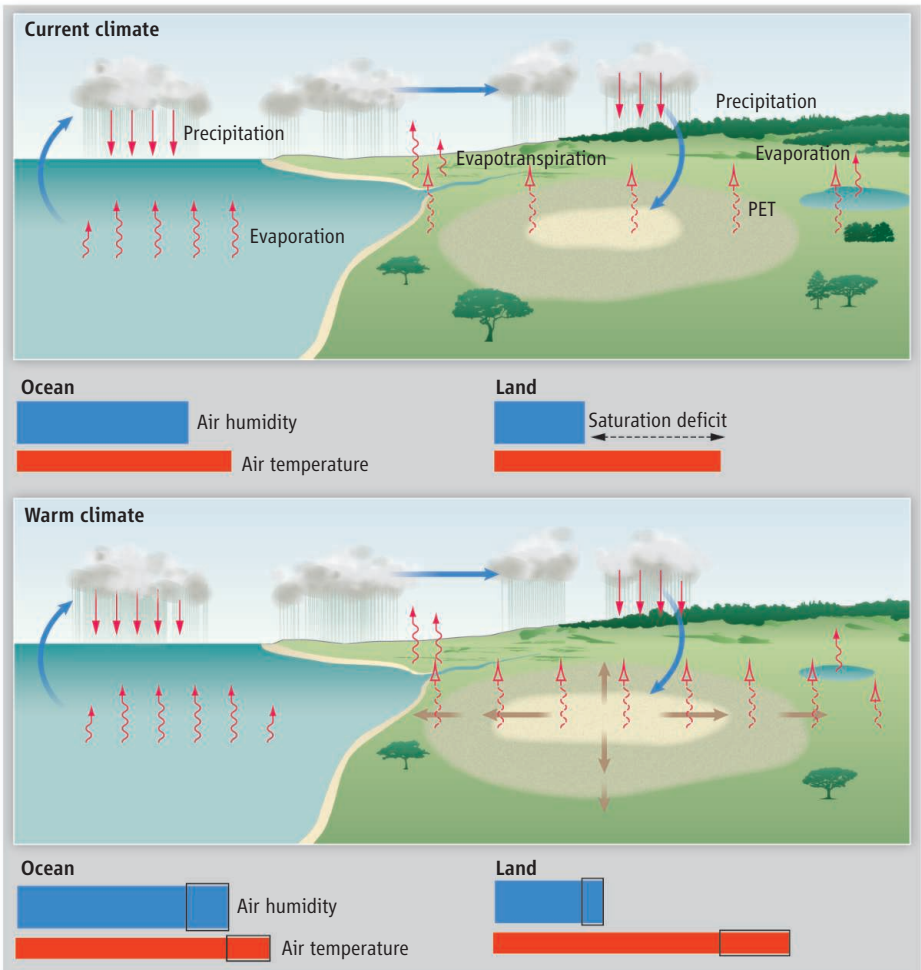
Challenges in CLIMATE SCIENCE
scim.ag/climatechall

A different way of approaching the problem is to try to capture the changes in background state, rather than temporary anomalies such as droughts. This can, for example, be done using the ratio of precipitation (P) to potential evapotranspiration (PET) based on the Penman-Monteith equation (*1*, *5*). PET is the evaporative demand of the atmosphere, calculated as the amount of evaporation one would get, with given air properties, from a completely wet surface. Over a body of water PET equals the true evaporation, but on land, the true evaporation will be less than PET unless the soil is saturated with water. The P/PET ratio may be near zero in a desert but can exceed unity in wet climates. If the P/PET ratio falls, it means that conditions get drier; if it rises, conditions are getting wetter.

Recent observational studies have shown that P/PET is decreasing on average as the globe warms (*5*, *6*). Climate model simulations (see fig. S1, panel A) (*5*) predict that by 2100 under a high-emissions scenario, when climate is projected to be several degrees warmer than it is now, P/PET will drop much further in most tropical and mid-latitude land regions (see the figure). Such drops can shift a region to the next drier climate category among humid, subhumid, semiarid, arid, and hyperarid conditions (the latter four together are denoted dryland). In one simulation, the area of global dryland is projected to expand by ~10% by 2100 (*5*). Models predict that India and northern tropical Africa will become wetter, but nearly all other land regions are predicted to become drier. Under most scenarios, the drying would further intensify during the 22nd century.

Global averages of precipitation and evaporation must remain equal to each other on climate time scales. The observed and predicted drying tendency in P/PET over land thus implies that PET there increases faster than does global evaporation (noting that precipitation changes similarly on land and oceans). If there were no land on Earth, PET globally could not increase faster than P; they would always be equal. Thus, the increase in P/PET must be peculiar to land surfaces. One might expect complex land-surface

[1]Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, 2052 Australia. [2]College of Atmospheric Sciences, Lanzhou University, Lanzhou, China. [3]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA. E-mail: ssherwood@alum.mit.edu



Warmer, drier. Aridity increases in warmer climates, leading to expansion of dry climate zones. Evaporation and precipitation increase modestly, but on land, evaporative demand (broken wavy arrows) increases faster than precipitation, because the strong increases in air temperature and consequently saturated water vapor concentration over land (red bars at lower right) exceed growth in actual water vapor concentration (blue bars). Increases in sensible and latent heat (associated, respectively, with temperature and water vapor, and represented by the area of each bar) have the same sum over land and ocean, with sensible heat increasing more over land than oceans and latent heat increasing more over oceans. Relative humidity (ratio of blue to red bar length) decreases over land.

responses involving soils or vegetation to be responsible, but recent research (*7*–*12*) suggests that the overall drying trend on land is rooted in relatively simple atmospheric thermodynamics.

The key factor causing drying is that land surfaces (and the air just above them) warm, on average, about 50% more than ocean surfaces (*7*). There is a simple and plausible explanation for this long-remarked phenomenon, at least for low and mid-latitudes. The atmosphere keeps convective instability (which gives rise to cumulus clouds) small over both land and ocean regions. This instability depends on the total latent and sensible heat in air near the surface. Because the latent heat (determined by atmospheric water vapor concentration) is smaller over land and changes less upon warming (see the figure), sensible heat (determined by air temperature) must change more, explaining the enhanced land warming (*7*, *8*). Indeed, if this enhanced warming did not happen, air over land would become less able to sustain clouds and precipitation, thus drying and warming the land via increased sunshine. Enhanced warming of land surfaces relative to oceans thus occurs simply because continental air masses are drier than maritime ones, which in turn is a consequence of the limited availability of surface water.

The second factor ensuring drying is that water vapor content over land does not increase fast enough relative to the rapid warming there. Nearly all water vapor in the atmosphere comes from the oceans, where the water vapor content of the overlying air increases by ~6% per degree Celsius of ocean surface temperature (*9*, *10*). When this air moves onto land, its typical water vapor content (though reduced) reflects the amount that it held originally (*11*). Because the land warms faster than the oceans, however, the humidity of the arriving air does not increase enough to maintain a constant relative humidity. The latter must therefore fall on average (see the figure), as indeed seen in model simulations (*11*, *12*) and observed on all continents (*10*). Therefore, the saturation deficit (gap between actual and saturation water vapor concentration; see the figure), which is the key factor controlling PET, grows much faster in percentage terms than do other hydrological quantities. This increases the aridity.

A map of the predicted change in annual mean near-surface relative humidity (see fig. S1, panel B) (*13*) not only confirms a general decrease over most land regions, but also shows a pattern nearly identical to that of the change in P/PET. These similarities show that regional changes in near-surface humidity, soil moisture, and precipitation are tightly coupled. Increases in PET are mainly attributable to overall land warming rather than relative humidity change (*14*), but the P/PET ratio on land is reduced largely by the enhanced land warming relative to oceans (see the figure) and by the decreases in relative humidity on land. The latter are negative over most land areas despite being slightly positive over oceans. Positive feedback from soil moisture changes is not needed to explain enhanced land warming, but likely amplifies it in some regions (*15*).

Regional variations in simulated aridity change may still be unreliable, or may reflect other changes such as poleward shifts of climate zones (*5*). But the general trend toward a drier land surface appears to rest on relatively firm foundations. The predicted drying would be sufficient to shift large portions of the Earth to new, drier climate categories (although the richer atmospheric CO_2 might mitigate the impact on some plants). The background drying is separate from, but may be compounded by, the expected trend toward more intermittent rainfall for a given mean rain rate (*16*).

As the above considerations show, focusing on changes in precipitation, as typical in high-profile climate reports (*2*), does not tell the whole story—or perhaps even the main story—of hydrological change. In particular, it obscures the fact that in a warmer climate, more rain is needed. Many regions will get more rain, but it appears that few will get enough to keep pace with the growing evaporative demand.

CREDIT: ADAPTED BY P. HUEY/*SCIENCE*

Published by AAAS

References and Notes

1. J. Sheffield *et al.*, *Nature* **491**, 435 (2012).
2. IPCC, Summary for Policymakers, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, Cambridge and New York, 2013).
3. C. Prudhomme *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 10.1073/pnas.1222473110 (2014).
4. J. Schewe *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 10.1073/pnas.1222460110 (2014).
5. S. Feng, Q. Fu, *Atmos. Chem. Phys.* **13**, 10081 (2013).
6. A. Dai, *Nat. Clim. Change* **3**, 52 (2013).
7. M. M. Joshi *et al.*, *Clim. Dyn.* **30**, 455 (2008).
8. M. P. Byrne, P. A. O'Gorman, *J. Clim.* **26**, 4000 (2013).
9. A. Dai, *J. Clim.* **19**, 3589 (2006).
10. A. J. Simmons, K. M. Willett, P. D. Jones, P. W. Thorne, D. P. Dee, *J. Geophys. Res.* **115**, D01110 (2010).
11. D. P. Rowell, R. G. Jones, *Clim. Dyn.* **27**, 281 (2006).
12. P. A. O'Gorman, C. J. Muller, *Environ. Res. Lett.* **5**, 025207 (2010).
13. M. Collins *et al.*, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, Cambridge and New York, 2013), chap. 12.
14. J. Scheff, D. M. W. Frierson, *J. Clim.* 10.1175/JCLI-D-13-00233.1 (2013).
15. S. I. Seneviratne *et al.*, *Geophys. Res. Lett.* **40**, 5212 (2013).
16. K. E. Trenberth, *Clim. Res.* **47**, 123 (2011).

Acknowledgments: This Perspective germinated at the Bert Bolin Centre 2013 Summer School on Subtropical Climate. We acknowledge the support of the National Basic Research Program of China (2012CB955303), the National Natural Science Foundation of China (grant 41275070), and the Australian Research Council Centre of Excellence for Climate System Science.

Supplementary Materials

www.sciencemag.org/content/343/6172/737/suppl/DC1
Fig. S1
References

10.1126/science.1247620

CHEMISTRY

Capturing Surface Processes

A modified surface x-ray diffraction geometry allows dynamic restructuring of surfaces to be studied.

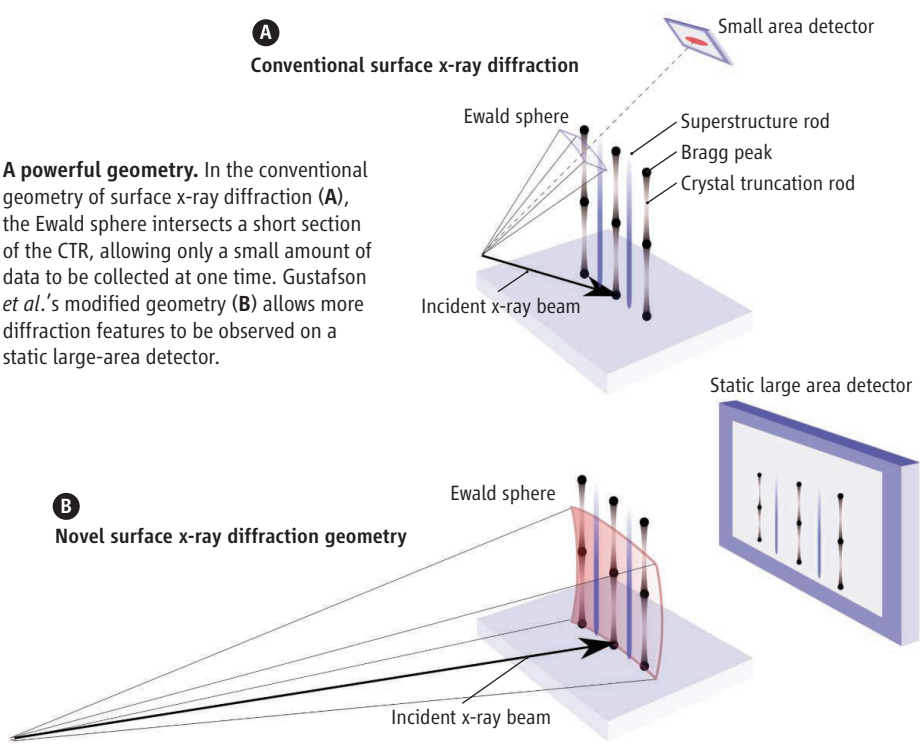
Chris Nicklin

The outer atomic layers of a solid or liquid play a central role in determining the properties of the sample as a whole, because it is here where the material interacts with the external environment. Detailed knowledge of the arrangement of atoms at a surface or interface between two materials is required to understand and tune the material's properties. This outer-layer structure is crucial for technological processes such as catalysis, lubrication, and electron transport. In surface x-ray diffraction, surface structures are investigated by directing high-energy x-rays at a sample at grazing angles of typically less than 1° (*1*). On page 758 of this issue, Gustafson *et al.* outline a different geometry for these measurements, using even higher-energy x-rays and shallower angles to allow faster data collection, enabling dynamic surface restructuring processes to be captured (*2*).

In surface x-ray diffraction, the diffracted intensity results from a combination of x-rays scattered from the bulk of the sample and x-rays scattered from its surface (see the figure). Intense Bragg peaks occur where the bulk scattering exhibits constructive interference. The truncation of the sample at the surface leads to streaking between the Bragg peaks in the direction perpendicular to the surface. These streaks, known as crystal truncation rods (CTRs) (*3*) show modulations in intensity that results from interference between the bulk-scattered and surface scattered x-rays. Additionally, ordered reconstructions of the outer atomic layers result in superstructure rods, which have an intensity profile that depends only on the surface scattering. Modeling these modulations can reveal the surface structure and registry with the bulk with a resolution of <0.05 Å. The ordered array of diffraction features (CTRs, superstructure rods, and Bragg peaks) formed by a single-crystal sample is known as the reciprocal space lattice.

Traditionally, surface x-ray diffraction measurements have required a high resolution diffractometer, which allows the sample and detector to be accurately positioned at specific angles relative to each other. These instruments usually have five or six independent rotation axes that enable a particular diffraction feature to be detected while maintaining the fixed angle of incidence (*4*). In most state of the art experiments, the scattered x-rays are collected by a small two-dimensional detector (*5*), where the intensity of the spot or streak on the detector results in a single point on the rod.

One way to understand the x-ray scattering process is through the Ewald construc-



A powerful geometry. In the conventional geometry of surface x-ray diffraction (**A**), the Ewald sphere intersects a short section of the CTR, allowing only a small amount of data to be collected at one time. Gustafson *et al.*'s modified geometry (**B**) allows more diffraction features to be observed on a static large-area detector.

Surface and Interface Diffraction, Beamline I07, Diamond Light Source, Didcot OX11 0DE, UK. E-mail: chris.nicklin@diamond.ac.uk

tion (*6*), which leads to a representation of all the possible points where diffraction could occur, known as the Ewald sphere. However, it is only in the small area where the Ewald sphere intersects the reciprocal space lattice, and with the detector placed appropriately, that a diffracted signal from the sample is actually measured (see the figure, panel A). Measuring the full CTR requires the coordinated movement of three or four of the diffractometer axes (detector and sample), which is relatively slow.

The geometry outlined by Gustafson *et al*. enables a much larger section of the CTR to be collected at once (see the figure, panel B). The higher x-ray energy increases the size of the Ewald sphere, meaning that it is flatter when it cuts through the CTR. This greater overlap explains why more of the rod is visible on the detector. A full CTR is recorded by rotating the sample perpendicular to its surface through a small angular range, which moves the intersection point of the Ewald sphere and the rod. The higher energies also lead to the reciprocal space lattice being much more contracted in real space; reducing the angular separation between the scattered diffraction features. A suitably large detector can therefore collect many reflections simultaneously.

This geometry has a number of benefits, as shown in Gustafson *et al*.'s elegant study. A large data set can be acquired quickly, allowing dynamic restructuring of the surface to be monitored in detail during in situ processing. As an example, the authors determine the changes to the Pd surface during catalytic oxidation of carbon monoxide with subsecond time resolution. Additionally, the impressive volume of data collected can be visualized in a number of ways. The authors, for example, show an "in-plane" projection. Due to the inherent high resolution of surface x-ray diffraction, they can clearly observe the small shifts in peak positions that result from strain in the reconstructed sample.

The novel experimental geometry reported by Gustafson *et al*. makes it possible to extend surface x-ray diffraction to experiments that are currently very difficult, if not impossible. In contrast to many other surface structural techniques, surface x-ray diffraction is not restricted to a vacuum environment. Ambient pressure studies, for example (*7*), better resemble real conditions in automotive catalysts or in atmospheric reactions. The technique could be used to characterize transient structural phases that may occur in a specific pressure or humidity range. It could also provide deeper insights into dynamic processes such as the mechanism of facet formation during gas exposure, by monitoring not only the intensity, but also the angle and splitting of the CTRs in three dimensions. The ability to visualize the data in simple-to-interpret ways adds to the benefits of this powerful surface structural technique.

References

1. R. Feidenhans'l, *Surf. Sci. Rep.* **10**, 105 (1989).
2. J. Gustafson *et al*., *Science* **343**, 758 (2014); 10.1126/1246834.
3. I. K. Robinson, *Phys. Rev. B* **33**, 3830 (1986).
4. E. Vlieg, *J. Appl. Cryst.* **31**, 198 (1998).
5. C. M. Schlepütz *et al*., *Acta Crystallogr. A* **61**, 418 (2005).
6. P. P. Ewald, *Acta Crystallogr. A* **25**, 103 (1969).
7. V. H. Grassian, *Surf. Sci.* **602**, 2955 (2008).

10.1126/science.1250472

MATERIALS SCIENCE

Graphene Oxide Membranes for Ionic and Molecular Sieving

Membranes made by properly spacing and bonding stacked graphene oxide nanosheets enable precise, superfast sieving of ions and molecules.

Baoxia Mi

Ionic and molecular sieving membranes that enable fast solute separations from aqueous solutions are essential for processes such as water purification and desalination, sensing, and energy production (*1*–*3*). The two-dimensional structure and tunable physicochemical properties of graphene oxide (GO) offer an exciting opportunity to make a fundamentally new class of sieving membranes by stacking GO nanosheets (*4*–*6*). In the layered GO membrane, water molecules permeate through the interconnected nanochannels formed between GO nanosheets and follow a tortuous path primarily over the hydrophobic nonoxidized surface rather than the hydrophilic oxidized region of GO (*7*). The nearly frictionless surface of the nonoxidized GO facilitates the extremely fast flow of water molecules (*5*). On page 752 of this issue, Joshi *et al*. (*8*) further report that ions smaller in size than the GO nanochannel can permeate in the GO membrane at a speed orders of magnitude faster than would occur through simple diffusion. Size exclusion appears to be the dominant sieving mechanism.

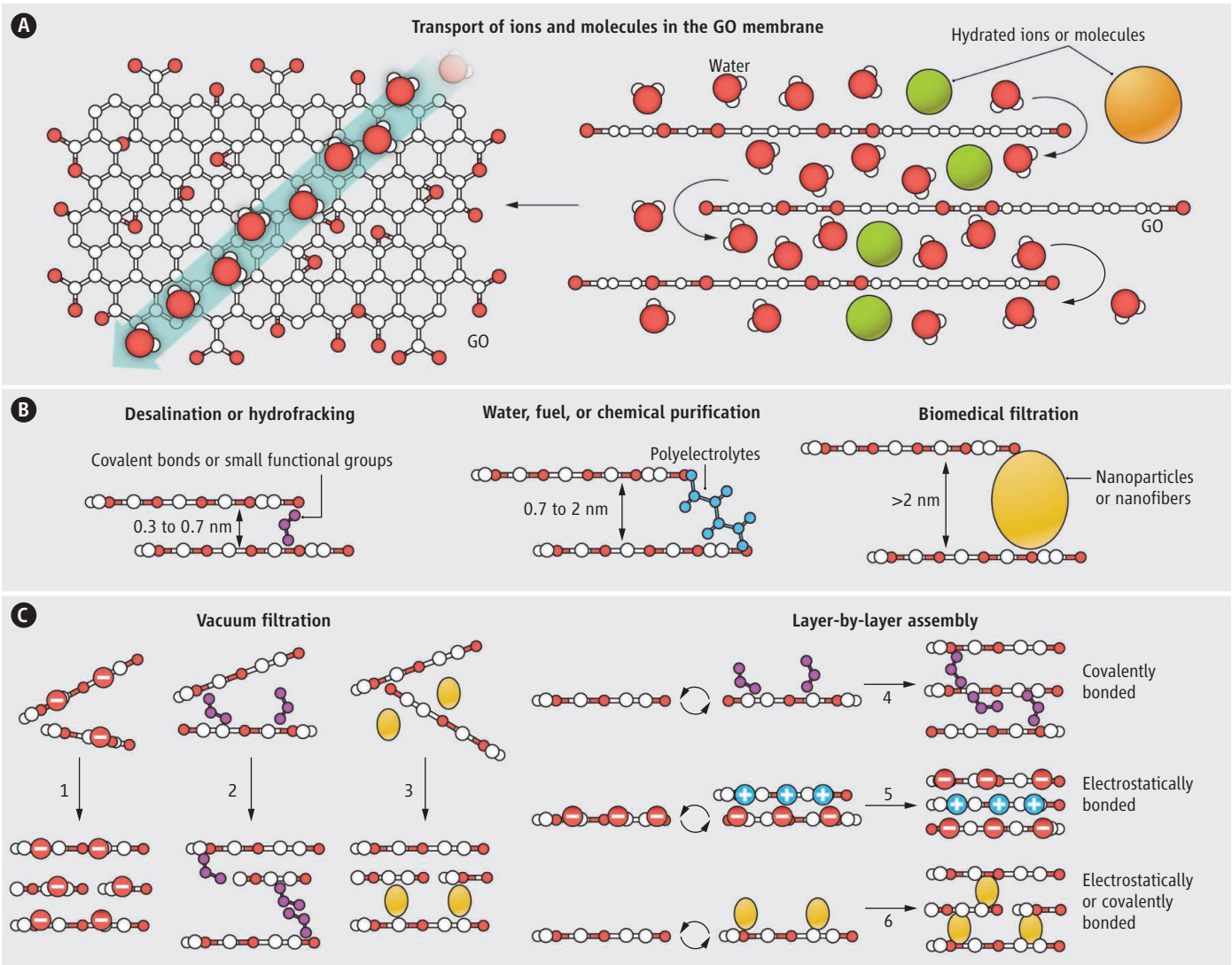
When dry, GO membranes made by vacuum filtration can be so tightly packed (with a void spacing of ~0.3 nm between GO nanosheets) that only water vapor aligned in a monolayer can permeate through the nanochannel (*5*). Joshi *et al*. found that when such a GO membrane was immersed in an ionic solution, hydration increased the GO spacing to ~0.9 nm (*8*). Any ion or molecule with a hydrated radius of 0.45 nm or less could enter the nanochannel, but all larger-sized species were blocked (see the figure).

Such a sharp size cutoff by the GO membrane has important implications in a myriad of separation applications. By adjusting the GO spacing through sandwiching appropriately sized spacers between GO nanosheets, a broad spectrum of GO membranes could be made, each being able to precisely separate target ions and molecules within a specific size range from bulk solution. Compared with the typically wide pore-size distribution of commonly used polymeric membranes, the narrow channel-size distribution of GO membranes is truly advantageous for precise sieving.

The hydration of GO in aqueous solution, however, makes it more challenging to manipulate the GO spacing within a subnanometer range than to enlarge it. For example, desalination requires that the GO spacing should be less than 0.7 nm to sieve the hydrated Na^+ (with a hydrated radius of 0.36 nm) from water. Such small spacing could be obtained by partially reducing GO to decrease the size of hydrated functional groups or by covalently bonding the stacked GO nanosheets with small-sized molecules to overcome the hydration force.

In contrast, an enlarged GO spacing (1 to 2 nm) can be conveniently achieved by

Department of Civil and Environmental Engineering, University of Maryland, 1161 Glenn L. Martin Hall, College Park, MD 20742, USA. E-mail: bmi@umd.edu



GO membranes. (**A**) Water and small-sized ions and molecules (compared with the void spacing between stacked GO nanosheets) permeate superfast in the GO membrane, but larger species are blocked. (**B**) The separation capability of the GO membrane is tunable by adjusting the nanochannel size. (**C**) Several methods for the synthesis of GO membranes have been reported or are envisioned; GO nanosheets can be physically packed by vacuum filtration (options 1 to 3), or they can be stabilized by covalent bonds, electrostatic forces, or both (options 4 to 6) during layer-by-layer assembly.

inserting large, rigid chemical groups (*6*) or soft polymer chains (e.g., polyelectrolytes) between GO nanosheets, resulting in GO membranes ideal for applications in water purification, wastewater reuse, and pharmaceutical and fuel separation. If even larger-sized nanoparticles or nanofibers are used as spacers, GO membranes with more than 2-nm spacing may be produced for possible use in biomedical applications (e.g., artificial kidneys and dialysis) that require precise separation of large biomolecules and small waste molecules.

GO membranes can be synthesized either by vacuum filtration or by layer-by-layer (LbL) assembly, both of which are conducted in aqueous solution without any organic solvent involved and, hence, are more environmentally friendly. The GO membranes prepared by vacuum filtration, either from a pure GO solution or a mixture of GO and spacers, might lack sufficient bonding between GO nanosheets. Because of the high hydrophilicity of GO, these membranes are likely to disperse in water, especially under cross-flow conditions typically encountered in membrane operations. In contrast, the LbL method is ideal for introducing an interlayer stabilizing force via covalent bonding (*6*), electrostatic interaction, or both effects during layer deposition.

The GO membrane thickness can be readily controlled by varying the number of LbL deposition cycles. Theoretically, as few as two stacked GO layers would be needed to create a sieving channel. In reality, however, deposition of additional GO layers is warranted to counteract the detrimental effects of possible defects and nonuniform deposition of GO nanosheets on the membrane's sieving capability. Finally, the LbL synthesis of GO membranes is highly scalable and cost-effective, unlike the challenging synthesis of monolayer graphene membranes, which requires the manufacturing of large-sized graphene sheets and the punching of nanopores with a narrow size distribution (*9*).

Indeed, the GO membrane represents a next generation of ultrathin, high-flux, and energy-efficient membranes for precise ionic and molecular sieving in aqueous solution, with applications in numerous important fields. Future research is needed to understand thoroughly the transport of water and solutes in the GO membrane, especially to fundamentally elucidate

Published by AAAS

other potential separation mechanisms (e.g., charge and adsorption effects) in addition to size exclusion. More research is also needed to address specific issues concerning various exciting yet challenging applications in desalination, hydrofracking water treatment, and energy production, as well as in biomedical and pharmaceutical fields. Other largely unexplored areas include making multifunctional GO membranes with exceptional antifouling, adsorptive, antimicrobial, and photocatalytic properties.

References and Notes

1. M. A. Shannon *et al*., *Nature* **452**, 301 (2008).
2. M. Elimelech, W. A. Phillip, *Science* **333**, 712 (2011).
3. D. L. Gin, R. D. Noble, *Science* **332**, 674 (2011).
4. L. Qiu *et al*., *Chem. Commun. (Camb.)* **47**, 5810 (2011).
5. R. R. Nair, H. A. Wu, P. N. Jayaram, I. V. Grigorieva, A. K. Geim, *Science* **335**, 442 (2012).
6. M. Hu, B. Mi, *Environ. Sci. Technol.* **47**, 3715 (2013).
7. D. W. Boukhvalov, M. I. Katsnelson, Y. W. Son, *Nano Lett.* **13**, 3930 (2013).
8. R. K. Joshi *et al*., *Science* **343**, 752 (2014).
9. E. N. Wang, R. Karnik, *Nat. Nanotechnol.* **7**, 552 (2012).

Acknowledgments: Supported by NSF Awards CBET 1154572 and 1158601.

10.1126/science.1250247

ENGINEERING

Robots Acting Locally and Building Globally

Robots that act like termites can construct complex structures, guided only by simple rules and sensing their local environment.

Judith Korb

Termites are among the most fascinating animal architects in nature; their mounds were first described in a scientific journal more than 200 years ago (*1*). How can such tiny insects, each less than 1 cm in size and equipped only with a simple brain, construct air-conditioned buildings up to 500 times their size? Termites' construction principles differ fundamentally from those of human architecture. Humans build houses according to a blueprint, and the construction process is centrally guided by this plan. In contrast, social insects such as termites build in a decentralized, self-organized manner. Each individual works rather independently and follows a set of simple rules; the interactions among the workers and the interaction of each worker with its environment ensure an organized process without a central blueprint (*2*–*4*). On page 754 of this issue, Werfel *et al*. (*5*) describe the use of such insect principles to guide simple robots in constructing user-defined structures for human purposes.

Central to the work of Werfel *et al*. is the principle of stigmergy (*6*): Social insects use local information at the building site to coordinate building activity. As this information changes during the building process, the behavior is adjusted accordingly. An example in termites is the proposed deposition of chemical volatiles with the building particles that guide individuals to local construction sites. Similarly, Werfel *et al*.'s autonomous constructing robots move along a grid system and deposit building bricks next to other bricks. The robots are simple, even more so than termite workers. The robots can only sense bricks and the other robots next to them. They can move backward or forward, turn in place, and climb one step up or down; they can pick up, carry, and deposit bricks.

The robots adjust their behavior according to what they perceive locally when they move along the grid system; the possibilities include "nothing," other robots, and bricks. The exact "traffic rules" depend upon the structure to be built, and these rules are derived by an offline compiler that transforms three-dimensional representations of a desired structure into two-dimensional

PHOTOS: (A) V. SALEWSKI AND J. KORB; (B) HARVARD SEAS

Coordinated construction. (**A**) This termite mound, 3 m high, is the air-conditioned home of a *Macrotermes bellicosus* colony. The mound is constructed by thousands of tiny workers that coordinate their building activity through local information at the construction site. (**B**) The robots developed by Werfel *et al*. use similar principles to construct complex structures.

Evolution and Ecology, University of Freiburg, D-79104 Freiburg, Germany. E-mail: judith.korb@biologie.uni-freiburg.de

projections and recursively determines the rules for the robots to follow. This system is extremely elegant, as it allows the autonomous construction of any predefined structures with simple robots. It is robust to failure because of its decentralization, and is very flexible in its adjustments. The approach by Werfel *et al*. differs from others in that it develops the lower-level rules from the final structure to be built. Commonly, the reverse approach is pursued in complex systems science, but determining the emergent higher-level result from lower-level rules has proved difficult. Hence, it is also still unclear how termites can construct their impressive castles from the simple behaviors that researchers observe (*3*, *4*).

In both nature's construction works and the structures created by the robots in the approach of Werfel *et al*., the properties of the final product are crucial. A termite mound's architecture can determine the success of a colony (*7*). Mounds that are better adapted to local environments will, as a rule, have more offspring; thus, improved building rules that are genetically encoded will spread over time through a population. What is different in nature is that it starts with "mutations" in the building rules that are then tested in the evolutionary process. Over the millennia, evolution tested different rules, and what we observe today are those that worked. They might not be perfect, and the algorithms of Werfel *et al*. might also show us whether termites could still "learn" from humans.

References

1. H. Smeathman, *Philos. Trans. R. Soc. London* **71**, 139 (1781).
2. E. Bonabeau, G. Theraulaz, J. L. Deneubourg, S. Aron, S. Camazine, *Trends Ecol. Evol.* **12**, 188 (1997).
3. S. Camazine *et al*., *Self-Organization in Biological Systems* (Princeton Univ. Press, Princeton, NJ, 2001).
4. J. Korb, in *Biology of Termites: A Modern Synthesis*, D. E. Bignell, Y. Roisin, N. Lo, Eds. (Springer, Dordrecht, Netherlands, 2011), pp. 349–376.
5. J. Werfel, K. Petersen, R. Nagpal, *Science* **343**, 754 (2014).
6. P. P. Grassé, *Insectes Soc.* **6**, 41 (1959).
7. J. Korb, K. E. Linsenmair, *Oecologia* **118**, 183 (1999).

10.1126/science.1250721

BIOCHEMISTRY

Protein Folding, Interrupted

An antifreeze protein with a core consisting mostly of ordered water molecules violates widely held ideas about protein stabilization.

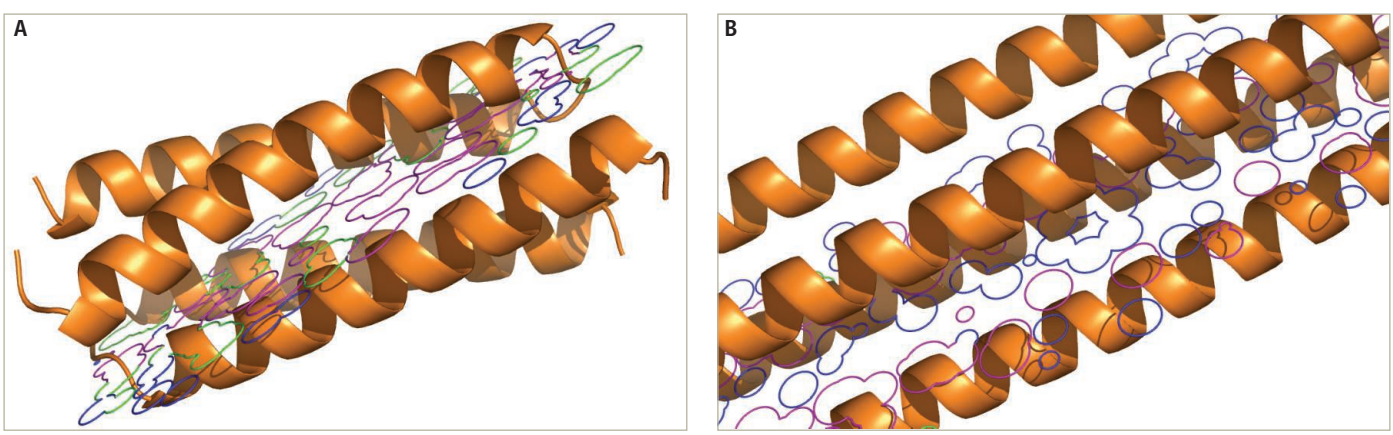
Kim A. Sharp

Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. E-mail: sharpk@mail.med.upenn.edu

Globular proteins start their lives as linear chains of amino acids coming off the ribosome. Proteins must then fold into specific three-dimensional structures to be functional. In 1957, the first such structure, of myoglobin, was determined at atomic resolution (*1*). Fifty-six years and 90,000-plus protein structures later (*2*), we have a very good idea of the necessary requirements for a stable, specific structure. Key to these requirements is the formation of a well-packed, largely anhydrous core (*3*). Yet, on page 795 of this issue, Sun *et al*. (*4*) report an antifreeze protein with a core mostly consisting of water.

In globular proteins, the anhydrous protein core provides both structural specificity and energetic stabilization (see the figure, panel A). Burial of apolar amino acid side chains inside the core relieves their unfavorable interaction with water, a process known as the hydrophobic effect (*5*, *6*). Even integral membrane proteins, which function in the nonaqueous lipid bilayer of the membrane and adopt structural motifs that are quite different from those of globular proteins, conform to this general principle. Here, the apolar side chains are on the outside of the structure, but by forming close contacts with the apolar lipid tails, they are still removed from water (*7*).

The remarkable structure of the antifreeze protein Maxi reported by Sun *et al*. flaunts its violation of the anhydrous-core principle. Maxi is a 145 Å–long four-helix bundle formed as a dimer of two-helix monomers. More than 400 highly organized water molecules form an integral part of Maxi's structure. The water is interleaved as a roughly two-molecule-thick layer between both intra- and intermonomer helix inter-



Core questions. In a typical globular protein, the four-helix bundle Rop [Protein Data Bank code: 4DO2 (*10*)] (**A**), water (blue) is excluded from the core because of efficient, interdigitated packing of apolar side chains (magenta), surrounded by polar side chains (green). By contrast, highly ordered water remains in the core in the antifreeze protein Maxi reported by Sun *et al*. (**B**). Secondary structure is rendered in orange. Slices through the van der Waals atomic surface taken through the core of the two proteins are shown in wire frame. Van der Waals surface slices produced by Pdb2DSlice and rendered in Pymol (*11*). In (B), the central third of the 145 Å–long Maxi protein is shown.

faces, effectively forming the core of the protein and extending to the ice-binding surface (see the figure, panel B).

Sun *et al*. show that the structure determined from crystallography persists in solution as the active form of the protein. Although ordered water molecules can be detected in most high-resolution x-ray crystallographic structures, they are usually located between replica molecules in the crystal lattice and are probably absent or quite rare when the protein is in native solution conditions, in contrast to the water inside Maxi. That the ordered water structure in Maxi extends to the ice-binding surface is suggestive of the function of this unusual core, given that Maxi, as an antifreeze protein, must bind ice nuclei and inhibit their growth to function. As Sun *et al*. suggest, this function may have driven the evolution of its unique water core, although clearly such a core is not necessary for antifreeze activity. No other known antifreeze protein has a water core like Maxi's.

When a protein folds, it forms van der Waals (packing) interactions, hydrogen bonds, and electrostatic interactions between charged and polar side chains within the protein. Each of these interactions competes with interactions of comparable strength and number between water and protein in the unfolded state. No such competition exists for the hydrophobic effect: Entropically favorable release of water upon burying apolar groups unambiguously favors the folded state. Thus, it is widely accepted that stabilization of globular proteins occurs primarily through the hydrophobic effect (*8*). It is therefore startling that Maxi retains the very structure of water—"semi-clathrate" in the words of Sun *et al*.—whose formation around apolar groups and subsequent disappearance was deemed to be the hallmark of the hydrophobic effect (*5*) and pivotal for protein stabilization. Clearly, the balance of interactions that stabilize Maxi is quite different from that used by most proteins.

Maxi's structure has intriguing implications not only for the energetics of protein folding, but for the mechanism and kinetics as well. Examination of the anhydrous core of any protein with its convoluted but well-packed atom-atom interfaces (see the figure, panel A) raises the question of how the water is removed so efficiently during folding. Removal of this water has been proposed as a potential rate-limiting step in protein folding. Two competing mechanisms have been proposed: Either the water is driven out as the protein collapses, or the unfavorable hydration free energy of apolar groups leads to their spontaneous dewetting (*9*). In the first mechanism, protein packing drives dehydration in the manner of a squeegee, whereas in the second, packing follows dehydration.

Resolution of this question depends on the relative magnitudes of packing and hydration forces, which have proved difficult to determine by experiment or theory. Moreover, proteins may well use both mechanisms. But it seems Maxi did not get the memo on how to fold: It chooses neither route to dehydration. Maxi folds to the point where water not in direct contact with the protein chain is removed from its core. It then arrests further folding to retain a beautifully ordered core of water interleaved between the protein helices. Further analysis of the energetics and kinetics of folding of Maxi will deepen our understanding of protein folding and stabilization.

References

1. J. C. Kendrew *et al*., *Nature* **181**, 662 (1958).
2. H. M. Berman *et al*., *Nucleic Acids Res.* **28**, 235 (2000).
3. F. M. Richards, *J. Mol. Biol.* **82**, 1 (1974).
4. T. Sun, F.-H. Lin, R. L. Campbell, J. S. Allingham, P. L. Davies, *Science* **343**, 795 (2014).
5. W. Kauzmann, *Adv. Protein Chem.* **14**, 1 (1959).
6. C. H. Tanford, *The Hydrophobic Effect* (Wiley, New York, 1973).
7. R. Henderson *et al*., *J. Mol. Biol.* **213**, 899 (1990).
8. K. A. Dill, *Biochem.* **29**, 7133 (1990).
9. Y. Levy, J. N. Onuchic, *Annu. Rev. Biophys. Biomol. Struct.* **35**, 389 (2006).
10. M. Ambrazi *et al*., *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **64**, 432 (2008).
11. W. L. DeLano, *PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA, 2002).

10.1126/science.1249405

NEUROSCIENCE

Genetic Resolutions of Brain Convolutions

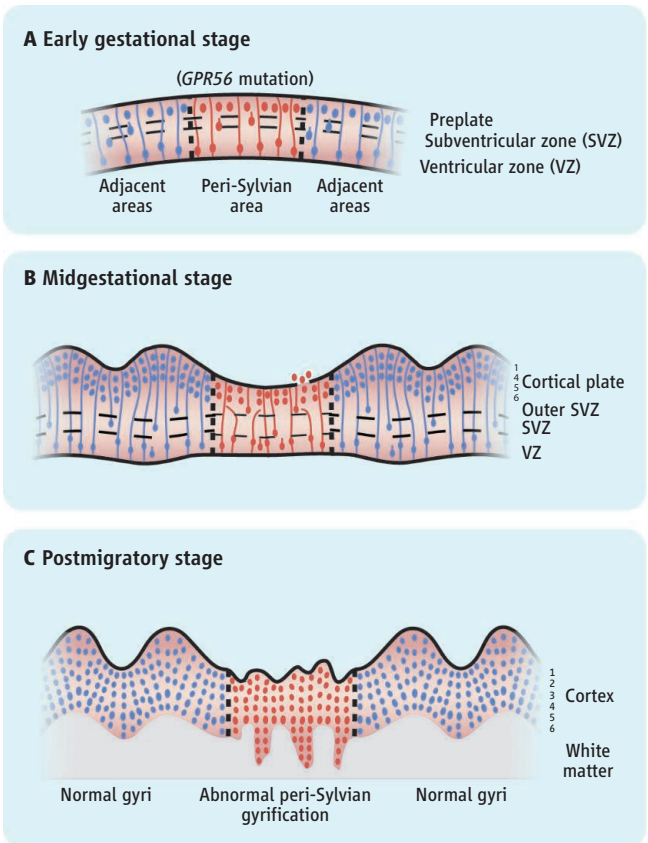
Genetic analysis of human brain abnormalities aids our understanding of how the cerebral cortex develops and evolves.

Brian G. Rash[1] and Pasko Rakic[1,2]

Cortical convolutions—prominent folds on the surface of the human brain—have a long history of speculation (*1*). The claims range from their function as a bodily cooling system to the attribution of Einstein's genius to the unusual shape of a single gyrus (the ridge of a cortical fold). Only recently, with advances in molecular genetics and brain imaging techniques, has it become possible to study the development, evolution, and abnormalities of cerebral convolutions in a scientifically rigorous manner (*2*). On page 764 of this issue, Bae *et al*. (*3*) show that a specific gene controls the number of gyri that form in a region of the cerebral cortex that includes Broca's area (the major language area). This begins to pinpoint mechanisms that underlie the development of specialized regions of the human brain and may be relevant to understanding human brain evolution.

Bae *et al*. examined individuals, from three consanguineous families, with abnormal cortical folding restricted to a region surrounding the Sylvian fissure, including Broca's area within the frontal lobe. Through a genome-wide linkage analysis, the authors traced the abnormality to mutations in the noncoding regulatory region of the *GPR56* gene. *GPR56* encodes a protein that functions in cell adhesion and guidance. Mutations caused the peri-Sylvian cortex to be thinner and smoother and exhibit multiple shallow indentations (polymicrogyria). Moreover, the authors discovered a natural, spontaneous mutation in the *GPR56* locus, which points to a mechanism that underlies both the formation of cortical maps and the process of gyrification. Bae *et al*. generated transgenic mice in which different expression patterns of a reporter gene (encoding β-galactosidase) could be driven by part of the noncoding region of *GPR56* (a minimal promoter) taken from human, marmoset, dolphin, cat, and mouse. This indicates evolutionary changes in cortical

[1]Department of Neurobiology, Yale University, New Haven, CT 06510, USA. [2]Kavli Institute for Neuroscience, Yale University, New Haven, CT 06510, USA. E-mail pasko.rakic@yale.edu



Localized gyral abnormalities. (**A**) Early embryonic stage in an individual with the *GPR56* mutation showing the prospective areas surrounding the Sylvian fissure (red) and adjacent cortex (blue) within the indicated zones. (**B**) Middle stage of corticogenesis indicating the prospective normal and affected peri-Sylvian cortical areas. (**C**) Postmigratory stage, showing abnormal gyrification and cytoarchitecture in the peri-Sylvian region flanked by normal cortical areas.

GPR56 expression. Overexpression of *GPR56* in mice increased proliferation of neuroprogenitor cells in the cortex, whereas loss of *GPR56* expression had the opposite effect.

The results of Bae *et al*. are predictable by the radial unit hypothesis and related protomap hypothesis, which concern the formation of the cortical areas and convolutions. These concepts have framed our current understanding of the development and evolution of cortical areas and their convolutions in the context of the genetic regulation of proliferation and migration of newborn neurons into columns within the overlying cortical plate (*4*). The models explain why the cortex normally expands not as a lump or globular nucleus (such as the striatum or thalamus), but as a relatively uniform thin, flat sheet composed of an array of radial units. Because the ventricular zone (VZ) of the cerebral cortex produces deep layers, and the subventricular zone (SVZ) produces neurons destined mostly for the superficial layers, both zones must be equally engaged in this complex process (*5*) (see the figure). The horizontal constraints provided by the radial glial cell scaffolding, which is so prominent in the gyrencephalic brains of human and nonhuman primates, explains why the larger cortex became convoluted and forms primary, and initiates secondary, furrows (sulci) on the brain's surface even before birth (*2*, *4*).

Animal studies have shown that the size and position of various cortical areas can be manipulated in a region-specific manner by molecules that control morphogenesis (*6*–*10*), indicating that the pattern of the primordial cortical protomap is malleable at early embryonic stages. Mechanistically, for example, a local increase in the proliferative rate of a prepatterned regional population of radial glial cells (cortical neural stem cells) at early stages will add radial units and enlarge the surface area for that population compared to the rest of the cortical primordium. By contrast, a local decrease in proliferation at early stages would diminish the cortical surface, and if continued at later stages, would also decrease cortical thickness (see the figure). The results of Bae *et al*., in which both VZ and SVZ are involved, are predictable by these models and support the idea that cortical size and its foldings are an extended property of local cell proliferation in the transient embryonic zones. The grossly distorted and disorganized cortical columns observed by Bae *et al*. could be attributed to the abnormal shape of radial glial cells, some of which lose their attachments to the pial surface (the thin membrane that surrounds the brain).

After the protomap is established, the final pattern of gyri in each area is probably a product of both differential cellular proliferation and subsequent formation of connections during cortical maturation (*11*). Bae *et al*. show that gyrification in human cortex is a regional event controlled by local proliferation that starts at early stages, before the initiation of neuronal connections (*2*, *4*). This supports the existence of mechanisms that locally regulate gyrification, but the extent to which they are intermingled with the mechanisms of protomap patterning or the subsequent formation of abnormal connections requires further investigation.

The study by Bae *et al*. raises questions about how new convolutions and even new cortical lobes can be created during evolution by simple genomic mutations (*2*). That areas and gyri can be modified or created at will in the laboratory by altering a single gene or factor that controls morphogenesis provides insight into the possible mechanisms of cortical development (*6*–*10*, *12*, *13*), but whether such experimentally induced changes are biologically useful is another question. Bae *et al*. have shown that a single gene mutation can reconfigure the cortex in a functionally deleterious manner, but through evolution, such a mutation may prove to be advantageous for the survival of species. The *GPR56* promoter can drive gene expression in the lateral cortex in nonhuman species, indicating that it is important, but not sufficient, to create a new area such as Broca's, so the search for the decisive gene(s) should continue. The study by Bae *et al*. demonstrates the potential of using advanced methods in human to identify specific genes and test their function in animals in order to obtain information on the origin of human uniqueness (*2*, *14*, *15*).

References and Notes

1. S. Gould, *The Mismeasure of Man* (Norton, New York, 1996).
2. D. H. Geschwind, P. Rakic, *Neuron* **80**, 633 (2013).
3. B.-I. Bae *et al*., *Science* **343**, 764 (2014).
4. P. Rakic, *Nat. Rev. Neurosci.* **10**, 724 (2009).
5. R. F. Hevner, T. F. Haydar, *Cereb. Cortex* **22**, 465 (2012).
6. T. Fukuchi-Shimogori, E. A. Grove, *Science* **294**, 1071 (2001).
7. J. A. Cholfin, J. L. Rubenstein, *J. Comp. Neurol.* **509**, 144 (2008).
8. D. D. O'Leary, D. Borngasser, *Cereb. Cortex* **16** (suppl. 1), i46 (2006).
9. P. Rakic, *Science* **241**, 170 (1988).
10. B. G. Rash, S. Tomasi, H. D. Lim, C. Y. Suh, F. M. Vaccarino, *J. Neurosci.* **33**, 10802 (2013).
11. D. C. Van Essen, *Nature* **385**, 313 (1997).
12. A. Chenn, C. A. Walsh, *Science* **297**, 365 (2002).
13. T. F. Haydar, C. Y. Kuan, R. A. Flavell, P. Rakic, *Cereb. Cortex* **9**, 621 (1999).
14. G. Clowry, Z. Molnár, P. Rakic, *J. Anat.* **217**, 276 (2010).
15. K. Y. Kwan *et al*., *Cell* **149**, 899 (2012).

Acknowledgments: We thank A. Ayoub and J. Arellano for helpful discussion. B.G.R. and P.R. are supported by NIH grants DA023999 and NS014841.

10.1126/science.1250246

Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors

Andrea Pauli,* Megan L. Norris, Eivind Valen, Guo-Liang Chew, James A. Gagnon, Steven Zimmerman, Andrew Mitchell, Jiao Ma, Julien Dubrulle, Deepak Reyon, Shengdar Q. Tsai, J. Keith Joung, Alan Saghatelian, Alexander F. Schier*

READ THE FULL ARTICLE ONLINE
http://dx.doi.org/10.1126/science.1248636

Cite this article as A. Pauli *et al.*, *Science* **343**, 1248636 (2014).
DOI: 10.1126/science.1248636

Introduction: Embryogenesis is thought to be directed by a small number of signaling pathways with most if not all embryonic signals having been identified. However, the molecular control of some embryonic processes is still poorly understood. For example, it is unclear how cell migration is regulated during gastrulation, when mesodermal and endodermal germ layers form. The goal of our study was to identify and characterize previously unrecognized signals that regulate embryogenesis.

Methods: To identify uncharacterized signaling molecules, we mined zebrafish genomic data sets for previously non-annotated translated open reading frames (ORFs). One such ORF encoded a putative signaling protein that we call Toddler (also known as Apela/Elabela/Ende). We analyzed expression, production, and secretion of Toddler using RNA in situ hybridization, mass spectrometry, and Toddler–GFP fusion proteins, respectively. We used transcription activator-like effector (TALE) nucleases to generate frame-shift mutations in the *toddler* gene. To complement loss-of-function analyses with gain-of-function studies, Toddler was misexpressed through mRNA or peptide injection. We characterized phenotypes using marker gene expression analysis and in vivo imaging, using confocal and lightsheet microscopy. *Toddler* mutants were rescued through global or localized *toddler* production. The relationship between Toddler and APJ/Apelin receptors was studied through genetic interaction and receptor internalization experiments.

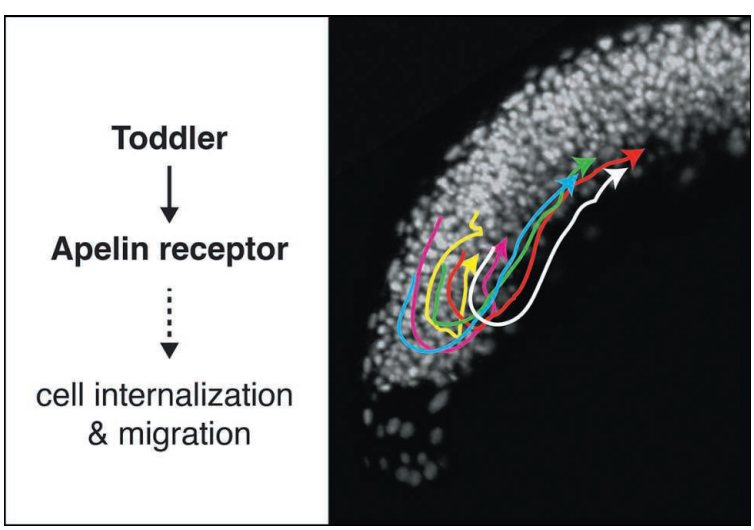
Results: We identified several hundred non-annotated candidate proteins, including more than 20 putative signaling proteins. We focused on the functional importance of the short, conserved, and secreted peptide Toddler. Loss or overproduction of Toddler reduced cell movements during zebrafish gastrulation; mesodermal and endodermal cells were slow to internalize and migrate. Both the local and ubiquitous expression of Toddler were able to rescue gastrulation movements in *toddler* mutants, suggesting that Toddler acts as a motogen, a signal that promotes cell migration. Toddler activates G-protein–coupled APJ/Apelin receptor signaling, as evidenced by Toddler-induced internalization of APJ/Apelin receptors and rescue of *toddler* mutants through expression of the known receptor agonist Apelin.

Discussion: These findings indicate that Toddler promotes cell movement during zebrafish gastrulation by activation of APJ/Apelin receptor signaling. Toddler does not seem to act as a chemo-attractant or -repellent, but rather as a global signal that promotes the movement of mesendodermal cells. Both loss and overproduction of Toddler reduce cell movement, revealing that Toddler levels need to be tightly regulated during gastrulation. The discovery of Toddler helps explain previous genetic studies that found a broader requirement for APJ/Apelin receptors than for Apelin. We propose that in these cases, Toddler—not Apelin—activates APJ/Apelin receptor signaling. Our genomics analysis identifying a large number of candidate proteins that function during embryogenesis suggests the existence of other previously uncharacterized embryonic signals. Applying similar genomic approaches to adult tissues might identify additional signals that regulate physiological and behavioral processes.

FIGURES IN THE FULL ARTICLE

SUPPLEMENTARY MATERIALS

Materials and Methods
Figs. S1 to S19
Full Reference List
Table S1
Data files S1 and S2



Toddler promotes gastrulation movements via Apelin receptor signaling. Toddler is an essential, short, conserved embryonic signal that promotes cell migration during zebrafish gastrulation. The internalization movement highlighted by the colored cell tracks requires Toddler signaling. Toddler signals via the G-protein–coupled APJ/Apelin receptor and may be one of several uncharacterized embryonic signals.

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: pauli@fas.harvard.edu (A.P.); schier@fas.harvard.edu (A.F.S.)

Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors

Andrea Pauli,[1]* Megan L. Norris,[1]† Eivind Valen,[1]† Guo-Liang Chew,[1] James A. Gagnon,[1] Steven Zimmerman,[1] Andrew Mitchell,[2] Jiao Ma,[2] Julien Dubrulle,[1] Deepak Reyon,[3,4] Shengdar Q. Tsai,[3,4] J. Keith Joung,[3,4,5] Alan Saghatelian,[2] Alexander F. Schier[1,5,6,7,8]*

It has been assumed that most, if not all, signals regulating early development have been identified. Contrary to this expectation, we identified 28 candidate signaling proteins expressed during zebrafish embryogenesis, including Toddler, a short, conserved, and secreted peptide. Both absence and overproduction of Toddler reduce the movement of mesendodermal cells during zebrafish gastrulation. Local and ubiquitous production of Toddler promote cell movement, suggesting that Toddler is neither an attractant nor a repellent but acts globally as a motogen. Toddler drives internalization of G protein–coupled APJ/Apelin receptors, and activation of APJ/Apelin signaling rescues *toddler* mutants. These results indicate that Toddler is an activator of APJ/Apelin receptor signaling, promotes gastrulation movements, and might be the first in a series of uncharacterized developmental signals.

Many of the inductive events during early development are directed by a small number of signaling pathways whose agonists have been known for more than a decade (*1*). Therefore, it has been assumed that most, if not all, embryonic signals have been identified. However, the molecular control of some embryonic processes is still poorly understood. For example, it is largely unclear how cell migration is regulated during gastrulation or how cells coalesce into discrete tissues during organogenesis (*2*–*5*), suggesting that some of the involved signals are yet to be identified. Moreover, recent genomic studies have suggested that translation of short open reading frames (ORFs) and the generation of small peptides are much more pervasive than previously assumed (*6*, *7*). To search for new candidate signaling molecules, we used the Translated ORF Classifier (TOC) (*7*) to examine zebrafish transcript annotations and ribosome profiling data sets (*7*–*9*) for non-annotated translated ORFs (Fig. 1A) (materials and methods in the supplementary materials). This analysis identified 700 novel protein-coding transcripts (399 loci) (supplementary data files S1 and S2), of which 81% (562 transcripts in 325 loci) shared nucleotide sequence alignments with other vertebrates (table S1). Notably, this approach identified 28 candidate signaling proteins (40 transcript isoforms) characterized by the presence of putative signal sequences and lack of predicted transmembrane domains (table S1). Ribosome profiling and phylogenetic analysis suggest that these RNAs can generate secreted peptides with lengths ranging from 32 to 556 amino acids (Fig. 1A, fig. S1, and table S1). Although these genes have not been identified previously or are annotated in the zebrafish Ensembl database as noncoding RNAs, the majority (24 of 28) appear to be conserved in other vertebrates (fig. S1 and table S1).

Toddler Encodes a Short, Conserved, and Secreted Peptide

To test the functional potential of these candidate signals, we focused on a gene that we named *toddler* on the basis of the phenotype described below (Fig. 1B). *Toddler* (*tdl*) mRNA is expressed ubiquitously during late blastula and gastrula stages and becomes restricted to the lateral mesoderm, endoderm, and anterior and posterior notochord after gastrulation (Fig. 1C). *Toddler* is annotated as a noncoding RNA in zebrafish (*ENSDARG00000094729*), mouse [*Gm10664*; also called *Ende* (*10*)], and human (*LOC100506013*) (fig. S2) and is present in two lncRNA catalogs (*9*, *11*); however, it contains a 58–amino acid ORF with a predicted signal sequence and high conservation in vertebrates, including human (Fig. 1D and fig. S3). Sequence comparisons with the highly conserved C-terminal portion did not identify homology to any other known proteins, raising the possibility that this gene encodes an uncharacterized embryonic signal.

Six lines of evidence indicate that *toddler* is translated and encodes a secreted peptide. First, phylogenetic comparisons of synonymous versus nonsynonymous codon changes reveal strong amino acid preservation in the *toddler* ORF (PhyloCSF score of 98 (*8*); see Fig. 1, B and D, and table S1). Second, previous ribosome profiling data in mouse (*6*) and zebrafish (*7*) indicate that the *toddler* ORF is protected by actively translating ribosomes in vivo (Fig. 1B). Third, mass spectrometric analysis of nontrypsinated protein extracts from embryos expressing *toddler* mRNA detected the 11–amino acid C-terminal Toddler peptide fragment that is predicted to be a convertase cleavage product (Fig. 1D and fig. S4). Fourth, enhanced green fluorescent protein (eGFP) fusion proteins containing the wild-type signal sequence of Toddler are found extracellularly, whereas signal peptide cleavage site mutants are retained in the cell (Fig. 1E). Fifth, as described below, extracellular injection of in vitro–synthesized Toddler peptide (C-terminal 21 amino acids) elicits the same gain-of-function phenotypes as excess of *toddler* mRNA. Sixth, wild-type but not frameshifted *toddler* mRNA rescues *toddler* mutants (see below), providing direct evidence that it is the peptide product rather than the RNA that is functional in vivo. Together, these findings identify Toddler as a short, conserved, and secreted peptide.

Toddler Is Essential for Embryogenesis

To disrupt *toddler* function, we generated mutants by TALEN-mediated mutagenesis (fig. S5 and materials and methods) (*12*, *13*). Seven *toddler* alleles were recovered, each of which introduces a frameshift immediately after the signal peptide sequence (fig. S5, B and C). The vast majority of homozygous *toddler* mutants die between 5 and 7 days of development and display small or absent hearts, posterior accumulation of blood cells, malformed pharyngeal endoderm, and abnormal left-right positioning and formation of the liver (Fig. 2, A and B, and fig. S6). Penetrance and expressivity of *toddler* mutants vary, including occasional escapers that live to adulthood and rare instances of *toddler* mutants that display more severe defects in endoderm and mesoderm formation (fig. S7). Notably, the lethality of *toddler* mutants (survival, 0 of 25 animals) was rescued by injection of low amounts (2 pg) of wild-type (survival, 23 of 30 animals) but not frameshifted (survival, 0 of 32 animals) *toddler* mRNA (Fig. 2, A, C, and D). Embryonically rescued *toddler* mutants survived to adulthood and were fertile in the absence of any later source of Toddler peptide, indicating that zebrafish Toddler is only essential during early embryogenesis.

Toddler Is Required for Normal Gastrulation Movements

To determine when Toddler function is required during early embryogenesis, we used a heat shock–inducible transgene. Induction of *toddler* expression during late blastula and early gastrula stages,

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. [2]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. [3]Molecular Pathology Unit, Center for Computational and Integrative Biology, and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA 02129, USA. [4]Department of Pathology, Harvard Medical School, Boston, MA 02115, USA. [5]The Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA. [6]FAS Center for Systems Biology, Harvard University, Cambridge, MA 02138, USA. [7]Center for Brain Science, Harvard University, Cambridge, MA 02138, USA. [8]Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138, USA.

*Corresponding author. E-mail: pauli@fas.harvard.edu (A.P.); schier@fas.harvard.edu (A.F.S.)

†These authors contributed equally to this work.

but not at later times, rescued *toddler* mutants (fig. S8 and materials and methods).

The early requirement for Toddler, together with its expression peak during gastrulation (Fig. 1C), suggested that the later phenotypes originate from earlier defects. We therefore analyzed morphology and gene expression during blastula and gastrula stages and discovered that *toddler* mutant mesendodermal progenitors did not move properly toward the animal pole during gastrulation. Although ventral and lateral mesendodermal cells in wild-type embryos internalized at the margin and moved toward the animal pole (Fig. 2, C and E), these cells were closely packed and confined to a band near the margin in *toddler* mutant embryos (Fig. 2, C and D, and fig. S9). These defects were apparent by analysis of endodermal (*sox17*) and mesodermal (*fibronectin1/fn1*, *spadetail/tbx16*, *fascin*, *draculin/drl*) markers (Fig. 2C and fig. S9). In contrast, ectodermal (*sox3*), dorsal mesodermal (*goosecoid/gsc*, *hgg1*), and tail mesodermal (*ntla*) markers were largely unchanged in their expression domains (fig. S10). In addition to the ventrolateral movement defects, *toddler* mutants contained ~20% fewer endodermal cells at mid-gastrulation (Fig. 2, C and D, and fig. S9A). The initial expression of mesendodermal markers appeared unaffected (fig. S10B), suggesting that mesendodermal cells are specified normally in *toddler* mutant embryos but proliferate less. Notably, the *toddler* gastrulation phenotypes could be rescued by injecting low levels (2 pg) of *toddler* mRNA at the one-cell stage (Fig. 2, C and D, and fig. S9, A and C). These results reveal an important role for Toddler in the movement of ventral and lateral mesendodermal cells during gastrulation.

Toddler Promotes Endodermal and Mesodermal Cell Migration

To determine how Toddler affects the movement of cells during gastrulation, we performed live cell imaging and followed cell trajectories in wild-type and

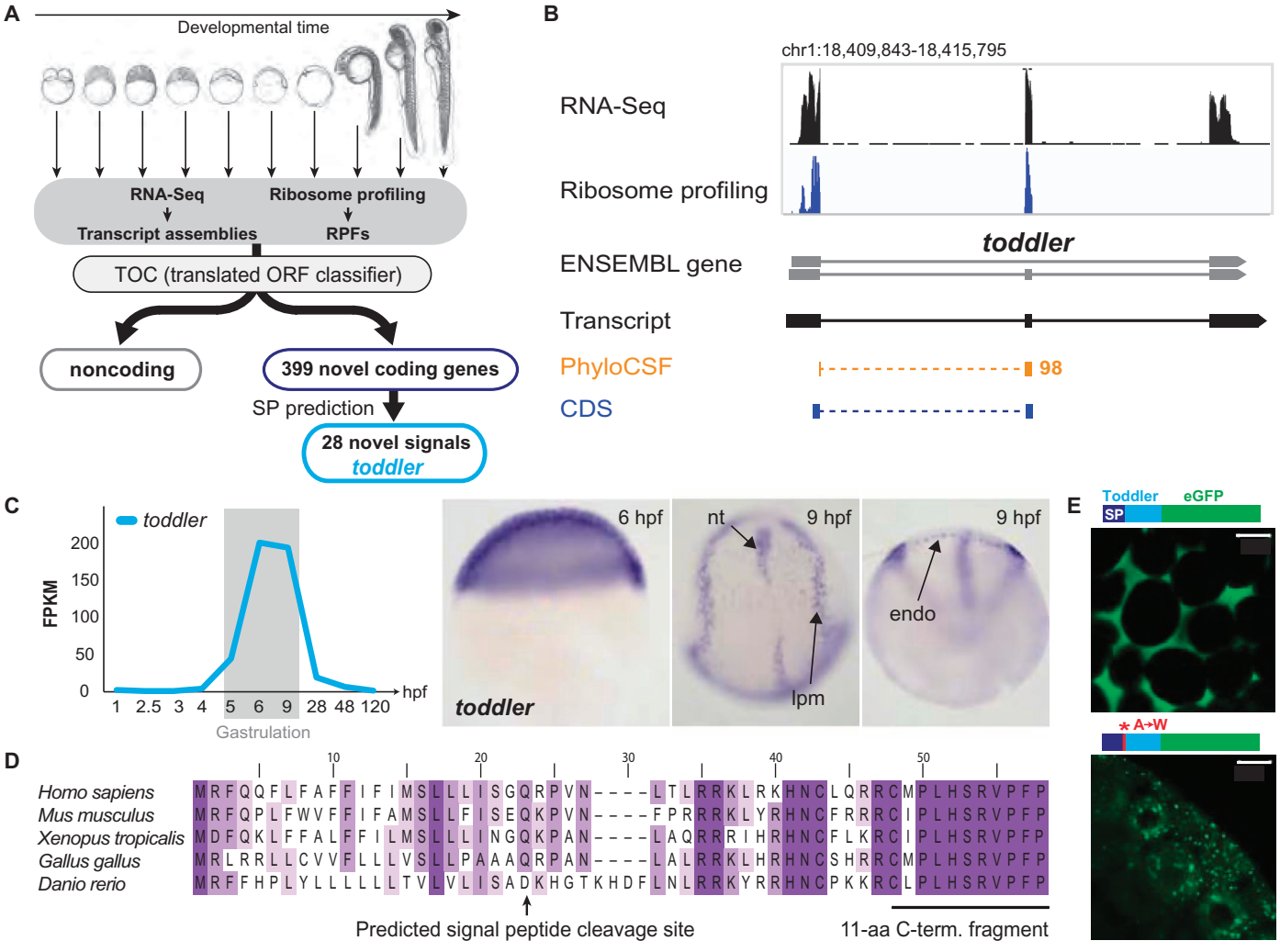


Fig. 1. Identification of the novel embryonic signal Toddler. (**A**) Overview of the individual steps used to identify novel coding and noncoding transcripts. SP, signal peptide; RPFs, ribosome protected fragments. (**B**) Genomic features of *toddler*. Coverage tracks for RNA-Seq (black) and ribosome profiling (blue), and tracks outlining the highest scoring regions in PhyloCSF (orange). Note that both PhyloCSF (*8*) and ribosome profiling (*7*) predict *toddler* to be protein-coding. (**C**) Expression analysis of *toddler* transcripts during embryogenesis. *toddler* transcripts peak during gastrulation [RNA-Seq data (*8*)]. FPKM, fragments per kilobase of transcript per million mapped reads. RNA in situ hybridization reveals ubiquitous expression of *toddler* transcripts at the beginning of gastrulation [6 hours postfertilization (hpf)]; expression becomes restricted to mesendodermal cells toward the end of gastrulation (9 hpf). nt, notochord; lpm, lateral plate mesoderm; endo, endoderm. (**D**) Toddler is conserved in vertebrates. ClustalW2 multiple protein sequence alignment of Toddler peptide sequences from five vertebrates. Darker shading indicates higher percentage identity of the amino acid. The predicted signal peptide cleavage site and the highly conserved C-terminal 11–amino acid (aa) peptide fragment that was detected by mass spectrometry are indicated. (**E**) Toddler signal sequence drives secretion. Injection of mRNAs encoding C-terminal Toddler-eGFP fusion proteins reveals that the wild-type Toddler signal sequence drives secretion (extracellular localization of eGFP), whereas mutation of A→W in the signal peptide cleavage site causes Toddler-eGFP to remain intracellularly (top, wild-type Toddler ORF; bottom, A→W mutant Toddler ORF). Fusion protein diagrams are not drawn to scale. Scale bars, 20 μm.

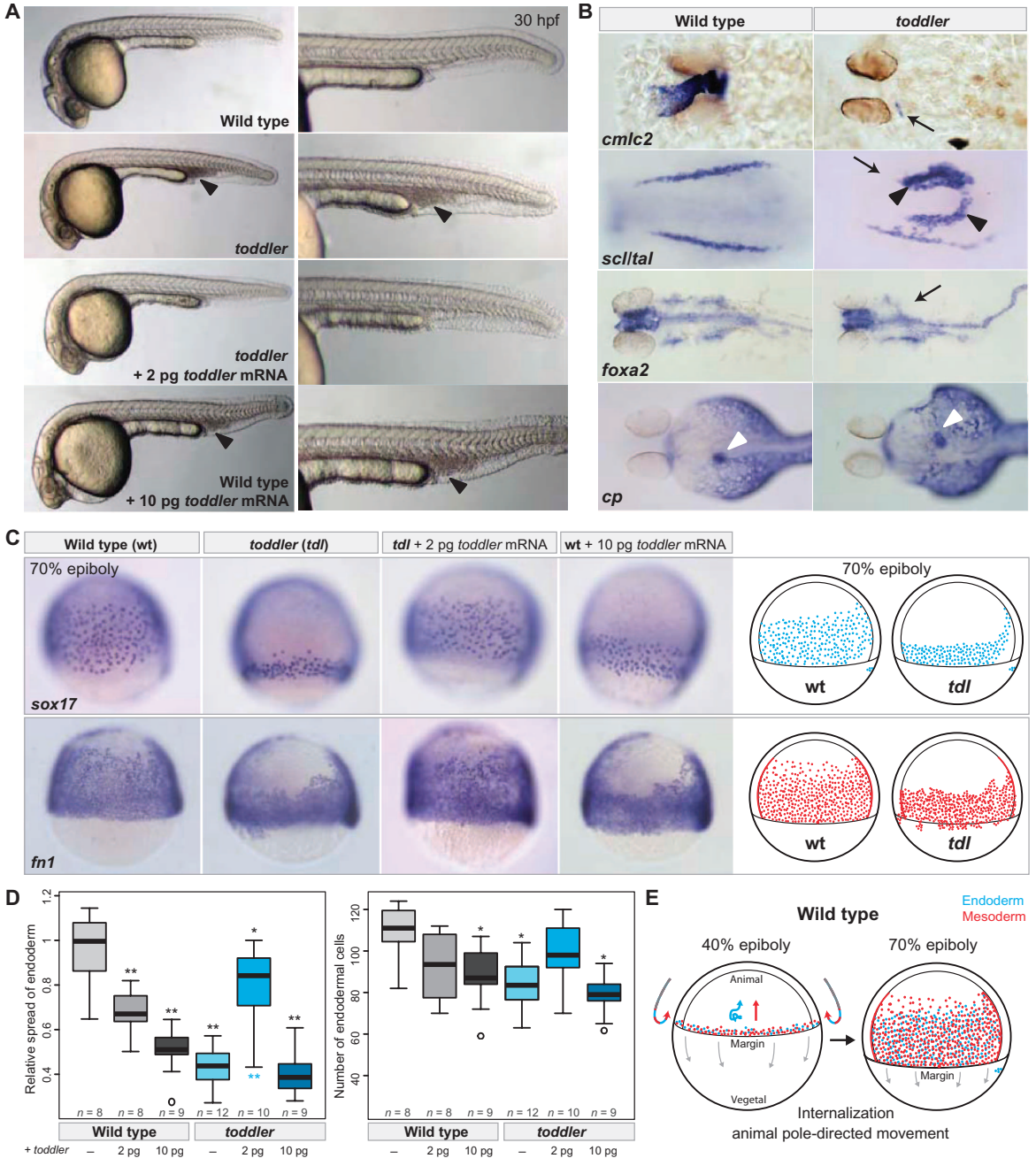


Fig. 2. Toddler is essential for embryogenesis. (**A**) Morphological analysis of *toddler* mutants. TALEN-induced *toddler* null mutants (see fig. S5) lack a functional heart, have no blood circulation, and accumulate blood posteriorly (black arrowheads). Defects in *toddler* mutant embryos are rescued by low doses (2 pg) of *toddler* mRNA. Injection of higher doses of *toddler* mRNA (≥10 pg) causes phenotypes in wild-type embryos reminiscent of *toddler* loss-of-function mutants. Shown are lateral views of embryos of the indicated genotypes at 30 hpf. (**B**) Marker gene analysis in wild-type and *toddler* mutant embryos at 36 hpf (*cmlc2*), at the 8 to 10 somite stage (*scl/tal*), at 30 hpf (*foxa2*), and at 3 days postfertilization [*ceruloplasmin* (*cp*)]. Black arrows indicate lack of or reduced staining in *toddler* mutant embryos; black arrowheads indicate ectopic expression; white arrowheads point to the liver in wild-type (>70% on left side) and *toddler* mutant embryos (expression: 45% right, 15% medial, 40% none/nonspecific). (**C**) Toddler is required for movement of ventrolateral endoderm and mesoderm toward the animal pole. Both absence of Toddler (*toddler*) and overexpression of *toddler* mRNA (wild-type embryos + 10 pg of *toddler* mRNA) reduce the movement of endodermal (*sox17*) and mesodermal [*fibronectin 1* (*fn1*)] cells toward the animal pole, as detected by in situ hybridization. All in situ images are lateral views of embryos at 70% epiboly (dorsal to the right). Illustrations of the observed endodermal (blue) and mesodermal (red) phenotypes in wild-type (wt) and *toddler* mutant (*tdl*) embryos are shown on the right. (**D**) Quantification of the endodermal defects at 70% epiboly. Left, relative spread of lateral endoderm along the animal-vegetal axis (that is, height of lateral band of *sox17*-expressing cells divided by the wild-type mean); right, number of endodermal cells within a lateral, fixed-size area. Gray, wild-type genomic background; cyan, *toddler* mutant genomic background. *P* values for pairwise comparisons with wild-type (black, top) or *toddler* mutant (cyan, bottom) were calculated on the basis of a standard Welch's *t* test (*$P < 0.01$; **$P < 0.00001$). (**E**) Illustration of early gastrulation movements in wild-type zebrafish embryos. Mesodermal (red) and endodermal (blue) cells are induced and internalized at the margin (40% epiboly stage). Whereas internalized cells migrate toward the animal pole in either a directional (mesoderm) or random walk–like pattern (endoderm) (*3*, *15*), epiboly movements are directed toward the vegetal pole (gray arrows). At 70% epiboly, mesodermal and endodermal cells have moved animally and cover most of the lateral side of the embryo.

toddler mutant embryos (movies S1 to S6). *Toddler* mutant endodermal cells [*sox17::GFP* (*14*)] displayed reduced movement toward the animal pole (Fig. 3A, fig. S11, and movies S1 and S2), migrated more slowly, and showed reduced net (start-to-end) displacement compared to wild-type cells (Fig. 3B and fig. S11). During early gastrulation, *toddler* mutant endodermal cells exhibited the characteristic random walk–

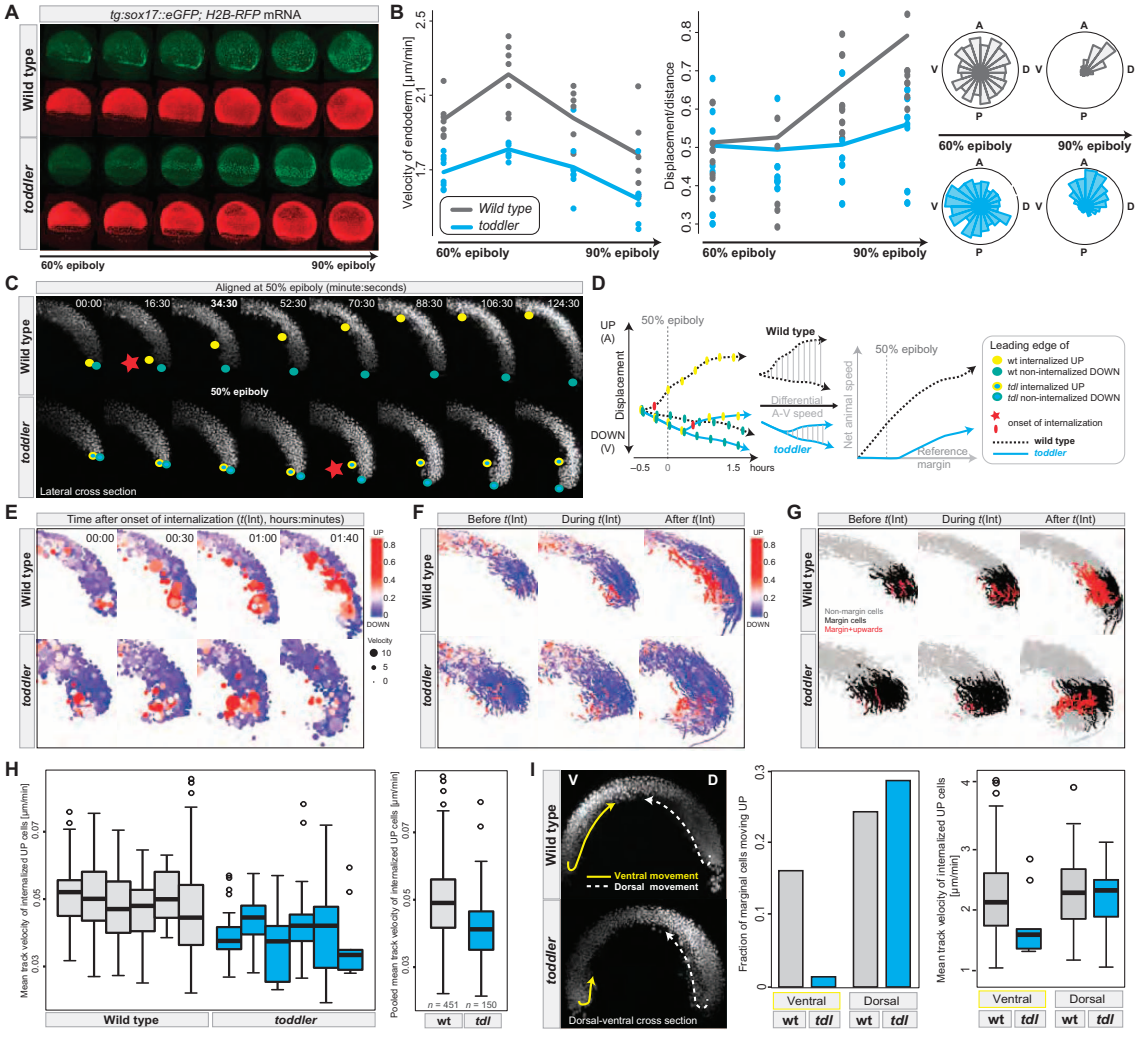


Fig. 3. Abnormal gastrulation movements in *toddler* mutants. (**A** and **B**) Analysis of endodermal cell migration in *sox17::eGFP* transgenic wild-type and *toddler* mutant embryos by confocal microscopy. Green, endodermal cells (marked by *sox17::eGFP*); red, nuclei [*human histone2B-RFP* (*H2B-RFP*) mRNA injection]. (A) Still images of maximum intensity projections of a time-lapse movie from 60 to 90% epiboly (movies S1 and S2). (B) Quantification of the average (median) velocity of endodermal cells (left), displacement versus distance travelled (middle), and directionality (rose-plots; right) in wild-type (gray) and *toddler* mutant (cyan) embryos. Each dot represents the average speed (or the ratio between displacement versus distance travelled) of all endodermal cells tracked within a single embryo during a 45-min time interval with respect to its previous position [speed = actual distance (micrometers)/time (min)]. Shown are the data for four consecutive 45-min time windows. Roseplots display the random movement of endodermal cells during early gastrulation and the more directional migration at later stages [animal (A), posterior (P), dorsal (D), ventral (V)]. (**C** to **I**) Analysis of early gastrulation movements in *H2B-RFP* mRNA injected wild-type and *toddler* mutant embryos by light-sheet microscopy (single-plane illumination microscopy). (C to H) Internalization and animal pole–directed movement of lateral mesendodermal cells are reduced in *toddler* mutants. Analyses are shown for lateral cross sections of a time-lapse movie (movie S4) of a wild-type–*toddler* mutant embryo pair, imaged in parallel at 90-s intervals within a single experiment. (C) Still images of maximum intensity projections of 40-µm lateral cross sections (20 z-slices) during the time of internalization (time in minutes:seconds). Movies were aligned at 50% epiboly (48:00). Leading edges of internalizing mesendodermal cells (yellow dots) and vegetally moving cells (green dots) highlight the opposing paths of cells during gastrulation. Red stars mark the onset of cell internalization. (D) Comparison of animally and vegetally directed migratory paths of the wild-type and mutant embryo pair shown in (C). Frame-to-frame displacements (plotted on the left) were used to derive the net animal pole–directed cell movement. *Toddler* mutants (cyan) show delayed onset of internalization and reduced step-to-step and net animal pole–directed movement. (E to G) Cell tracking and digital analysis of gastrulation movements. (E) Position, speed (dot size), and directionality [color-coded from blue (vegetal movement) to red (animal movement)] of tracked cells during and after the time of internalization [*t*(Int)]. Movies were aligned to the onset of internalization [*t*(Int) = 00:00; time in hours:minutes]. (F and G) Cell tracks before ($t < -5$ min), during (-5 min $< t <$ 1 hour), and after ($t > 1$ hour) internalization in wild-type and *toddler* mutant embryos. In (F), tracks were color-coded on the basis of the total number of animal pole–directed (red) or vegetal pole–directed (blue) movements, normalized to the total number of frames per track. In (G), tracks were color-coded on the basis of their relative position and directionality with respect to the margin at the time of internalization (margin cells: cells located within 100 µm above the margin at the onset of internalization). Gray, nonmargin cells; black, margin cells; red, internalizing and upward-moving margin cells. (H) Quantification of the mean velocity of internalizing, animal pole–directed movement in wild-type and toddler mutant embryos. Mean track velocities were obtained from cell-tracking data derived from lateral cross sections of six wild-type (gray) and six *toddler* mutant (cyan) embryos, imaged in parallel in three independent experiments. Pooled wild-type and *toddler* mutant mean track velocities are plotted on the right (n = number of cell tracks). (I) *Toddler* mutant embryos are defective in ventrolateral but not dorsal internalization. (Left) Still image of maximum intensity projections of 40-µm dorsal-ventral cross sections (20 z-slices) of a wild-type–*toddler* mutant embryo pair 110 min after the onset of internalization. Arrows highlight the paths that the leading internalizing cells took on dorsal (D, dashed white line) and ventral (V, solid yellow line) sides of the embryo. Ventral movement toward the animal pole is severely reduced in the *toddler* mutant embryo, whereas dorsal internalization occurs normally. (Right) Quantification of the fraction and speed of internalizing marginal cells based on their positioning in the embryo (dorsal versus ventral) and genotype [wild type (gray) versus *toddler* mutant (cyan)] (see also movie S6).

like migration pattern observed in wild-type embryos (*3*, *15*), but they migrated in a less directional fashion than their wild-type counterparts during later gastrulation (movie S1 and Fig. 3B).

To analyze the earliest steps of mesendoderm movement, we followed the paths of H2B-RFP–labeled nuclei by light-sheet microscopy in wild-type and *toddler* mutant embryos (movie S3 and fig. S12). Analysis of 10 wild-type and 11 *toddler* mutant embryos confirmed that the movement of ventrolateral but not dorsal internalizing cells toward the animal pole was impaired in *toddler* mutants (Fig. 3, C to I, figs. S12 to S14, and movies S3 to S6). Internalization of ventrolateral cells at the margin was delayed (Fig. 3, C and D, fig. S13A, and movies S4 and S5) and reduced (Fig. 3, E to G and I, fig. S13, and movies S3 to S6). Although internalization in wild-type embryos started about 30 min before embryos reached 50% epiboly, it often commenced only after the 50% epiboly stage in *toddler* mutants (Fig. 3, C and D, fig. S13A, and movies S4 and S5). Ventrolateral internalized cells moved more slowly (Fig. 3, H and I) and often piled up at the margin (Fig. 3, C and E, figs. S13 to S15, and movies S3 to S6). In addition, epiboly movements were often delayed in *toddler* mutants, particularly during the time of internalization (fig. S13, A and B). In rare cases, we observed an almost complete absence of animal pole–directed ventrolateral cell movements; in these embryos, ventral and lateral marginal cells instead moved vegetally (movies S3, S5, and S6), likely contributing to the ectopic accumulation of posteriorly located blood cells at later stages (Fig. 2, A and B). These results identify Toddler as a key signal that promotes the internalization and animal pole–directed movement of mesendodermal cells during gastrulation.

Overexpression of Toddler Phenocopies *Toddler* Mutants

In contrast to inducers of specific cell fates, many signals involved in cell migration or tissue morphogenesis share loss- and gain-of-function phenotypes. For example, both reduction and increase in Wnt/planar cell polarity signaling disrupt convergence and extension movements during gastrulation (*2*–*5*). To determine whether Toddler might share this feature, we carried out overexpression analyses. Injection of *toddler* mRNA at levels only five times higher (≥10 pg) than needed for rescue caused phenotypes in wild-type embryos that resembled *toddler* loss-of-function mutants, including gastrulation and heart defects (Fig. 2, A, C, and D, and fig. S9, A and C). Similar phenotypes were observed upon extracellular injection of an in vitro–synthesized Toddler peptide fragment (C-terminal 21 amino acids; fig. S16). These observations reveal that proper levels of Toddler are required for normal mesendodermal movement and provide further evidence of an important role for Toddler in cell migration.

Toddler Functions as a Motogen

Most genes encoding signals that attract or repel cells are expressed in specific domains (*16*), and ubiquitous production of such signals interferes with guided cell migration. In contrast, we find that *toddler* RNA is expressed ubiquitously (Fig. 1C and fig. S17A) and that ubiquitous expression of *toddler* mRNA upon injection at the one-cell stage promotes the normal movement of mesendodermal cells in *toddler* mutants (Fig. 2, C and D). To further test the role of Toddler in cell migration, we locally expressed Toddler in the vegetal or animal regions of *toddler* mutants. In both scenarios, localized Toddler production was able to promote the migration of mesendodermal cells and rescue *toddler* mutants (Fig. 4). Although more complex scenarios are formally possible [for example, local processing (*17*) and self-generated gradient formation (*18*, *19*)], these results suggest that Toddler does not attract cells to or repel cells from specific sites. Instead, Toddler appears to act as a motogen (*20*–*22*)—a general promoter of mesendodermal cell migration.

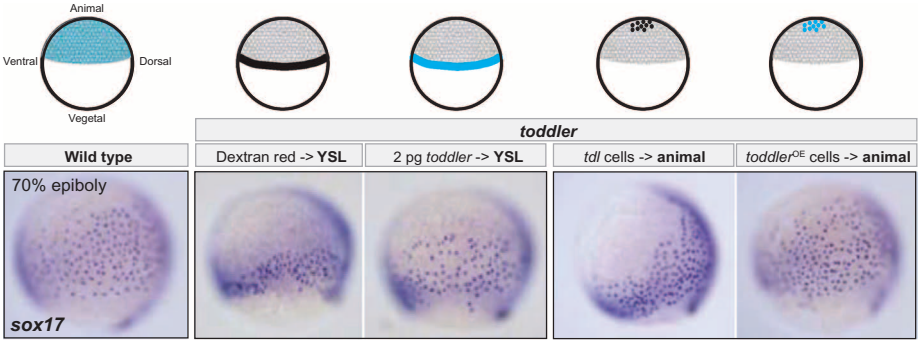


Fig. 4. Toddler functions as a motogen. Ubiquitous or localized expression of Toddler promotes animal pole–directed endodermal cell migration in *toddler* mutant embryos. Toddler was expressed either vegetally from the yolk syncytial layer (YSL) (injection of *toddler* mRNA into the YSL) or animally from a *toddler*-overexpressing (OE) clone of cells transplanted into the animal pole. Dextran red injections into the YSL and transplantation of uninjected *toddler* mutant cells served as controls. Different treatments are illustrated on top; *toddler* expression domains are highlighted in cyan. All *sox17* in situ hybridization images are lateral views of embryos at 70% epiboly (dorsal to the right).

Toddler Acts via APJ/Apelin Receptors

To identify candidate receptors for Toddler, we compared the *toddler* phenotype to previously described receptor mutant phenotypes. On the basis of the small size of Toddler peptide and the involvement of G protein signaling in gastrulation movements, we focused on G protein–coupled receptors (GPCRs) as candidate Toddler receptors (*14*, *23*–*30*). Four observations raised the possibility that the G protein–coupled APJ/Apelin receptor might mediate Toddler signaling. First, loss of APJ/Apelin receptor signaling results in small hearts and affects lateral mesoderm migration in zebrafish (*24*–*26*), phenotypes reminiscent of some aspects of the *toddler* mutant phenotype. However, in contrast to the broad roles of Toddler in mesendoderm migration, APJ/Apelin receptor signaling had been specifically implicated in cardiovascular development (*24*–*26*, *31*–*36*). Second, overexpression of Apelin, the only known ligand for the APJ/Apelin receptor (*35*–*38*), interferes with gastrulation movements in zebrafish (*24*–*26*). Third, the expression levels of Apelin receptors and Toddler peak during gastrulation (Fig. 5A), and Apelin receptors are expressed in mesendodermal cells [fig. S16A and (*24*, *25*, *39*)], the cell types affected in *toddler* mutants. Fourth, we found that Apelin is expressed only at the end of gastrulation [Fig. 5A and (*24*)], after the *toddler* and *APJ/apelin receptor* phenotypes (*24*, *25*, *40*) are apparent. These observations, together with the milder phenotypes observed in the absence of Apelin as compared to loss of APJ/Apelin receptors (*24*–*26*, *34*, *36*, *41*–*46*), raised the hypothesis that Toddler might be the bona fide activator of APJ/Apelin receptor signaling during gastrulation. We tested three predictions of this model.

First, we determined whether the absence of Apelin receptor function phenocopies *toddler* mutants. We reexamined *aplnra* and *aplnrb* double morphants (*24*–*26*) and found phenotypes that were highly similar to *toddler* mutants, including reduced movement of ventrolateral mesendoderm during gastrulation (Fig. 5, B and C). We also confirmed and extended previous analyses of the effects of Apelin overexpression (*24*–*26*) and found defects very similar to those caused by Toddler overexpression (Fig. 5, B and C). In addition, we observed that coexpression of Toddler and Apelin receptor at levels that individually did not cause major defects resulted in abnormal gastrulation movements reminiscent of Toddler and Apelin (*24*–*26*) overexpression phenotypes (Fig. 5D). These results reveal shared morphogenetic activities of the Apelin receptor and Toddler signaling pathways.

Second, we tested the epistatic relationship between Toddler and Apelin receptor signaling. The similarity of gain- and loss-of-function phenotypes precluded standard tests such as overexpression of Toddler in Apelin receptor mutants. Instead, we tested whether activation of Apelin receptor signaling can bypass the requirement

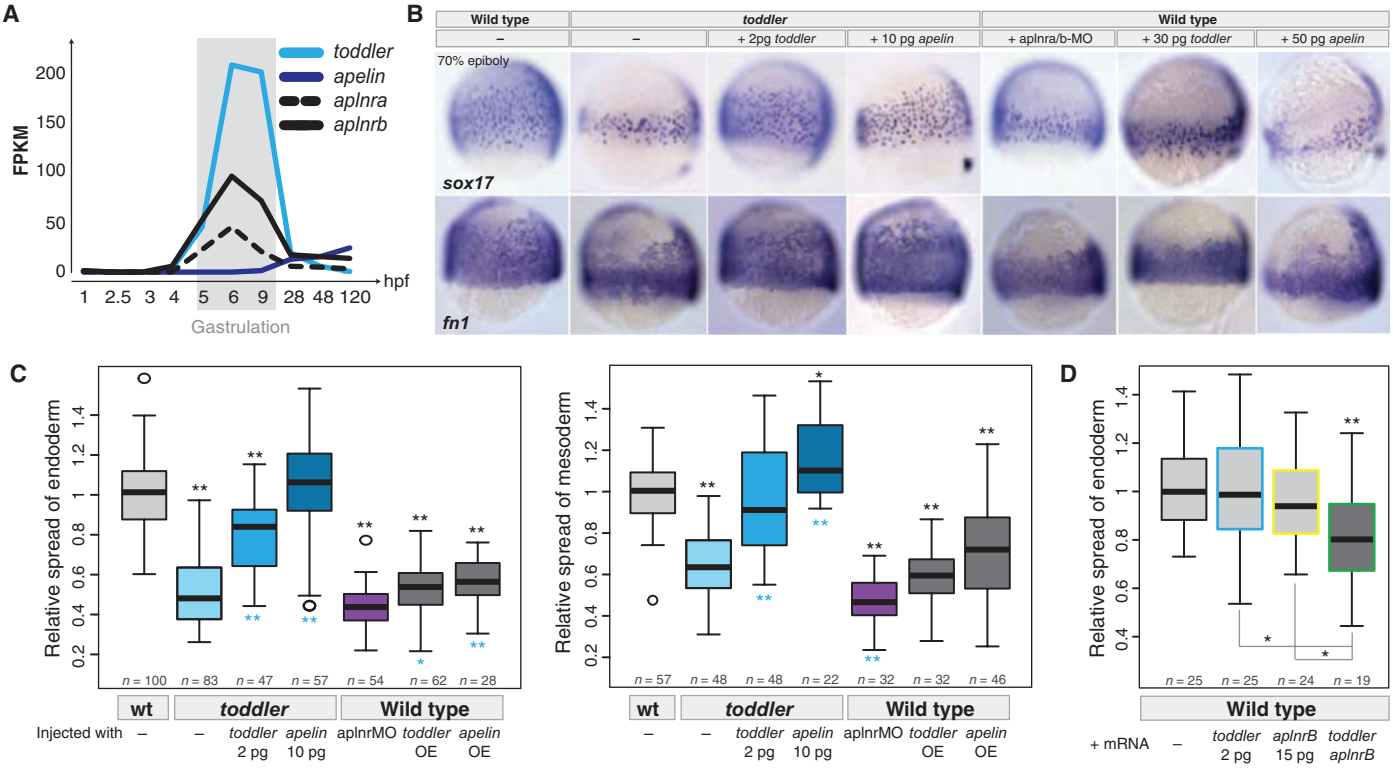


Fig. 5. Toddler acts via Apelin receptors. (**A**) RNA-Seq–based expression levels of *toddler*, *apelin*, and *apelin receptors* (*aplnra* and *aplnrb*) during embryogenesis. (**B**) Genetic evidence for Toddler signaling via the Apelin receptor. Endodermal (*sox17*) and mesodermal [*fibronectin 1* (*fn1*)] cell distributions were analyzed by in situ hybridization at 70% epiboly. Apelin receptor knockdown [aplnra/b morpholino (MO) injection] phenocopies *toddler* mutants, and Apelin production can rescue *toddler* mutants. Overexpression of Apelin causes phenotypes resembling *toddler* mRNA overexpression. (**C**) Quantification of the relative lateral spread of endoderm (left) and mesoderm (right). Quantifications are from multiple experiments (n = number of embryos per category). P values for pairwise comparisons with wild type (black, top) or *toddler* mutant (cyan, bottom) were calculated on the basis of a standard Welch's t test (*$P < 0.01$; **$P < 0.00001$). (**D**) Synergistic effect of Toddler and Apelin receptor b on endodermal cell migration. Injection of *toddler* or *aplnrb* mRNA at low concentrations (2 and 15 pg, respectively) did not cause significant defects in animal pole–directed movement of endodermal cells (different batch of *toddler* mRNA than used in Fig. 2D). However, coinjection of both mRNAs reduced the extent of endoderm movement. Shown are the combined data of two independent experiments. P values for pairwise comparisons with wild type (top) or individual mRNA injections (bottom) were calculated on the basis of a standard Welch's t test (*$P < 0.01$; **$P < 0.00001$).

for Toddler. *Apelin* mRNA injection into *toddler* mutant embryos restored normal mesendoderm migration (Fig. 5, B and C), cardiac development, and survival to adulthood. These results suggest that Toddler and Apelin activate the same signaling pathway.

Third, we tested whether Toddler can drive the internalization of Apelin receptors (Fig. 6), a hallmark of activated GPCR signaling (*47*–*50*). We misexpressed *toddler* mRNA with eGFP-tagged Apelin receptor a or b and observed strong internalization of the receptors from the plasma membrane (Fig. 6B). This effect was specific because other signaling proteins (chemokines Sdf1a/Cxcl12a or Sdf1b/Cxcl12b) did not alter the distribution of membrane-bound Apelin receptors, nor did Toddler alter the distribution of other chemokine receptors (Cxcr4a-eGFP, Cxcr4b-eGFP, and Cxcr7b-eGFP) (Fig. 6B and fig. S18). Moreover, Toddler produced from a local clone of cells was sufficient to cause Aplnrb-eGFP internalization at a distance from the source, suggesting that secreted Toddler peptide can act on neighboring cells (Fig. 6C). This conclusion was further strengthened by the observation that extracellular injection of in vitro–synthesized C-terminal Toddler or Apelin peptides induced efficient internalization of Aplnr-eGFP (Fig. 6D). These results indicate that Toddler activates Apelin receptors.

Discussion

Our study indicates that Toddler is an activator of APJ/Apelin receptor signaling, promotes gastrulation movements (see summary in Fig. 6E), and may be the first in a series of previously unknown developmental signals. While this study was under review, Toddler (named ELABELA) was independently reported to signal via APJ/Apelin receptors during endoderm differentiation and heart formation (*51*). The HUGO Gene Nomenclature Committee (HGNC) has recently designated the name Apela (apelin receptor early endogenous ligand) as the standardized symbol for Toddler/ELABELA/Ende. Our results lead to four major conclusions.

First, Toddler is a previously unrecognized signal that promotes cell movement during gastrulation. The rescue of *toddler* mutants by ubiquitous Toddler expression suggests that Toddler acts neither as a chemoattractant nor as a chemorepellent, but rather as a nondirectional signal to promote the internalization and movement of ventrolateral mesendodermal cells. Dorsal mesendoderm movement is largely unaffected in *toddler* mutants, consistent with the absence of Apelin receptor expression in this region and the role of other pathways in dorsal gastrulation movements (*3*). Both loss and overproduction of Toddler reduce cell movement, revealing that Toddler levels need to be tightly regulated to allow for normal gastrulation. It remains to be determined whether Toddler promotes motility by regulating cell shape, cellular protrusions, cell-substrate interactions, and cell-cell adhesion or through other means.

Second, Toddler-Apelin receptor signaling provides a long-sought link between mesendoderm induction and migration. Nodal signaling not only induces mesendoderm formation (*52*) but also activates the expression of Apelin receptors [fig. S17B and (*39*)]. Thus, Nodal-mediated induction of Apelin receptor expression might render cells competent to respond to Toddler and to become more motile (Fig. 6E). In this scenario, the activation of Apelin receptor expression in

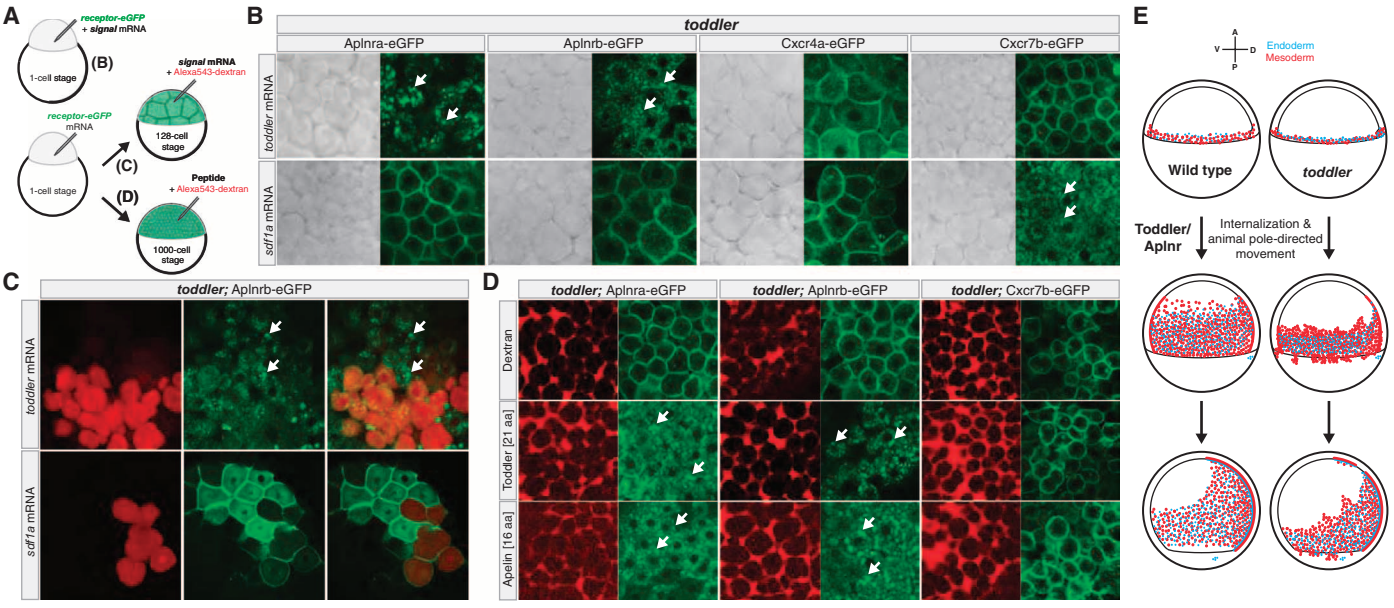


Fig. 6. Toddler drives internalization of Apelin receptors. (**A**) Schematic illustration of different treatments used to test for Toddler-mediated Apelin receptor internalization. (**B**) Test for signal-mediated internalization of eGFP-tagged receptors in zebrafish by coinjection of *signal* and *receptor-eGFP* mRNA into one-cell stage *toddler* mutant embryos. Receptor internalization was monitored by confocal microscopy. White arrows point to fluorescent foci of internalized receptors. In the absence of signal peptide overexpression, ectopically expressed receptors localize to the plasma membrane in pregastrulation *toddler* mutant embryos [see control Alexa543-dextran injections in (D)]. (**C**) Generation of a local source of Toddler or Sdf1a by injection of *toddler* or *sdf1a* mRNA (together with Alexa543-dextran as tracer) into a single cell at the 128-cell stage. Local expression of Toddler is sufficient to cause Aplnrb-eGFP internalization in cells that do not express *toddler* mRNA (non-red cells). (**D**) Extracellular injection of in vitro–synthesized C-terminal Toddler or Apelin peptide fragments is sufficient to drive internalization of Apelin receptors. (**E**) Model of the role of Toddler-Apelin receptor signaling in mesendodermal cell migration during zebrafish gastrulation. Left, wild type; right, *toddler*; top, 40% epiboly (mesendoderm specification and internalization); middle, 70% epiboly (animal pole–directed cell movement); bottom, 90% epiboly (dorsal convergence).

cells located at the margin at the end of the blastula stage would restrict the motogenic effects of Toddler and prevent ectopic and premature cell motility.

Third, Toddler is a novel agonist of APJ/Apelin receptor signaling, as evidenced by Toddler-induced internalization of Apelin receptors and rescue of *toddler* mutants by production of the known receptor agonist Apelin. Additionally, a fusion protein of alkaline phosphatase and Toddler binds to cells expressing Apelin receptors (*51*). Previous studies have implicated APJ/Apelin receptor signaling in a variety of biological processes, including the regulation of cardiovascular development and physiology, the control of fluid homeostasis, or even as a co-receptor for HIV infection (*53*, *54*). Although Apelin has previously been the only known agonist of the APJ/Apelin receptor, genetic studies have found discrepancies between the roles of Apelin and its receptor in mouse (*34*, *36*, *41*, *45*, *55*) and zebrafish (*24*–*26*). For example, Apelin knockout mice are viable and fertile (*45*, *46*, *56*), whereas APJ/Apelin receptor mutant mice are born at sub-Mendelian ratios (*34*). Our studies suggest that both Toddler and Apelin can activate APJ/Apelin receptors and indicate that it is endogenous Toddler—not Apelin—that activates APJ/Apelin receptor signaling during zebrafish gastrulation. Analogously to the promise of Apelin in biomedical applications (*53*, *54*), Toddler and its derivatives may take the place of Apelin in therapeutic contexts. Indeed, Toddler may also activate mammalian APJ/Apelin receptors because misexpression of zebrafish, mouse, and human Toddler induces similar overexpression phenotypes in zebrafish (fig. S19).

Fourth, our RNA-Seq and ribosome profiling data indicate that Toddler might just be one of several poorly characterized developmental signals that may have been missed in mutagenesis screens because of their small size. Applying similar genomic approaches to adult tissues might identify additional previously unknown signals that regulate physiological and behavioral processes.

References and Notes

1. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell* (Garland Science, New York, ed. 5, 2007).
2. C. P. Heisenberg, Y. Bellaïche, Forces in tissue morphogenesis and patterning. *Cell* **153**, 948–962 (2013). doi: 10.1016/j.cell.2013.05.008; pmid: 23706734
3. L. Solnica-Krezel, D. S. Sepich, Gastrulation: Making and shaping germ layers. *Annu. Rev. Cell Dev. Biol.* **28**, 687–717 (2012). doi: 10.1146/annurev-cellbio-092910-154043; pmid: 22804578
4. J. B. Wallingford, Planar cell polarity and the developmental control of cell behavior in vertebrate embryos. *Annu. Rev. Cell Dev. Biol.* **28**, 627–653 (2012). doi: 10.1146/annurev-cellbio-092910-154208; pmid: 22905955
5. S. Nowotschin, A. K. Hadjantonakis, Cellular dynamics in the early mouse embryo: From axis formation to gastrulation. *Curr. Opin. Genet. Dev.* **20**, 420–427 (2010). doi: 10.1016/j.gde.2010.05.008; pmid: 20566281
6. N. T. Ingolia, L. F. Lareau, J. S. Weissman, Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011). doi: 10.1016/j.cell.2011.10.002; pmid: 22056041
7. G. L. Chew *et al.*, Ribosome profiling reveals resemblance between long non-coding RNAs and 5′ leaders of coding RNAs. *Development* **140**, 2828–2834 (2013). doi: 10.1242/dev.098343; pmid: 23698349
8. A. Pauli *et al.*, Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* **22**, 577–591 (2012). doi: 10.1101/gr.133009.111; pmid: 22110045
9. I. Ulitsky, A. Shkumatava, C. H. Jan, H. Sive, D. P. Bartel, Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550 (2011). doi: 10.1016/j.cell.2011.11.055; pmid: 22196729
10. A. S. Hassan, J. Hou, W. Wei, P. A. Hoodless, Expression of two novel transcripts in the mouse definitive endoderm. *Gene Expr. Patterns* **10**, 127–134 (2010). doi: 10.1016/j.gep.2010.02.001; pmid: 20153842
11. M. Guttman *et al.*, lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300 (2011). doi: 10.1038/nature10398; pmid: 21874018
12. D. Reyon *et al.*, FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* **30**, 460–465 (2012). doi: 10.1038/nbt.2170; pmid: 22484455
13. J. D. Sander *et al.*, Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.* **29**, 697–698 (2011). doi: 10.1038/nbt.1934; pmid: 21822241
14. T. Mizoguchi, H. Verkade, J. K. Heath, A. Kuroiwa, Y. Kikuchi, Sdf1/Cxcr4 signaling controls the dorsal migration of endodermal cells during zebrafish gastrulation. *Development* **135**, 2521–2529 (2008). doi: 10.1242/dev.020107; pmid: 18579679

15. G. Pézeron *et al.*, Live analysis of endodermal layer formation identifies random walk as a novel gastrulation movement. *Curr. Biol.* **18**, 276–281 (2008). doi: 10.1016/j.cub.2008.01.028; pmid: 18291651
16. M. Doitsidou *et al.*, Guidance of primordial germ cell migration by the chemokine SDF-1. *Cell* **111**, 647–659 (2002). doi: 10.1016/S0092-8674(02)01135-2; pmid: 12464177
17. M. Haskel-Ittah *et al.*, Self-organized shuttling: Generating sharp dorsoventral polarity in the early *Drosophila* embryo. *Cell* **150**, 1016–1028 (2012). doi: 10.1016/j.cell.2012.06.044; pmid: 22939625
18. G. Venkiteswaran *et al.*, Generation and dynamics of an endogenous, self-generated signaling gradient across a migrating tissue. *Cell* **155**, 674–687 (2013). doi: 10.1016/j.cell.2013.09.046; pmid: 24119842
19. E. Donà *et al.*, Directional tissue migration through a self-generated chemokine gradient. *Nature* **503**, 285–289 (2013). doi: 10.1038/nature12635; pmid: 24067609
20. E. M. Powell, W. M. Mars, P. Levitt, Hepatocyte growth factor/scatter factor is a motogen for interneurons migrating from the ventral to dorsal telencephalon. *Neuron* **30**, 79–89 (2001). doi: 10.1016/S0896-6273(01)00264-1; pmid: 11343646
21. P. Giacobini *et al.*, Hepatocyte growth factor acts as a motogen and guidance signal for gonadotropin hormone-releasing hormone-1 neuronal migration. *J. Neurosci.* **27**, 431–445 (2007). doi: 10.1523/JNEUROSCI.4979-06.2007; pmid: 17215404
22. G. Kirfel, B. Borm, A. Rigort, V. Herzog, The secretory β-amyloid precursor protein is a motogen for human epidermal keratinocytes. *Eur. J. Cell Biol.* **81**, 664–676 (2002). doi: 10.1078/0171-9335-00284; pmid: 12553667
23. S. Nair, T. F. Schilling, Chemokine signaling controls endodermal migration during zebrafish gastrulation. *Science* **322**, 89–92 (2008). doi: 10.1126/science.1160038; pmid: 18719251
24. X. X. I. Zeng, T. P. Wilm, D. S. Sepich, L. Solnica-Krezel, Apelin and its receptor control heart field formation during zebrafish gastrulation. *Dev. Cell* **12**, 391–402 (2007). doi: 10.1016/j.devcel.2007.01.011; pmid: 17336905
25. I. C. Scott *et al.*, The G protein-coupled receptor Agtrl1b regulates early development of myocardial progenitors. *Dev. Cell* **12**, 403–413 (2007). doi: 10.1016/j.devcel.2007.01.012; pmid: 17336906
26. S. Paskaradevan, I. C. Scott, The Aplnr GPCR regulates myocardial progenitor development via a novel cell-non-autonomous, $G\alpha_{i/o}$ protein-independent pathway. *Biol. Open* **1**, 275–285 (2012). doi: 10.1242/bio.2012380; pmid: 23213418
27. X. Li *et al.*, Gpr125 modulates Dishevelled distribution and planar cell polarity signaling. *Development* **140**, 3028–3039 (2013). doi: 10.1242/dev.094839; pmid: 23821037
28. F. Lin *et al.*, $G\alpha_{12/13}$ regulate epiboly by inhibiting E-cadherin activity and modulating the actin cytoskeleton. *J. Cell Biol.* **184**, 909–921 (2009). doi: 10.1083/jcb.200805148; pmid: 19307601
29. M. Costa, E. T. Wilson, E. Wieschaus, A putative cell signal encoded by the *folded gastrulation* gene coordinates cell shape changes during *Drosophila* gastrulation. *Cell* **76**, 1075–1089 (1994). doi: 10.1016/0092-8674(94)90384-0; pmid: 8137424
30. S. Parks, E. Wieschaus, The *Drosophila* gastrulation gene *concertina* encodes a Gα-like protein. *Cell* **64**, 447–458 (1991). doi: 10.1016/0092-8674(91)90652-F; pmid: 1899050
31. C. D'Aniello *et al.*, G protein-coupled receptor APJ and its ligand apelin act downstream of Cripto to specify embryonic stem cells toward the cardiac lineage through extracellular signal-regulated kinase/p70S6 kinase signaling pathway. *Circulation* **105**, 231–238 (2009). doi: 10.1161/CIRCRESAHA.109.201186; pmid: 19574549
32. I. N. Wang *et al.*, Apelin enhances directed cardiac differentiation of mouse and human embryonic stem cells. *PLOS One* **7**, e38328 (2012). doi: 10.1371/journal.pone.0038328; pmid: 22675543
33. D. Tempel *et al.*, Apelin enhances cardiac neovascularization after myocardial infarction by recruiting Aplnr+ circulating cells. *Circ. Res.* **111**, 585–598 (2012). doi: 10.1161/CIRCRESAHA.111.262097; pmid: 22753078
34. Y. Kang *et al.*, Apelin-APJ signaling is a critical regulator of endothelial MEF2 activation in cardiovascular development. *Circ. Res.* **113**, 22–31 (2013). doi: 10.1161/CIRCRESAHA.113.301324; pmid: 23603510
35. M. Inui, A. Fukui, Y. Ito, M. Asashima, Xapelin and Xmsr are required for cardiovascular development in *Xenopus laevis*. *Dev. Biol.* **298**, 188–200 (2006). doi: 10.1016/j.ydbio.2006.06.028; pmid: 16876154
36. D. N. Charo *et al.*, Endogenous regulation of cardiovascular function by apelin-APJ. *Am. J. Physiol. Heart Circ. Physiol.* **297**, H1904–H1913 (2009). doi: 10.1152/ajpheart.00686.2009; pmid: 19767528
37. K. Tatemoto *et al.*, Isolation and characterization of a novel endogenous peptide ligand for the human APJ receptor. *Biochem. Biophys. Res. Commun.* **251**, 471–476 (1998). doi: 10.1006/bbrc.1998.9489; pmid: 9792798
38. D. K. Lee *et al.*, Characterization of apelin, the ligand for the APJ receptor. *J. Neurochem.* **74**, 34–41 (2000). doi: 10.1046/j.1471-4159.2000.0740034.x; pmid: 10617103
39. B. Tucker *et al.*, Zebrafish *Angiotensin II Receptor-like 1a* (*agtrl1a*) is expressed in migrating hypoblast, vasculature, and in multiple embryonic epithelia. *Gene Expr. Patterns* **7**, 258–265 (2007). doi: 10.1016/j.modgep.2006.09.006; pmid: 17085078
40. S. Nornes, B. Tucker, M. Lardelli, Zebrafish *aplnra* functions in epiboly. *BMC Res. Notes* **2**, 231 (2009). doi: 10.1186/1756-0500-2-231; pmid: 19922670
41. M. C. Scimia *et al.*, APJ acts as a dual receptor in cardiac hypertrophy. *Nature* **488**, 394–398 (2012). doi: 10.1038/nature11263; pmid: 22810587
42. J. Ishida *et al.*, Regulatory roles for APJ, a seven-transmembrane receptor related to angiotensin-type 1 receptor in blood pressure in vivo. *J. Biol. Chem.* **279**, 26274–26279 (2004). doi: 10.1074/jbc.M404149200; pmid: 15087458
43. E. M. Roberts *et al.*, Abnormal fluid homeostasis in apelin receptor knockout mice. *J. Endocrinol.* **202**, 453–462 (2009). doi: 10.1677/JOE-09-0134; pmid: 19578099
44. H. Kidoya *et al.*, Spatial and temporal role of the apelin/APJ system in the caliber size regulation of blood vessels during angiogenesis. *EMBO J.* **27**, 522–534 (2008). doi: 10.1038/sj.emboj.7601982; pmid: 18200044
45. K. Kuba *et al.*, Impaired heart contractility in Apelin gene–deficient mice associated with aging and pressure overload. *Circ. Res.* **101**, e32–e42 (2007). doi: 10.1161/CIRCRESAHA.107.158659; pmid: 17673668
46. A. Y. Sheikh *et al.*, In vivo genetic profiling and cellular localization of apelin reveals a hypoxia-sensitive, endothelial-centered pathway activated in ischemic heart failure. *Am. J. Physiol. Heart Circ. Physiol.* **294**, H88–H98 (2008). doi: 10.1152/ajpheart.00935.2007; pmid: 17906101
47. N. A. Evans *et al.*, Visualizing differences in ligand-induced β-arrestin–GFP interactions and trafficking between three recently characterized G protein-coupled receptors. *J. Neurochem.* **77**, 476–485 (2001). doi: 10.1046/j.1471-4159.2001.00269.x; pmid: 11299310
48. D. K. Lee, S. S. G. Ferguson, S. R. George, B. F. O'Dowd, The fate of the internalized apelin receptor is determined by different isoforms of apelin mediating differential interaction with β-arrestin. *Biochem. Biophys. Res. Commun.* **395**, 185–189 (2010). doi: 10.1016/j.bbrc.2010.03.151; pmid: 20353754
49. A. Reaux *et al.*, Physiological role of a novel neuropeptide, apelin, and its receptor in the rat brain. *J. Neurochem.* **77**, 1085–1096 (2001). doi: 10.1046/j.1471-4159.2001.00320.x; pmid: 11359874
50. N. Zhou *et al.*, Cell–cell fusion and internalization of the CNS-based, HIV-1 co-receptor, APJ. *Virology* **307**, 22–36 (2003). doi: 10.1016/S0042-6822(02)00021-1; pmid: 12667811
51. S. C. Chng, L. Ho, J. Tian, B. Reversade, ELABELA: A hormone essential for heart development signals via the apelin receptor. *Dev. Cell* **27**, 672–680 (2013). doi: 10.1016/j.devcel.2013.11.002; pmid: 24316148
52. A. F. Schier, W. S. Talbot, Molecular genetics of axis formation in zebrafish. *Annu. Rev. Genet.* **39**, 561–613 (2005). doi: 10.1146/annurev.genet.37.110801.143752; pmid: 16285872
53. G. Barnes, A. G. Japp, D. E. Newby, Translational promise of the apelin–APJ system. *Heart* **96**, 1011–1016 (2010). doi: 10.1136/hrt.2009.191122; pmid: 20584856
54. M. J. Kleinz, A. P. Davenport, Emerging roles of apelin in biology and medicine. *Pharmacol. Ther.* **107**, 198–211 (2005). doi: 10.1016/j.pharmthera.2005.04.001; pmid: 15907343
55. A. Kasai *et al.*, Retardation of retinal vascular development in apelin-deficient mice. *Arterioscler. Thromb. Vasc. Biol.* **28**, 1717–1722 (2008). doi: 10.1161/ATVBAHA.108.163402; pmid: 18599802
56. H. Kidoya, H. Naito, N. Takakura, Apelin induces enlarged and nonleaky blood vessels for functional recovery from ischemia. *Blood* **115**, 3166–3174 (2010). doi: 10.1182/blood-2009-07-232306; pmid: 20185589

Acknowledgments: We thank D. Richardson and C. Kraft from the Harvard Center for Biological Imaging for technical support; F. Merkle for providing human and mouse embryonic stem cell RNA; M. Lin for the initial PhyloCSF analysis; L. Solnica-Krezel, E. Raz, C. Houart, members of the 2013 MBL Zebrafish Course, and the Schier laboratory for helpful discussions; and S. Mango, W. Talbot, R. Losick, J. Farrell, and K. Rogers for comments on the manuscript. Obtaining the TALEN plasmids will require the completion of a Uniform Biological Material Transfer Agreement with the Massachusetts General Hospital. The Massachusetts General Hospital has applied for a patent that covers the FLASH method used to make the TALENs and J.K.J. is an inventor on this patent. J.K.J. has financial interests in Editas Medicine and Transposagen Biopharmaceuticals. J.K.J. is a member of the Scientific Advisory Board of Transposagen Biopharmaceuticals and is a co-founder and paid consultant of Editas Medicine and holds equity in both companies. J.K.J.'s interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. This research was supported by NIH (A.F.S., A.P., and A.S.), Human Frontier Science Program (A.P. and E.V.), Howard Hughes Medical Institute (G.-L.C.), and the American Cancer Society (J.A.G.). **Author contributions:** A.P. and A.F.S. conceived the study and wrote the paper. A.P. collected and analyzed the data, with contributions from A.F.S., M.L.N. (phenotypic characterization), E.V. (computational analyses), G.-L.C. (ribosome profiling), J.A.G., S.Z., D.R., S.Q.T., J.K.J. (TALEN-mediated mutagenesis), A.M., J.M., A.S. (mass spectrometry), and J.D. (MATLAB cell tracking).

Supplementary Materials

www.sciencemag.org/content/343/6172/1248636/suppl/DC1
Materials and Methods
Figs. S1 to S19
References (*57*–*67*)
Table S1
Data Files S1 and S2

18 November 2013; accepted 26 December 2013
10.1126/science.1248636

A Genetic Atlas of Human Admixture History

Garrett Hellenthal,[1] George B. J. Busby,[2] Gavin Band,[3] James F. Wilson,[4] Cristian Capelli,[2] Daniel Falush,[5]* Simon Myers[3,6]*†

Modern genetic data combined with appropriate statistical methods have the potential to contribute substantially to our understanding of human history. We have developed an approach that exploits the genomic structure of admixed populations to date and characterize historical mixture events at fine scales. We used this to produce an atlas of worldwide human admixture history, constructed by using genetic data alone and encompassing over 100 events occurring over the past 4000 years. We identified events whose dates and participants suggest they describe genetic impacts of the Mongol empire, Arab slave trade, Bantu expansion, first millennium CE migrations in Eastern Europe, and European colonialism, as well as unrecorded events, revealing admixture to be an almost universal force shaping human populations.

Diverse historical, archaeological, anthropological, and linguistic sources of information indicate that human populations have interacted throughout history, because of the rise and fall of empires, invasions, migrations, slavery, and trade. These interactions can result in sudden or gradual transfers of genetic material, creating admixed populations. However, the genetic legacy of these interactions remains unknown in most cases, and the historical record is incomplete. We have developed an approach that provides a detailed characterization of the mixture events in the ancestry of sampled populations on the basis of genetic data alone.

Admixed populations should have segments of DNA from all contributing source groups (Fig. 1A), whose sizes decrease over successive generations because of recombination, and approaches have been developed to date admixture events by inferring the size of ancestry segments (*1*–*5*). Between-population frequency differences of individual alleles may provide information on ancestry sources (*6*, *7*). On the basis of these principles, we developed an integrated approach by using genome-wide patterns of ancestry to infer jointly both fine-scale information about groups involved in admixture and its timing, allowing for the fact that migration and admixture events can occur at multiple times or involve numerous groups.

The GLOBETROTTER Method

Our approach gains power and resolution by using alleles at multiple successive single-nucleotide

[1]UCL Genetics Institute, University College London, Gower Street, London WC1E 6BT, UK. [2]Department of Zoology, Oxford University, South Parks Road, Oxford OX1 3PS, UK. [3]Wellcome Trust Centre for Human Genetics, Oxford University, Roosevelt Drive, Oxford OX3 7BN, UK. [4]Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK. [5]Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. [6]Department of Statistics, Oxford University, 1 South Parks Road, Oxford OX1 3TG, UK.

*These authors contributed equally to this work.
†Corresponding author. E-mail: myers@stats.ox.ac.uk

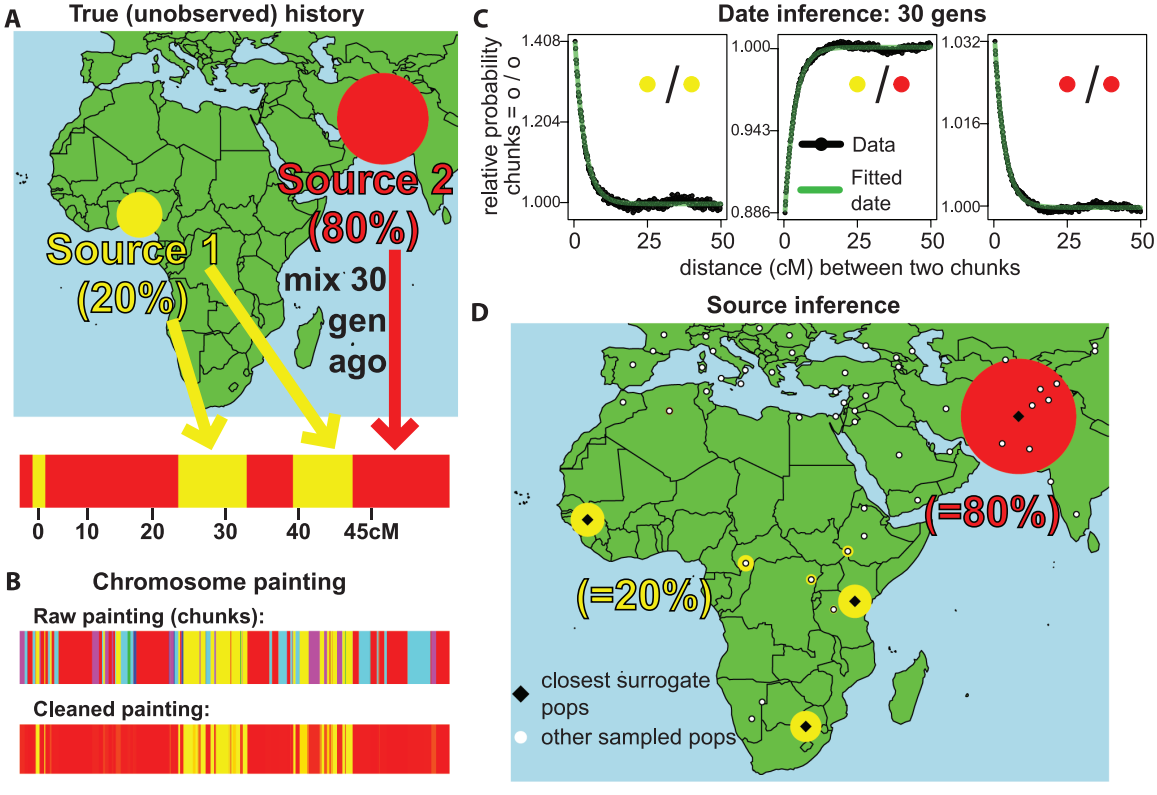


Fig. 1. Ancestry painting and admixture analysis of simulated admixture. (**A**) A simulated event 30 generations ago between Brahui (80%, red) and Yoruba (20%, yellow) resulted in admixed individuals having DNA segments from each source (bottom). The true sources are then treated as unsampled. cM, centimorgan. (**B**) CHROMOPAINTER's painting of the same region (yellow, Africa; green, America; red, Central-South Asia; blue, East Asia; cyan, Europe; pink, Near East; black, Oceania), showing haplotypic segments ("chunks") shared with these groups. Our model fitting narrows the donor set largely to Central-South Asia and Africa, generating a "cleaned" painting. (**C**) Coancestry curves (black line) show relative probability of jointly copying two chunks from red (Balochi; F_{ST} = 0.003 with Brahui) and/or yellow (Mandenka; F_{ST} = 0.009 with Yoruba) donors, at varying genetic distances. The curves closely fit an exponential decay (green line) with a rate of 30 generations (95% CI: 27 to 33). The positive slope for the Balochi-Mandenka curve (middle) implies that these donors represent different admixing sources. (**D**) GLOBETROTTER's source inference, with black diamonds indicating sampled populations with greatest similarity ($F_{ST} \leq 0.001$ over minimum) to true sources, white circles other sampled populations. Red and yellow circles, with areas summing to 20% and 80%, respectively, show inferred haplotypic makeup of the two admixing sources.

polymorphisms (SNPs) (haplotypes) (*8*). Given a focal population within a larger data set containing many such groups, the chromosomes of individuals in this population share ancestors with those in other populations, resulting in shared "chunks" of DNA. We used CHROMOPAINTER (*8*) to decompose each chromosome as a series of haplotypic chunks, each inferred to be shared with an individual from one of the other groups and colored (or painted) by this group (Fig. 1B). If the focal population is admixed, the changing colors along a chromosome noisily reflect true but unknown underlying ancestry (Fig. 1B) and so can be used to learn details of the source group(s) involved. To do this, we modeled haplotypes within each unsampled source group as being found across a weighted mixture of sampled "donor" populations (*9*). If a source group is genetically relatively similar to a single sampled population, then this population will dominate the inferred mixture. If there is no close proxy for the admixing group in the sample, especially likely for ancient admixture events or sparsely sampled regions, several donor populations will be needed to approximate its pattern of haplotype sharing. The focal population is then automatically a haplotypic mixture of the combined donors, because it is a mixture of the source groups. Inferring the reduced set of groups within the mixture allows us to produce a "cleaned" painting (Fig. 1B) using only these groups.

To assess the evidence for admixture and date events, informally we measured the scale at which the cleaned painting changes along the genome. Specifically, we produced a coancestry curve for each pair of donor populations, plotting genetic distance x against a measure of how often a pair of haplotype chunks separated by x come from each respective donor (Fig. 1C), analogously to ROLLOFF curves (*4*), and averaging over uncertain and typically computationally estimated haplotypic phase (*9*). In theory, given a single admixture event, ancestry chunks inherited from each source have an exponential size distribution, resulting in an exponential decay of these coancestry curves (*9*). The rate of decay in all curves will be equal to the time in generations since admixture (Fig. 1C) (*4*, *9*, *10*), allowing estimation of this date: Steeper decay corresponds to older admixture. Such a decay distinguishes true admixture from ancient spatial structure and should only occur in recipient but not donor groups involved in nonreciprocal admixture events. We test for evidence to reject ($P < 0.01$) a no-admixture

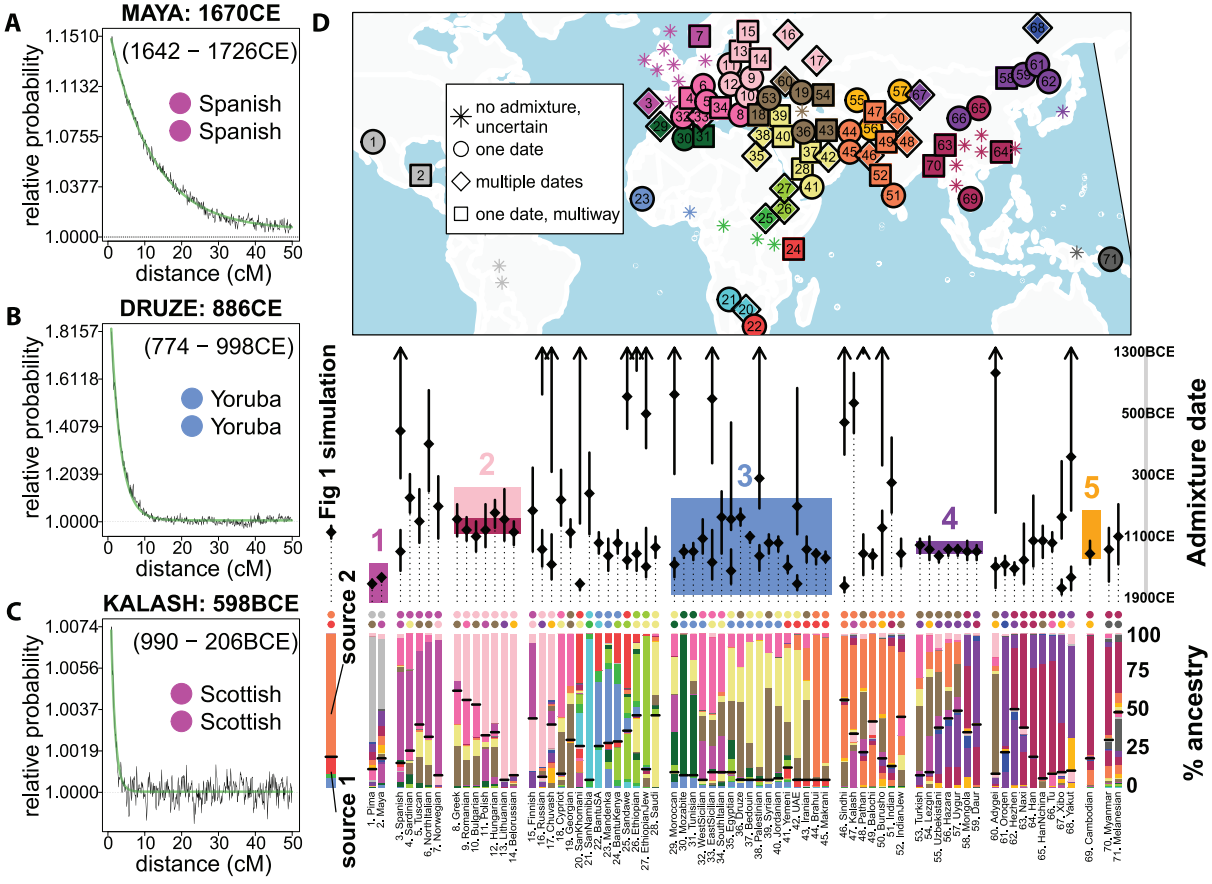


Fig. 2. Overview of inferred admixture for 95 human populations. (**A**) Coancestry curve for the Maya for Spanish donor group (inferred as closest to minor admixing source), with green fitted line showing inferred exponential decay curve and a corresponding recent admixture date (with 95% CIs). (**B** and **C**) As (A), but showing the Druze and Kalash, respectively, with different indicated donors (donors indicated are proxies for minor admixing source, inferred as closest to Yoruba and Germany/Austria, respectively) and with successively older admixture. (**D**) On the map (locations approximate in densely sampled regions), shapes (see legend) indicate inference: no admixture, a single admixture event, or more complex admixture. Colors indicate fineSTRUCTURE clustering into 18 clades (table S11 and figs. S12 and S13). Inferred date(s) and 95% CIs are directly below the map, with two inferred admixing sources (dots and vertical bars) shown below each date (see example for simulation of Fig. 1 at left). For multiple admixture times, these two sources correspond to the more recent event; for multiple groups, they reflect the strongest admixture "direction." Colored dots above each bar indicate clades best representing the major (top) and minor (bottom) sources. The bar is split at the inferred admixture fraction (horizontal line, fractions <5% shown as 5%). Each bar section indicates the inferred donor group haplotypic makeup, colored as the map, for one source. Shaded boxes on the inferred admixture times denote events referred to in the text, specifically (label 1) European colonization of the Americas (1492 CE to present, fuchsia); (2) Slavic (500 to 900 CE, pink) and Turkic (500 to 1100 CE, maroon) migrations; (3) Arab slave trade (650 to 1900 CE, cyan); (4) Mongol empire (1206 to 1368 CE, purple); and (5) Khmer empire (802 to 1431 CE, orange).

null model, that is, no exponential decay in (normalized) coancestry curves, via bootstrapping (*9*). Multiple admixture times result in a mixture of exponentials (*9*); if admixture is detected, we test for evidence of multiple admixture times (e.g., two episodes of admixture or more continuous admixture over a longer period; empirical $P < 0.05$ in simulations), comparing the fit of a single exponential decay rate versus a mixture of rates.

The curve heights (intercepts) provide complementary information to deconvolve the number and genetic composition of the ancestral sources before admixture (*11*). Fitted curves for all pairs of donor groups (Fig. 1C shows three examples) specify a pairwise intercept matrix, which, after normalization, we decompose by using a series of eigenvectors. Analogous to the standard use of eigenvector decomposition in principal components analysis (PCA) in genetics to estimate relative ancestry source contributions

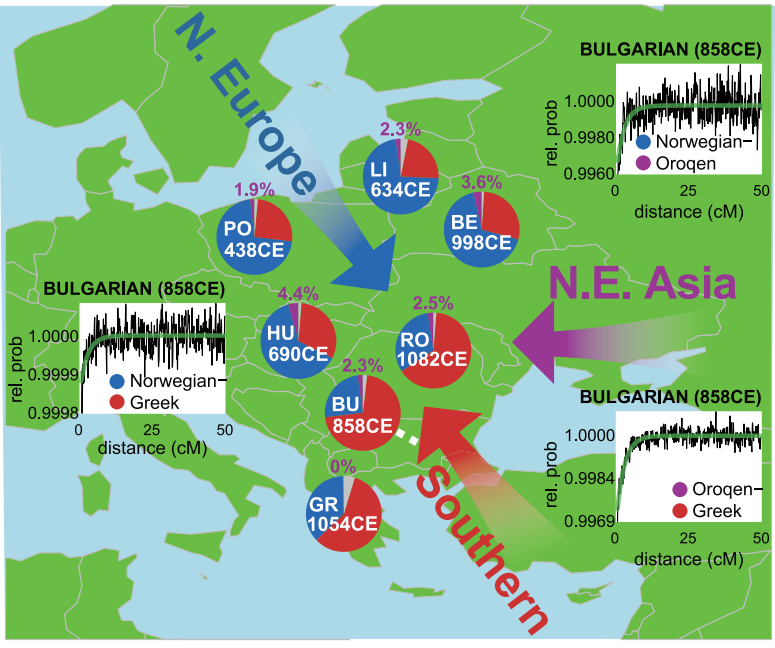


Fig. 3. Multiway admixture in Eastern Europe. Mixing percentages (pie graphs) and dates (white text) inferred by using the strongest admixture "direction" for six eastern European groups—Belarus (BE), Bulgaria (BU), Hungary (HU), Lithuania (LI), Poland (PO), Romania (RO), analyzed when disallowing copying from nearby groups—and Greece (GR), analyzed by using the full set of 94 donors. Mixing percentages indicate percentages for three geographic regions: "N. Europe" (Northwest Europe and East Europe from clades of table S11; blue), "Southern" (South Europe and West Asia; red), and "N.E. Asia" (Northeast Asia and Yakut; purple, also given above each pie), plus other (gray). All groups except Greece show evidence ($P < 0.05$) of multiway admixture involving sources along the approximate directions show by the arrows. Coancestry curves (black lines) for Bulgaria, fitted with an exponential decay curve (green lines), exemplify this multiway signal. Each pairing of the three donor groups, each a proxy for the admixture source from a different region (Norway, northeast Europe; Oroqen, Northeast Asia; and Greece, South Europe and West Asia), exhibits negative correlation (a dip) in ancestry weights at short genetic distances, implying at least three identifiably distinct ancestral sources mixing (approximately) simultaneously (*9*).

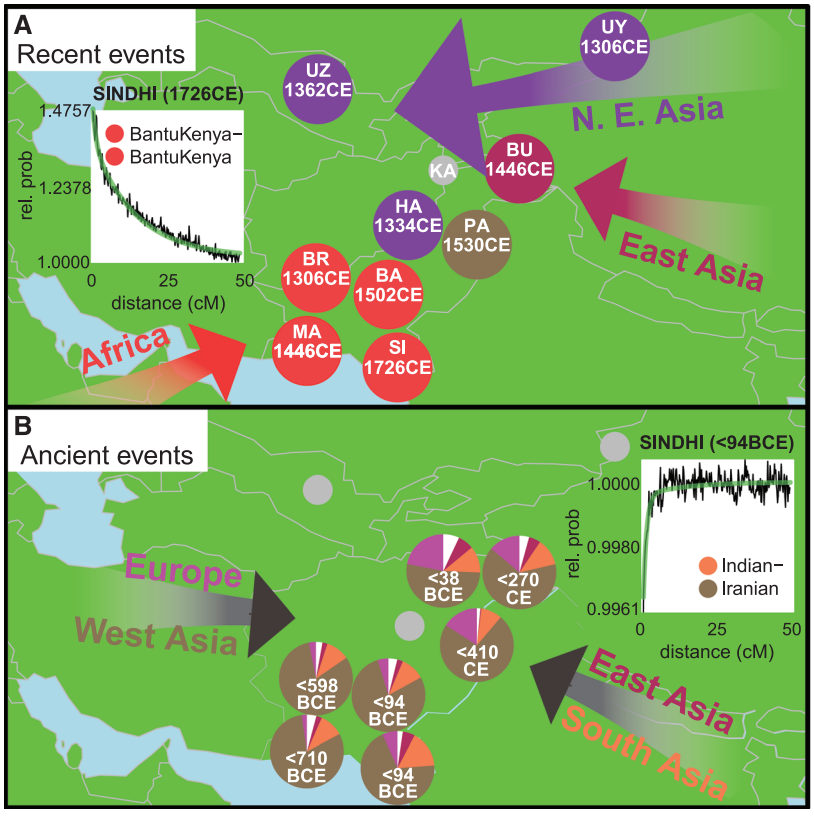


Fig. 4. Ancient and modern admixture in Central Asia. (**A**) Dates (white text) and minority contributing sources for recent inferred events in nine populations (circles), analyzed disallowing copying from nearby groups, show contributions from Northeast Asia (purple) in the Hazara (HA), Uygur (UY), and Uzbekistani (UZ); East Asia (maroon) in Burusho (BU); West Asia (brown) in Pathan (PA); and Africa (red) in Balochi (BA), Brahui (BR), Makrani (MA), and Sindhi (SI). Kalash (KA, gray) have no inferred recent event. (**B**) Inferred mixing percentages (pie graphs) and dates (white text gives upper CI bound) for additional, possibly shared, ancient events in seven groups (HA, UY, and UZ have no inferred ancient events). Pie graphs show inferred donor makeup of each group after removing the recent event contribution from (A), if any, with colors referring to donors from "East Asia" (Southeast Asia from clades of table S11; maroon), "Europe" (Northwest, East, and South Europe, fuchsia), "Central South Asia" (orange), "West Asia" (brown), and other (white). Arrows indicate "directions" of ancient admixture, with donor regions splitting into two pairs that represent different sources. Coancestry curves (black lines) for Sindhi are superimposed for two different donor pairs representing proxies for admixing groups with ancestry indicated by the solid circles, indicating highly different exponential decay rates fit as a mixture of 7 and 94 generations (green lines).

for different individuals (*12*), the eigenvectors allow us to estimate the relative contribution to different admixing sources (e.g., source 1 versus source 2) for each different donor group (*9*). Also as for PCA, admixture between K distinct source populations produces $K - 1$ significant eigenvectors (*13*), and we test for three or more admixing sources by testing (empirically) for evidence of two or more such eigenvectors ($P < 0.05$) (*9*). After iterative modeling to improve results, this allows us to attempt to "reverse" the admixture process (Fig. 1D) and to infer the haplotypic makeup of admixing source groups as well as admixture date(s) in our method, which we call GLOBETROTTER.

Simulations

To test our approach under diverse single, complex, and no-admixture scenarios, incorporating many of the complexities (such as unsampled or admixed donor groups) likely to be present in real data, we simulated admixture scenarios involving real (but hidden to our analysis) human populations (*4*, *9*) and populations generated under a coalescent framework (*14*) incorporating inferred (*15*–*18*) past demographic events. Admixture was simulated between 7 and 160 generations [200 to 4400 years, assuming 28 years per human generation (*19*)] ago, with admixture fractions 3 to 50% and genetic differentiation (F_{ST}) between the admixing groups varying from 0.018 (similar to Europe versus Central Asia) to 0.185 (similar to West Africa versus Europe). Results are detailed online (figs. S3 to S7 and tables S1 and S5). All populations simulated without admixture, including those with long-term migration, showed no admixture evidence ($P > 0.1$). Power to detect admixture ($P < 0.01$) when present was 94%, and 95% of our 95% bootstrapped confidence intervals (CIs) contained the true admixture date, including cases with two distinct incidents of admixture or multiple groups admixing simultaneously. Inferred source accuracy was very high (*9*), with, for example, the mixture representation predicting a haplotype composition more correlated to the true, typically unsampled, source population than to any single sampled population >80% of the time. However, source accuracy was lower for admixing sources contributing only 5% of DNA, with around 40% of such scenarios yielding elevated (>25%) rates of falsely inferring multiple admixture times and/or admixing groups. Further testing demonstrated robustness of GLOBETROTTER, in simulations and real data, to haplotypic phase inference approach used, inclusion/exclusion of particular chromosomes, genetic map chosen to provide genetic distances, and the presence of population bottlenecks since admixture, whereas GLOBETROTTER admixture dating was improved relative to ROLLOFF (*4*, *9*).

Nevertheless, there are multiple settings that we believe are challenging for our approach. First, although the admixing sources need not be sampled—often impossible because of genetic drift, extinction, or later admixture into the sources themselves—source inference is improved when more similar extant groups are sampled, and GLOBETROTTER may miss events where we lack any extant group that can separate sources. Second, sampling of several genetically very similar groups can mask admixture events they share. Similarly, a caveat is that where genuine, recent bidirectional gene flow has occurred, admixture fractions are difficult to define and interpret. However, date estimation is predicted to still be useful, and in real data the majority of our inferred events do not appear to be bidirectional in this manner. Third, even in theory our approach finds it challenging to distinguish distinct continuous "pulses" of admixture and continuous migration over some time frame (*9*), because of the difficulty of separating exponential mixtures (*20*). If the time frame were narrow, we expect to infer a single admixture time within the range of migration dates. Where we infer two admixture dates, in particular with the same source groups, the exponential decay signal could also be consistent with more continuous migration, and so we conservatively refer to this as admixture at multiple dates. Last, we only attempt to analyze populations with signals consistent with at most three groups admixing and infer at most two admixture times, and we can provide only less precise inference of sources for the weaker or older admixture signal in these complex cases (*9*).

Analysis of Worldwide Admixture

By using GLOBETROTTER, we analyzed 1490 individuals from 95 worldwide human groups (table S10 and fig. S12) (*9*), composed of 17 newly genotyped groups (*21*), 53 from the Human Genome Diversity Panel (HGDP) (*22*), and 25 from other sources (*23*, *24*), filtered to 474,491 autosomal SNPs. We phased the individuals by using IMPUTE2 (*9*, *25*) and used fineSTRUCTURE (*8*) to verify homogeneity within labeled populations, to identify genetically similar and clustered groups, and to remove outlying individuals (figs. S12 to S14 and tables S10 and S11). Of the 95 populations, 80 showed evidence ($P < 0.01$) of admixture, although nine could not be characterized by our approach (table S12). More than half of these have evidence of multiple waves of admixture ($P < 0.05$), and estimated admixture times vary from <10 to >150 generations (Fig. 2). We present individual results, for each population, via an interactive map online (*26*). We tested consistency of our results against a previous analysis of the 53 groups within the HGDP (*11*), which identified 34 groups with evidence of recent admixture. We identified ($P < 0.01$) admixture evidence in all 34 cases (with multiple event evidence in 15 cases) and obtained 95% admixture date CIs narrower than, but consistent with, those estimated by using ROLLOFF (*9*, *11*). For 10 of 19 HGDP groups lacking previous support for recent admixture, GLOBETROTTER also identifies no events: In the remaining populations, admixture is inferred as occurring between genetically similar sources ($F_{ST} < 0.02$), a challenging setting where simulations suggest our method is more powerful (*9*).

In several instances, GLOBETROTTER clarifies or extends previous genetic analyses. For example, a previous study (*27*) inferred admixture in the Maya, with best source populations the Mozabites from North Africa and the Native American Surui, speculating on the basis of historical events that this might actually represent a mixture of European, West African, and Native American ancestry sources. GLOBETROTTER inferred admixture between three groups in the Maya dating to around 1670 CE (9 generations ago) (*28*) (Fig. 2, A and D, fuchsia box 1), with distinct sources from Europe (most genetically similar to the Spanish), West Africa (the Yoruba), and the Americas (the Pima, the nearest sampled group in the Americas). A different method, which aims to detect but not date admixture, concluded that Cambodians trace ~16% of their DNA to a group equally related to modern-day Europeans and East Asians (*29*). GLOBETROTTER infers a ~19% contribution from a similar source related to modern-day Central, South, and East Asians and an ~81% contribution from a source related specifically to modern-day Han and Dai, the latter a branch of the Tai people who entered the region in historical times (*30*) (Fig. 2D, orange box 5). Further, this event dates to 1362 CE (1194 to 1502 CE), a period spanning the end of the Indianized Khmer empire (802 to 1431 CE) (*30*), one of the most powerful empires in Southeast Asia, whose fall was hypothesized to relate to a Tai influx (*30*).

A comparison with the historical record becomes progressively more difficult for older episodes. Even when events are well attested, their exact genetic impacts (if any) are rarely if ever known, motivating our approach. Nevertheless, we have identified nine groups of populations showing related events, incorporating almost all (19/20) with the strongest GLOBETROTTER admixture evidence (*9*). Results are presented as online maps (*26*). Some events appear to match well with particular historical occurrences, such as the so-called Bantu Expansion into Southern Africa (*9*). Events affecting a group of seven populations (Fig. 2D, purple box 4) correspond in time to the rapid expansion, initiated by Genghis Khan, of the Mongol empire (1206 to 1368 CE) (*31*), one of the most dramatic events in human history. These populations, including the Hazara (*32*, *33*), the Uygur (*34*), and the Mongola themselves, were sampled from within the range of the Mongol empire and show an admixture event dating within the Mongol Period, with one source closely genetically related to the Mongola that progressively decreases in proportion westward, to 8% in the Turkish (Fig. 2D).

Seventeen populations from the Mediterranean, the Near East, and countries bordering the Arabian Sea (Fig. 2D, blue box 3) show signals of admixture from sub-Saharan Africa, with most recent dates in the range 890 to 1754 CE (Fig. 2, B

and D). We interpret these signals, consistent with overlapping results of previous studies (*4*, *20*), as resulting from the Arab expansion and slave trade, which originated around the seventh century CE (*35*). Our event dates are highly consistent with this but also imply earlier sub-Saharan African gene flow into, for example, the Moroccans. The highest-contributing sub-Saharan donor is West African for all 12 Mediterranean populations and an East or South African Bantu-speaking group for all five Arabian Sea populations (Fig. 2D), confirming genetically different sources for these slave trades (*35*).

A population group centered around Eastern Europe shows signals of complex admixture. FineSTRUCTURE did not fully separate groups from this region, suggesting masked shared events might be present. We therefore repainted them excluding each other as donors: We performed similar reanalyses of five additional geographic regions for the same reason (table S16 and figs. S16 to S21). The easterly Russians and Chuvash both show evidence ($P < 0.05$) of admixture at more than one time (Fig. 2D), at least partially predating the Mongol empire, between groups with ancestry related to Northeast Asians (e.g., the Oroqen, Mongola, and Yakut) and Europeans, respectively (table S16). Six other European populations (Fig. 2D, pink/maroon box 2) independently show evidence after the repainting for similar admixture events involving more than two groups ($P < 0.02$) at about the same time (Fig. 3). CIs for the admixture time(s) overlap but predate the Mongol empire, with estimates from 440 to 1080 CE (Fig. 3). In each population, one source group has at least some ancestry related to Northeast Asians, with ~2 to 4% of these groups' total ancestry linking directly to East Asia. This signal might correspond to a small genetic legacy from invasions of peoples from the Asian steppes (e.g., the Huns, Magyar, and Bulgars) during the first millennium CE (*36*). The other two source groups appear much more local. One is more North European in the repainting, when we exclude other East European groups as donors, and is largely replaced by northern Slavic-speaking groups in our original analysis (Fig. 2D and table S12). The other source is more southerly (e.g., Greeks and West Asians). This local migration could explain a recent observation of an excess of identity-by-descent sharing in Eastern Europe—including in the Greeks, in whom we infer admixture involving a group represented by Poland, at the same time—that was dated to a wide range between 1000 and 2000 years ago (*37*). We speculate that these events may correspond to the Slavic expansion across this region at a similar time, perhaps related to displacement caused by the Eurasian steppe invaders (*38*).

Last, Central Asia shows a particularly complex inferred history after a reanalysis of 10 groups excluding each other as donors, with 9 of 10 groups showing diverse recent events (Fig. 4A). The exception is the Kalash, a genetically isolated (*39*) population from the Hindu Kush mountains of Pakistan (*40*). Distinct, ancient, and partially shared admixture signals (always dated older than 90 BCE) are seen in six groups (Fig. 4B), including the Kalash (Fig. 2C), whose strongest signal suggests a major admixture event (990 to 210 BCE) from a source related to present-day Western Eurasians, although we cannot identify the geographic origin precisely. This period overlaps that of Alexander the Great (356 to 323 BCE), whose army, local tradition holds, the Kalash are descended from (*40*), but these ancient events predate recorded history in the region, precluding confident interpretation.

Our results demonstrate that it is possible to elucidate the effect of ancient and modern migration events and to provide fine-scale details of the sources involved, the complexity of events, and the timing of mixing of groups by using genetic information alone. Where independent information exists from alternative historical or archaeological sources, our approach provides results consistent with known facts and determines the amount of genetic material exchanged. In other cases, novel mixture events we infer are plausible and often involve geographically nearby sources, supporting their validity. Admixture events within the past several thousand years affect most human populations, and this needs to be taken into account in inferences aiming to look at the more distant history of our species. Future improvements in whole-genome sequencing, greater sample sizes, and incorporation of ancient DNA, together with additional methodological extensions, are likely to allow better understanding of ancient events where little or no historical record exists, to identify many additional events, to infer sex biases, and to provide more precise event characterization than currently possible. We believe our approach will extend naturally to these settings, as well as to other species.

References and Notes

1. D. Falush, M. Stephens, J. K. Pritchard, *Genetics* **164**, 1567–1587 (2003).
2. S. J. E. Baird, *Heredity* **97**, 81–83 (2006).
3. J. E. Pool, R. Nielsen, *Genetics* **181**, 711–719 (2009).
4. P. Moorjani *et al.*, *PLOS Genet.* **7**, e1001373 (2011).
5. I. Pugach, R. Matveyev, A. Wollstein, M. Kayser, M. Stoneking, *Genome Biol.* **12**, R19 (2011).
6. J. K. Pritchard, M. Stephens, P. Donnelly, *Genetics* **155**, 945–959 (2000).
7. D. H. Alexander, J. Novembre, K. Lange, *Genome Res.* **19**, 1655–1664 (2009).
8. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, *PLOS Genet.* **8**, e1002453 (2012).
9. Information on materials and methods is available on *Science* Online.
10. L. L. Cavalli Sforza, W. Bodmer, *The Genetics of Human Populations* (Freeman, San Francisco, CA, 1971).
11. P.-R. Loh *et al.*, *Genetics* **193**, 1233–1254 (2013).
12. A. L. Price *et al.*, *Nat. Genet.* **38**, 904–909 (2006).
13. N. Patterson, A. L. Price, D. Reich, *PLOS Genet.* **2**, e190 (2006).
14. G. K. Chen, P. Marjoram, J. D. Wall, *Genome Res.* **19**, 136–142 (2009).
15. I. Gronau, M. J. Hubisz, B. Gulko, C. G. Danko, A. Siepel, *Nat. Genet.* **43**, 1031–1034 (2011).
16. H. Li, R. Durbin, *Nature* **475**, 493–496 (2011).
17. R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, C. D. Bustamante, *PLOS Genet.* **5**, e1000695 (2009).
18. A. Keinan, J. C. Mullikin, N. Patterson, D. Reich, *Nat. Genet.* **39**, 1251–1255 (2007).
19. J. N. Fenner, *Am. J. Phys. Anthropol.* **128**, 415–423 (2005).
20. B. M. Henn *et al.*, *PLOS Genet.* **8**, e1002397 (2012).
21. G. B. J. Busby *et al.*, *Proc. Biol. Sci.* **279**, 884–892 (2012).
22. J. Z. Li *et al.*, *Science* **319**, 1100–1104 (2008).
23. D. M. Behar *et al.*, *Nature* **466**, 238–242 (2010).
24. B. M. Henn *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5154–5162 (2011).
25. B. N. Howie, P. Donnelly, J. Marchini, *PLOS Genet.* **5**, e1000529 (2009).
26. An online atlas of worldwide human genetic admixture history is available at http://admixturemap.paintmychromosomes.com.
27. N. Patterson *et al.*, *Genetics* **192**, 1065–1093 (2012).
28. We converted g inferred generations to the admixture year: $1950 - (g + 1) \times 28$.
29. J. K. Pickrell, J. K. Pritchard, *PLOS Genet.* **8**, e1002967 (2012).
30. D. Chandler, *A History of Cambodia* (Westview, Boulder, CO, ed. 4, 2007).
31. C. P. Atwood, *Encyclopedia of Mongolia and the Mongol Empire* (Facts on File, New York, 2004).
32. E. E. Bacon, *Southwest. J. Anthropol.* **7**, 230–247 (1951).
33. T. Zerjal *et al.*, *Am. J. Hum. Genet.* **72**, 717–721 (2003).
34. S. Xu, L. Jin, *Am. J. Hum. Genet.* **83**, 322–336 (2008).
35. M. Gordon, *Slavery in the Arab World* (New Amsterdam, New York, ed. 1, 1989).
36. P. Heather, *Empires and Barbarians: Migration, Development and the Birth of Europe* (Macmillan, London, 2009).
37. P. Ralph, G. Coop, *PLOS Biol.* **11**, e1001555 (2013).
38. C. I. Beckwith, *Empires of the Silk Road: A History of Central Eurasia from the Bronze Age to the Present* (Princeton Univ. Press, Princeton, NJ, 2006).
39. N. A. Rosenberg *et al.*, *Science* **298**, 2381–2385 (2002).
40. M. Lines, *The Kalasha People of North-Western Pakistan* (Emjay Books International, Peshawar, Pakistan, 1999).

Acknowledgments: We are grateful for the John Fell Fund–University of Oxford, the NIH, the Wellcome Trust (S.M., grant 098387/Z/12/Z), the Biotechnology and Biological Sciences Research Council, the Royal Society/Wellcome Trust (G.H., grant 098386/Z/12/Z), and the Istituto Italiano di Antropologia for funding. J.F.W. is a director, stockholder, and employee of ScotlandsDNA (and formerly of EthnoAncestry). We thank S. Karachanak, D. Toncheva, P. Anagnostou, F. Cali, F. Brisighelli, V. Romano, G. LeFranc, C. Buresi, J. Ben Chibani, A. Haj-Khelil, S. Denden, R. Ploski, T. Hervig, T. Moen, P. Krajewski, and R. Herrera for providing samples for our genotyping and the blood donors and the staff of the Unità Operativa Complessa di Medicina Trasfusionale, Azienda Ospedaliera Umberto I, Siracusa (Italy). Data analyzed in this study may be downloaded via http://admixturemap.paintmychromosomes.com/. Raw genotype data are available at the Gene Expression Omnibus database online (www.ncbi.nlm.nih.gov/geo/), series accession number GSE53626.

Supplementary Materials

www.sciencemag.org/content/343/6172/747/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S21
Tables S1 to S16
Appendix
References (*41*–*82*)

22 July 2013; accepted 20 December 2013
10.1126/science.1243518

Precise and Ultrafast Molecular Sieving Through Graphene Oxide Membranes

R. K. Joshi,[1] P. Carbone,[2] F. C. Wang,[3] V. G. Kravets,[1] Y. Su,[1] I. V. Grigorieva,[1] H. A. Wu,[3] A. K. Geim,[1]* R. R. Nair[1]*

Graphene-based materials can have well-defined nanometer pores and can exhibit low frictional water flow inside them, making their properties of interest for filtration and separation. We investigate permeation through micrometer-thick laminates prepared by means of vacuum filtration of graphene oxide suspensions. The laminates are vacuum-tight in the dry state but, if immersed in water, act as molecular sieves, blocking all solutes with hydrated radii larger than 4.5 angstroms. Smaller ions permeate through the membranes at rates thousands of times faster than what is expected for simple diffusion. We believe that this behavior is caused by a network of nanocapillaries that open up in the hydrated state and accept only species that fit in. The anomalously fast permeation is attributed to a capillary-like high pressure acting on ions inside graphene capillaries.

Porous materials with a narrow distribution of pore sizes, especially in the angstrom range (*1*–*5*), are of interest for use in separation technologies (*5*–*7*). The observation of fast permeation of water through carbon nanotubes (*8*–*10*) and, more recently, through graphene-oxide (GO) laminates (*11*) has led to many proposals to use these materials for nanofiltration and desalination (*8*–*19*). GO laminates are particularly attractive because they are easy to fabricate and mechanically robust and should be amenable to industrial-scale production (*20*, *21*). They are made of impermeable functionalized graphene sheets that have a typical size of $L \approx 1$ μm and the interlayer separation, d, sufficient to accommodate a mobile layer of water (*11*–*25*). Nanometer-thick GO films have recently been tried for pressure-driven filtration, revealing promising characteristics (*15*–*18*). However, the results varied widely for different fabrication methods, and some observations relevant to the present report (permeation of large molecules) are inconsistent with the known structure of GO laminates (*20*, *21*). This suggests the presence of cracks or pin holes in those GO thin films, which obscured their intrinsic properties (*25*).

We studied micrometer-thick GO membranes prepared from GO suspensions using vacuum filtration as described in (*25*). The resulting membranes were checked for their continuity by using a helium leak detector before and after filtration experiments, which demonstrated that the membranes were vacuum-tight in the dry state (*11*). Schematics of our permeation experiments are shown in Fig. 1. The feed and permeate compartments were initially filled to the same height with different liquids, including water, glycerol, toluene, ethanol, benzene, and dimethyl sulfoxide (DMSO). No permeation could be detected over a period of many weeks by monitoring liquid levels and using chemical analysis (*25*). If both compartments were filled with water solutions, permeation through the same vacuum-tight membrane could be readily observed as rapid changes in liquid levels (several millimeters per day). For example, a level of a 1 M sucrose solution in the feed compartment rose, whereas it fell in the permeate compartment filled with deionized water. For a membrane with a thickness h of 1 μm, we found water flow rates of ≈0.2 L m^{-2} h^{-1}, and the speed increased with increasing the molar concentration C. Because a 1 M sucrose solution corresponds to an osmotic pressure of ≈25 bar at room temperature (the van't Hoff factor is 1 in this case), the flow rates agree with the evaporation rates of ≈10 L m^{-2} h^{-1} reported for similar membranes in (*11*), in which case, the permeation was driven by a capillary pressure of the order of 1000 bar. The hydrostatic pressures in our experiments never exceeded 10^{-2} bar and, therefore, could be neglected.

We next investigated the possibility that dissolved ions and molecules could diffuse through the capillaries simultaneously with water. We filled the feed compartment with various solutions to determine whether any of the solutes permeated into the deionized water on the other side of the GO membrane (Fig. 1B). As a quick test, ion transport can be probed by monitoring electrical conductivity of the permeate compartment (fig. S1). We found that for some salts (for example, NaCl), the conductivity increased with time, but for others {for example, $K_3[Fe(CN)_6]$}, it did not change over many days of measurements.

Depending on the solute, we used ion chromatography, inductively coupled plasma optical

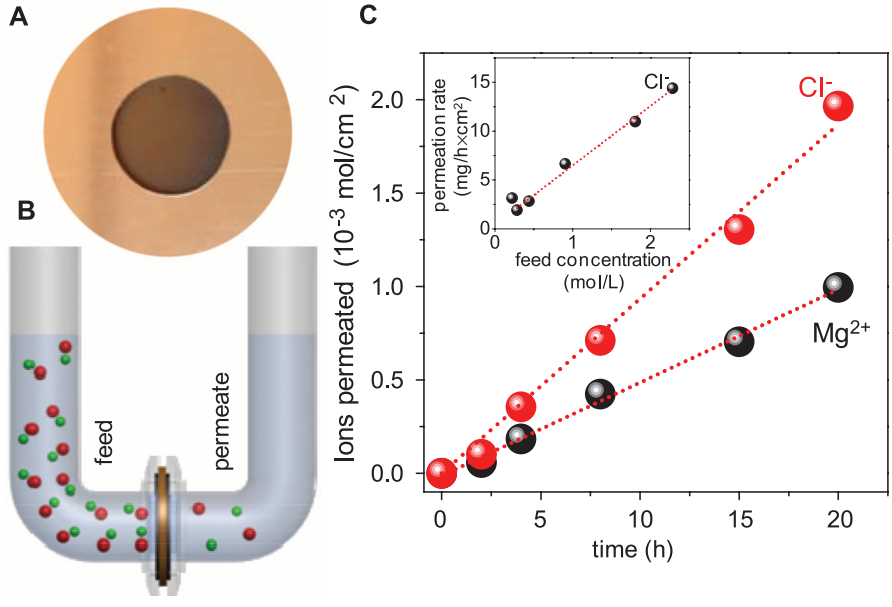


Fig. 1. Ion permeation through GO laminates. (**A**) Photograph of a GO membrane covering a 1-cm opening in a copper foil. (**B**) Schematic of the experimental setup. A U-shaped tube 2.5 cm in diameter is divided by the GO membrane into two compartments referred to as feed and permeate. Each is filled to a typical level of ~20 cm. Magnetic stirring is used so as to ensure no concentration gradients. (**C**) Permeation through a 5-μm-thick GO membrane from the feed compartment with a 0.2 M solution of $MgCl_2$. (Inset) Permeation rates as a function of C in the feed solution. Within our experimental accuracy (variations by a factor of <40% for membranes prepared from different GO suspensions), chloride rates were found the same for $MgCl_2$, KCl, and $CuCl_2$. Dotted lines are linear fits.

[1]School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK. [2]School of Chemical Engineering and Analytical Science, University of Manchester, Manchester M13 9PL, UK. [3]Chinese Academy of Sciences Key Laboratory of Mechanical Behavior and Design of Materials, Department of Modern Mechanics, University of Science and Technology of China, Hefei, Anhui 230027, China.

*Corresponding author. E-mail: rahul.raveendran-nair@manchester.ac.uk (R.R.N.); geim@man.ac.uk (A.K.G.)

emission spectrometry, total organic carbon analysis, and optical absorption spectroscopy (*25*) to measure permeation rates for a range of molecules and ions (table S1). An example of our measurements for $MgCl_2$ is shown in Fig. 1C, using ion chromatography and inductively coupled plasma optical emission spectrometry for Mg^{2+} and Cl^-, respectively. Concentrations of Mg^{2+} and Cl^- in the permeate compartment increased linearly with time, as expected. Slopes of such curves yield permeation rates. The observed rates depend linearly on concentration in the feed compartment (Fig. 1C, inset). Cations and anions move through membranes in stoichiometric amounts so that charge neutrality within each of the compartment is preserved. For example, in Fig. 1C permeation of one Mg^{2+} ion is accompanied by two Cl^- ions, and there is no electric field buildup across the membrane.

Our results obtained for different ionic and molecular solutions are summarized in Fig. 2. The small species permeate with approximately the same speed, whereas large ions and organic molecules exhibit no detectable permeation. The effective volume occupied by an ion in water is characterized by its hydrated radius. If plotted as a function of this parameter, our data are well described by a single-valued function with a sharp cutoff at ≈4.5 Å (Fig. 2). Species larger than this are sieved out. This behavior corresponds to a physical size of the mesh of ≈9 Å. Also, permeation rates do not exhibit any notable dependence on ion charge (Fig. 2) (*12*, *13*, *23*, *26*) because triply charged ions such as AsO_4^{3-} permeate with approximately the same rate as singly charged Na^+ or Cl^-. Last, to prove the essential role of water for ion permeation through GO laminates, we dissolved KCl and $CuSO_4$ in DMSO, the polar nature of which allows solubility of these salts. No permeation was detected, confirming the special affinity of GO laminates to water.

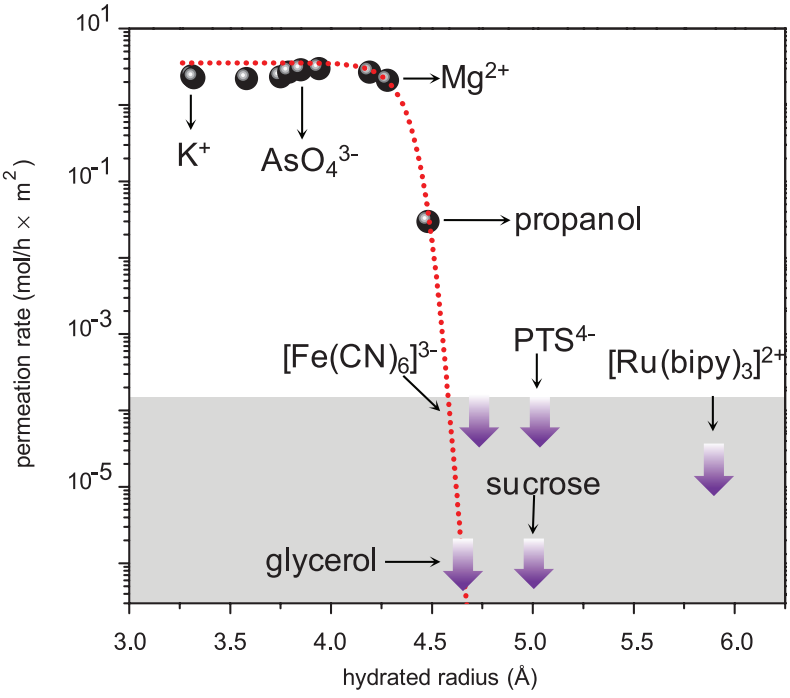
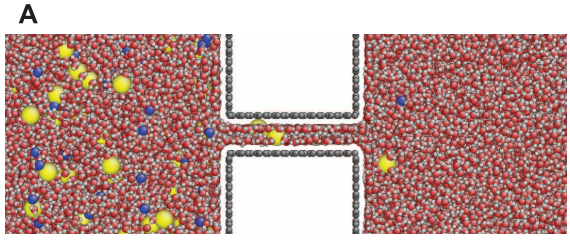


Fig. 2. Sieving through the atomic-scale mesh. The shown permeation rates are normalized per 1 M feed solution and measured by using 5-μm-thick membranes. Some of the tested chemicals are named here; the others can be found in table S1 (*25*). No permeation could be detected for the solutes shown within the gray area during measurements lasting for at least 10 days. The thick arrows indicate our detection limit, which depends on a solute. Several other large molecules—including benzoic acid, DMSO, and toluene—were also tested and exhibited no detectable permeation. The dashed curve is a guide to the eye, showing an exponentially sharp cutoff at 4.5 Å, with a width of ≈0.1 Å.

To explain the observed sieving properties, we use the model previously suggested to account for unimpeded evaporation of water through GO membranes (*11*). Individual GO crystallites have two types of regions: functionalized (oxidized) and pristine (*21*, *27*, *28*). The former regions act as spacers that keep adjacent crystallites apart and also prevent them from being dissolved. In a hydrated state, the spacers help water to intercalate between GO sheets, whereas pristine regions provide a network of capillaries that allow nearly frictionless flow of a layer of correlated water, similar to the case of water transport through carbon nanotubes (*8*–*10*). The earlier experiments using GO laminates in air (typical $d \approx 9 \pm 1$ Å) were explained by assuming one monolayer of moving water. For GO laminates soaked in water, d increases to ≈13 ± 1 Å (fig. S2), which allows two or three water layers (*19*, *22*, *23*, *29*). Taking into account the effective thickness of graphene of 3.4 Å (interlayer distance in graphite), this yields a pore size of ≈9 to 10 Å, which is in agreement with the mesh size found experimentally.

To support our model, we have used molecular dynamics (MD) simulations. The setup is shown in Fig. 3A, in which a graphene capillary separates feed and permeate reservoirs, and its width is varied between 7 and 13 Å to account for the possibility of one, two, or three layers of water (*25*). We find that the narrowest



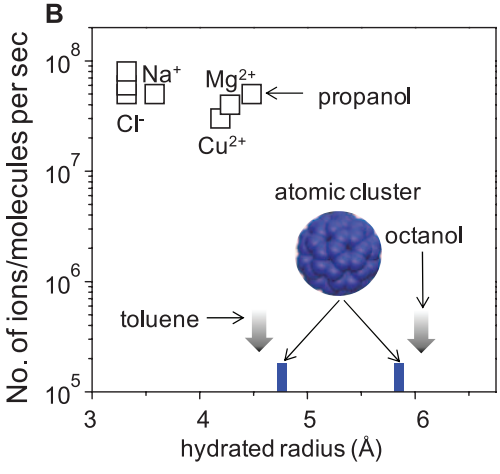


Fig. 3. Simulations of molecular sieving. (**A**) Snapshot of NaCl diffusion through a 9 Å graphene slit allowing two layers of water. Na^+ and Cl^- ions are in yellow and blue, respectively. (**B**) Permeation rates for NaCl, $CuCl_2$, $MgCl_2$, propanol, toluene, and octanol for such capillaries. For octanol poorly dissolved in water, the hydrated radius is not known, and we use its molecular radius. Blue marks indicate permeation cutoff for an atomic cluster (inset) for graphene capillaries accommodating two and three layers of water (widths of 9 and 13 Å, respectively).

capillaries become filled with a monolayer of water as described previously (*11*) and do not allow even such small ions as Na^+ and Cl^- inside. However, for two and three layers expected in the fully hydrated state (*25*) ions enter the capillaries and diffuse into the permeate reservoir. Their diffusion rates are found to be approximately the same for all small ions and show little dependence on ionic charge (Fig. 3B). Larger species (toluene and octanol) cannot permeate even through capillaries containing three layers of water (fig. S3). We have also modeled large solutes as atomic clusters of different size (*25*) and found that the capillaries accommodating two and three layers of water rejects clusters with the radius larger than ≈4.7 and 5.8 Å, respectively. This may indicate that the ion permeation through GO laminates is limited by regions containing two layers of water. The experimental and theory results in Figs. 2 and 3B show good agreement.

Following (*11*), we estimate that for our laminates with $h \approx 5$ μm and $L \approx 1$ μm, the effective length of graphene capillaries is $L \times h/d \approx 5$ mm and that they occupy $d/L \approx 0.1\%$ of the surface area. This estimate is supported by measuring the volume of absorbed water, which is found to match the model predictions (*25*). For a typical diffusion coefficient of ions in water ($\approx 10^{-5}$ cm^2 s^{-1}), the expected diffusion rate for a 1 M solution through GO membrane is $\approx 10^{-3}$ mol h^{-1} m^{-2} (*25*)—that is, several thousands of times smaller than the rates observed experimentally (Fig. 1C). Such fast transport cannot be explained by the confinement, which increases the diffusion coefficient by only a factor of 3/2, reflecting the change from bulk to two-dimensional water. Moreover, functionalized regions [modeled as graphene with randomly attached epoxy and hydroxyl groups (*20*, *21*)] do not enhance diffusion but rather suppress it (*25*, *29*) as expected because of the broken translational symmetry.

To understand the ultrafast ion permeation, we recall that graphite-oxide powders exhibit extremely high absorption efficiency with respect to many salts (*30*). Despite being densely stacked, our GO laminates are found to retain this property for salts with small hydrated radii [(*25*), section 6]. Our experiments show that permeating salts are absorbed in amounts reaching as much as 25% of the membrane's initial weight (fig. S2). The large intake implies highly concentrated solutions inside graphene capillaries (close to the saturation). Our MD simulations confirm that small ions prefer to reside inside capillaries (fig. S4). The affinity of salts to graphene capillaries indicates an energy gain with respect to the bulk water, and this translates into a capillary-like pressure that acts on ions within a water medium (*25*). Therefore, there is a large capillary force, sucking small ions inside GO laminates and facilitating their permeation. Our MD simulations provide an estimate for this ionic pressure as >50 bars (*25*).

The reported GO membranes exhibit extraordinary separation properties, and their full understanding will require further work both experimental and theoretical. With the ultrafast ion transport and atomic-scale pores, GO membranes already present an interesting material to consider for separation and filtration technologies, particularly those that target extraction of valuable solutes from complex mixtures. By avoiding the swelling of GO laminates in water (by using mechanical constraints or chemical binding), it may be possible to reduce the mesh size down to ~6 Å; in which case, one monolayer of water would still go through, but even the smallest salts would be rejected.

References and Notes

1. K. B. Jirage, J. C. Hulteen, C. R. Martin, *Science* **278**, 655–658 (1997).
2. N. B. McKeown, P. M. Budd, *Chem. Soc. Rev.* **35**, 675–683 (2006).
3. D. L. Gin, R. D. Noble, *Science* **332**, 674–676 (2011).
4. D. Cohen-Tanugi, J. C. Grossman, *Nano Lett.* **12**, 3602–3608 (2012).
5. S. P. Koenig, L. Wang, J. Pellegrino, J. S. Bunch, *Nat. Nanotechnol.* **7**, 728–732 (2012).
6. M. Ulbricht, *Polymer (Guildf.)* **47**, 2217–2262 (2006).
7. M. Elimelech, W. A. Phillip, *Science* **333**, 712–717 (2011).
8. J. K. Holt *et al.*, *Science* **312**, 1034–1037 (2006).
9. M. Majumder, N. Chopra, R. Andrews, B. J. Hinds, *Nature* **438**, 44 (2005).
10. G. Hummer, J. C. Rasaiah, J. P. Noworyta, *Nature* **414**, 188–190 (2001).
11. R. R. Nair, H. A. Wu, P. N. Jayaram, I. V. Grigorieva, A. K. Geim, *Science* **335**, 442–444 (2012).
12. M. Majumder, N. Chopra, B. J. Hinds, *J. Am. Chem. Soc.* **127**, 9062–9070 (2005).
13. F. Fornasiero *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17250–17255 (2008).
14. L. Qiu *et al.*, *Chem. Commun. (Camb.)* **47**, 5810–5812 (2011).
15. P. Sun *et al.*, *ACS Nano* **7**, 428–437 (2013).
16. Y. Han, Z. Xu, C. Gao, *Adv. Funct. Mater.* **23**, 3693–3700 (2013).
17. M. Hu, B. Mi, *Environ. Sci. Technol.* **47**, 3715–3723 (2013).
18. H. Huang *et al.*, *Chem. Commun. (Camb.)* **49**, 5963–5965 (2013).
19. K. Raidongia, J. Huang, *J. Am. Chem. Soc.* **134**, 16528–16531 (2012).
20. D. A. Dikin *et al.*, *Nature* **448**, 457–460 (2007).
21. G. Eda, M. Chhowalla, *Adv. Mater.* **22**, 2392–2415 (2010).
22. A. Lerf *et al.*, *J. Phys. Chem. Solids* **67**, 1106–1110 (2006).
23. D. W. Boukhvalov, M. I. Katsnelson, Y.-W. Son, *Nano Lett.* **13**, 3930–3935 (2013).
24. S. You, S. M. Luzan, T. Szabó, A. V. Taluzin, *Carbon* **52**, 171–180 (2013).
25. Materials and methods are available as supplementary materials on *Science* Online.
26. B. Van der Bruggen, J. Schaep, D. Wilms, C. Vandecasteele, *J. Membr. Sci.* **156**, 29–41 (1999).
27. N. R. Wilson *et al.*, *ACS Nano* **3**, 2547–2556 (2009).
28. D. Pacilé *et al.*, *Carbon* **49**, 966–972 (2011).
29. N. Wei, Z. Xu, Breakdown of fast water transport in graphene oxides. arXiv:1308.5367.
30. S. Wang, H. Sun, H. M. Ang, M. O. Tade, *Chem. Eng. J.* **226**, 336–347 (2013).

Acknowledgments: This work was supported by the European Research Council, the Royal Society, Engineering and Physical Research Council (UK), and the National Natural Science Foundation (China). We thank A. Mishchenko and J. Waters for help. R.K.J. also acknowledges support by the Marie Curie Fellowship, and R.R.N. acknowledges support by the Leverhulme Trust and Bluestone Global Tech.

Supplementary Materials

www.sciencemag.org/content/343/6172/752/suppl/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References (*31*–*52*)

9 September 2013; accepted 9 January 2014
10.1126/science.1245711

Designing Collective Behavior in a Termite-Inspired Robot Construction Team

Justin Werfel,[1]* Kirstin Petersen,[1,2] Radhika Nagpal[1,2]

Complex systems are characterized by many independent components whose low-level actions produce collective high-level results. Predicting high-level results given low-level rules is a key open challenge; the inverse problem, finding low-level rules that give specific outcomes. We present a multi-agent construction system inspired by mound-building termites, solving such an inverse problem. A user specifies a desired structure, and the system automatically generates low-level rules for independent climbing robots that guarantee production of that structure. Robots use only local sensing and coordinate their activity via the shared environment. We demonstrate the approach via a physical realization with three autonomous climbing robots limited to onboard sensing. This work advances the aim of engineering complex systems that achieve specific human-designed goals.

In contrast to the careful preplanning and regimentation that characterize human construction projects, animals that build in groups do so in a reactive and decentralized way. The most striking examples are mound-building termites, colonies of which comprise millions of independently behaving insects that build intricate structures orders of magnitude larger than themselves (*1*, *2*) (Fig. 1, A and B). These natural systems inspire us to envision artificial ones operating via similar principles (*3*, *4*), with independent agents

[1]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA 02138, USA. [2]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA.
*Corresponding author. E-mail: justin.werfel@wyss.harvard.edu

acting together to build elaborate large-scale structures, guided by reacting to the local situations they encounter. Such systems could enable construction in settings where human presence is dangerous or problematic, as in disaster areas or extraterrestrial environments.

Engineering an automated construction system that operates by termite-like principles rather than human-like ones requires an ability to design complex systems with desired collective behavior (e.g., producing a particular user-specified building). The hallmark of complex systems of independent agents (*5*–*7*) is unexpected collective behavior that emerges from their joint actions, not readily predictable from knowledge of agent rules. If a specific collective behavior is desired, no method in general is known to find agent rules that will produce it.

We present a decentralized multi-agent system for automated construction of user-specified structures, thereby providing a solution to such a problem of complex system design. An arbitrary number of independent robots follow an identical set of simple, local rules that collectively produce a specific structure requested by a user (Fig. 1, C and D). The rules are automatically generated from a high-level representation of the final target structure and provide provable guarantees of correct completion of that structure. The challenges associated with engineering a complex system are addressed by using principles drawn from social insects—in particular, indirect coordination through manipulation and sensing of a shared environment (stigmergy), and behavioral regularities that constrain the space of possible outcomes—which together make analysis and execution tractable. We first present the theoretical foundation for this work, followed by a physical implementation with three independent robots demonstrating autonomous construction.

The independence of individual robots stands in contrast to other work on automating construction with single (*8*–*10*) or multiple (*11*–*14*) robots with centralized sensing and/or control. Centralized systems that provide a global computing authority and/or precise positioning information during run time, in settings where such features are feasible, can have advantages in aspects such as efficiency and run-time flexibility. Conversely, decentralization provides advantages including opportunities for greater scalability (no coordinating authority that can become overloaded) and robustness (no single point of failure).

We distinguish between two types of building processes (Fig. 2). A system may produce a predetermined outcome, in which many possible system trajectories all lead to the same guaranteed final state. Alternatively, variation during the process may lead to a variable outcome, in which the final state is determined during the course of construction and can change if the process is rerun. In the context of human construction, single buildings are built via the first type of process, in which the order of operations might vary but the final result always matches a blueprint; cities develop via the second type of process, in which choices are contingent on previous decisions such that many distinct results are possible. Here, we focus on designing processes with fixed outcomes, but also show how our system can be used to generate structures that vary each time robots construct them.

Our system design is motivated by the goal of relatively simple, independent robots with limited capabilities (*15*), able to autonomously build a large class of nontrivial structures using a single type of prefabricated building material (solid "bricks"). We require a robot to be able to move forward, move back, and turn in place; climb up or down a step the height of one brick; and pick up one brick, move while carrying it, and attach it directly in front of itself at its own level. Robots can build staircases of bricks to climb to higher levels. Robots are limited to local sensing, able to perceive only bricks and other robots in their immediate vicinity. Information about the current state of the overall structure and the actions of more distant robots is not available. Robots obtain information about where bricks have been attached only through direct inspection; after they leave an area, this information is liable to become outdated as other robots modify the structure. The structure, built from square bricks in a nonoverlapping grid pattern, provides a reference that robots can use to keep track of their relative movement around it. A single "seed" brick, the initiation point from which the contiguous structure is built, provides a unique landmark.

We take an approach derived from the classically insect-inspired notion of stigmergy (*2*, *3*, *16*), in which, instead of any explicit broadcast or one-to-one communication between agents, all communication is implicit via the joint manipulation of a shared environment. In particular, we focus on qualitative stigmergy (*2*) in which actions are triggered by qualitatively different stimuli, such as distinct arrangements of building material. Robots in our system add bricks to the structure in response to existing configurations of bricks. In doing so, the rules they follow must be constructed in such a way that correct completion of the target structure is guaranteed, despite stale information about other parts of the structure, and irrespective of the (potentially variable) number

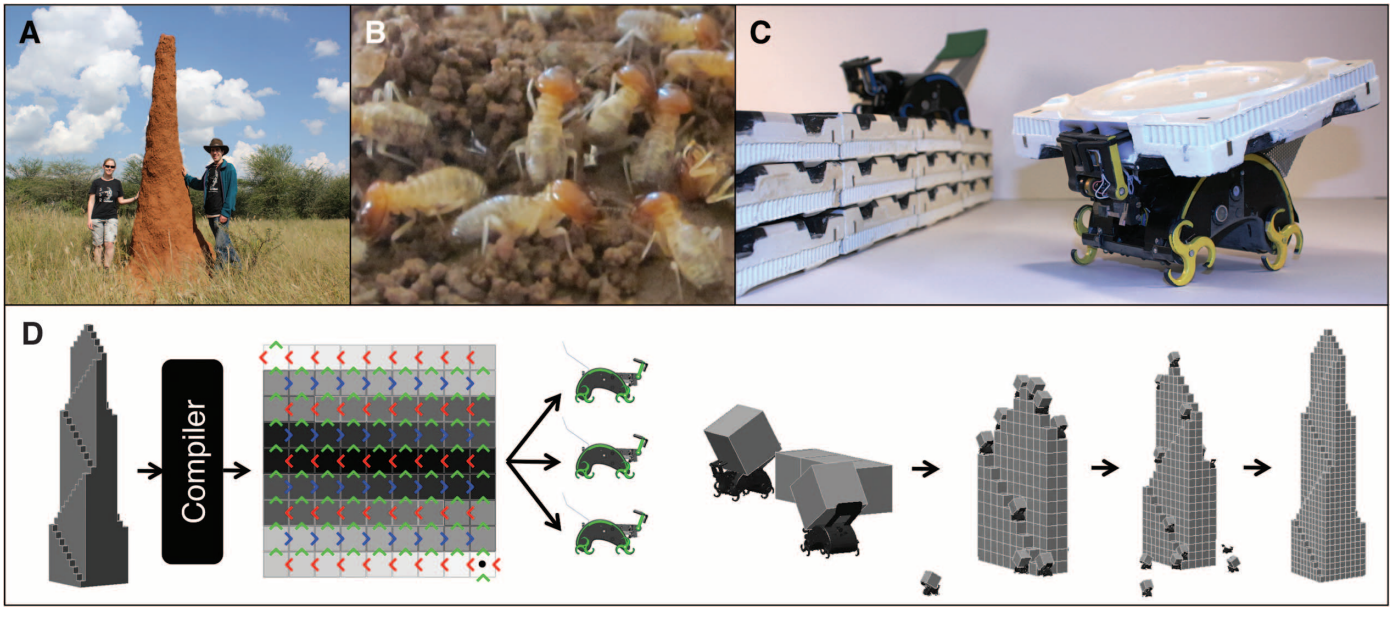


Fig. 1. Natural and artificial collective construction. (**A** and **B**) Complex meter-scale termite mounds (A) are built by millimeter-scale insects (B), which act independently with local sensing and limited information. (**C**) Physical implementation of our system, with independent climbing robots that build using specialized bricks. (**D**) System overview for building a specific predetermined result (Fig. 2, A and C): A user specifies a desired final structure; an offline compiler converts it to a "structpath" representation (Fig. 3), which is provided to all robots; robots follow local rules that guarantee correct completion of the target structure (movie S1).

of other robots and the order and timing of their own actions. The fact that all robots follow the same rules (*17*)—as all insects in the same colony follow consistent behavior patterns, or other animals obey intraspecific social conventions (*18*)—helps to constrain outcomes and restricts the space of possible situations that robots typically encounter and must be able to handle.

To ensure a predetermined outcome, robot rules are based on sensed brick configurations plus a static internal representation of the target structure. A user provides a picture or other high-level representation of the desired final structure, specifying what sites are ultimately meant to be occupied by bricks. An offline compilation step converts this to a "structpath" representation that provides movement guidelines for robots at each location—in a sense, traffic laws appropriate for that structure (*19*) (Fig. 3). Robots then follow a fixed set of simple rules (*19*), referring to the static structpath representation and otherwise identical for any target structure; these ensure the growth of that structure to completion in a way consistent with robot capabilities (movie S1). The rules rely on locally available information, preserve a robot's ability to move freely over the structure and opportunities for parallelism, and prevent deadlocks and other situations where the actions of one robot interfere with those of another. Robots act reactively (*20*, *21*); they do not preplan their actions, as is appropriate in this decentralized multirobot approach in which a robot setting out to perform a specific task might find it already completed by the time it gets there.

The structpath representation specifies a set of paths that robots can follow through the structure layout while respecting their movement constraints. In particular, all paths start from the seed and require climbing up or down a height of at most one brick at a time. The structpath specifies a fixed direction for robots to travel between each pair of neighboring sites; off the structure, robots follow its perimeter strictly counterclockwise. This directional restriction smooths traffic flow, ensures a flow of material into the growing structure (avoiding excessive backtracking from, e.g., laden robots making way for unladen ones to exit), and allows regularities in structure growth that let local rules ensure the preservation of global invariants. Paths may split and merge; a robot may have more than one way of leaving or entering a site. A multiplicity of possible paths helps the system exploit the parallelism of the swarm. The compiler performs a recursive search to identify a set of paths without cycles that meets these requirements, or to determine that none exists.

Individual robots then repeat the following routine: With a brick, circle the structure perimeter until reaching the seed; climb onto the structure and move along any legal path, keeping track of relative position with respect to the seed; attach the brick at any vacant site whose local neighborhood satisfies a fixed set of geometric requirements (*19*); continue to follow the path off the structure; obtain a new brick. These rules can be shown to guarantee successful completion of the target structure while ensuring that no intermediate state calls for a robot to perform tasks beyond its capabilities—in particular, climb or descend a height of greater than one brick, attach a brick at a higher or lower level than itself, or force a brick into place directly between two others (a mechanically difficult operation requiring high precision) (*19*). Direct interaction

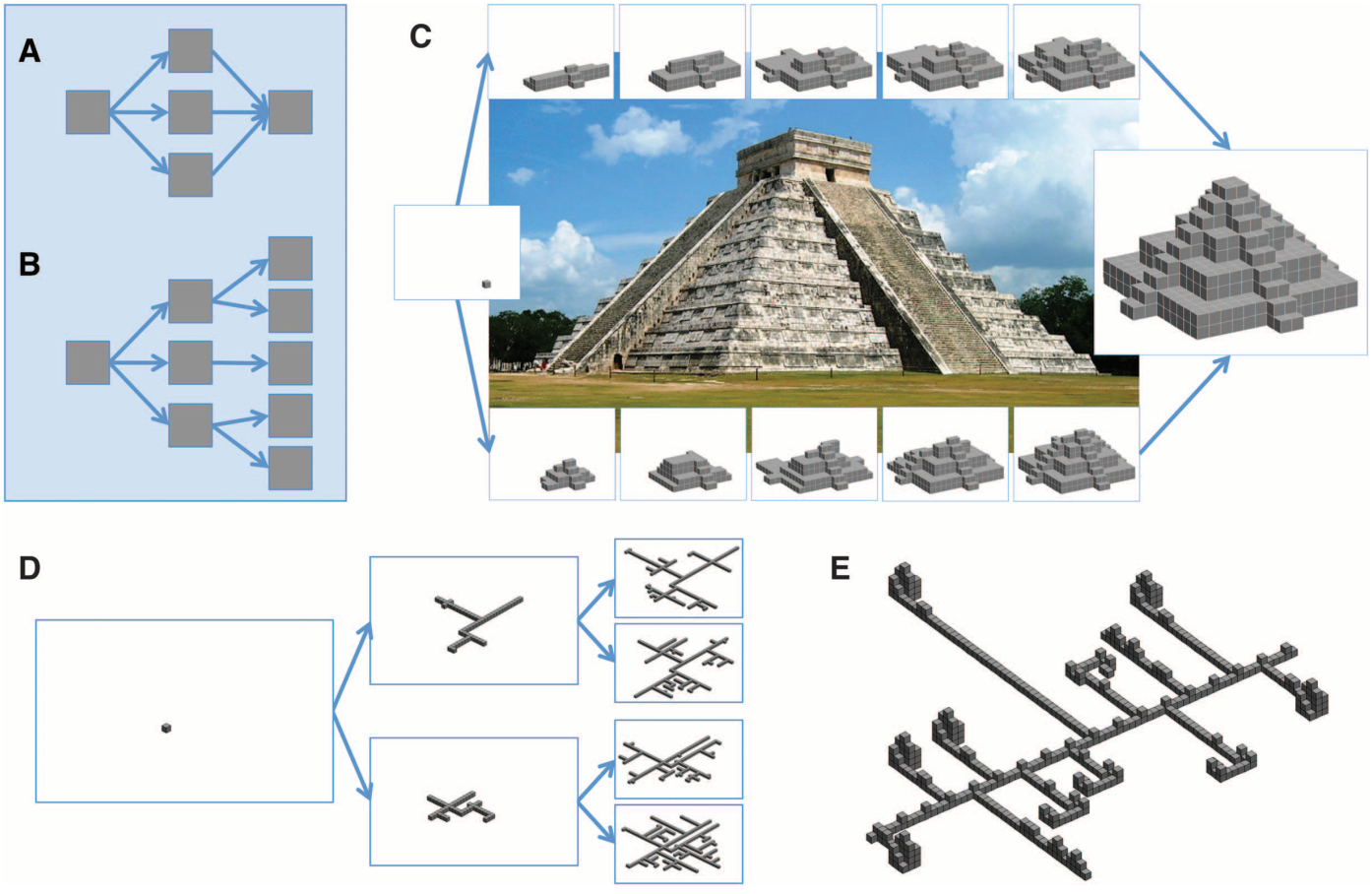


Fig. 2. Two types of building process. (**A**) Different possible sequences all lead to the same predetermined endpoint. (**B**) Different sequences lead to different results, which are determined only during the course of construction. (**C**) Example of the first type of process, building a step pyramid modeled after the main temple at Chichen Itza [photo by Kyle Simourd, CC BY 2.0]. Upper and lower panels show snapshots from different possible sequences (movie S2), at approximately 10%, 25%, 50%, 67%, and 80% completion of the common final structure. (**D**) Example of the second type of process, building a set of one-brick-high ramifying paths. (**E**) A hybrid system can combine elements of both types; in this example, paths of stochastically determined lengths lead to buildings chosen randomly from a set of predefined structures. See (*19*) for details of agent rules for all cases.

between robots is limited to each one yielding to the one ahead of it in this physical loop.

Because robots may take multiple possible paths through a structure, the ordering of the building process can occur in many different ways. Accordingly, the structure will emerge in different ways in different instances of building with the same structpath, with intermediate structures that may be observed in one instance but not another; however, the agent rules guide the process to always end in the same final structure (movie S2).

In addition to this approach for producing predetermined structures, the same robots can use different local rules to build structures whose detailed form emerges from the construction process. Multiple structures built with the same rules share qualitative features but differ in detail. Such a rule set could, for example, be used to generate a randomized street layout for a building complex. Figure 2E shows an example of a hybrid system built by such a rule set (*19*), where buildings chosen randomly from a set of predefined types are positioned at the ends of lanes of stochastically determined lengths. The robots again use stigmergy to coordinate their actions; for example, particular configurations of bricks constitute cues to agree on which building type should be constructed at the end of a given lane.

To demonstrate the feasibility of such a decentralized multirobot construction system, we present a proof-of-concept implementation in hardware (*19*) (Fig. 1C and Fig. 4). Design choices were driven by the requisite primitive operations that robots must perform: pick up a brick from a cache; attach a brick directly in front of themselves; detect nearby robots; when on the structure, move forward one site (while staying at the same level or climbing up or down one brick) or turn in place 90° left or right; when off the structure, circle its perimeter. For locomotion, we equipped robots with whegs [hybrid wheel-legs (*22*)], chosen for their empirical effectiveness in climbing (*23*). Each robot is equipped with seven active infrared sensors to detect black-and-white patterns on the bricks and ground for navigation; an accelerometer to register tilt angle for climbing and descent; an arm to lift and lower

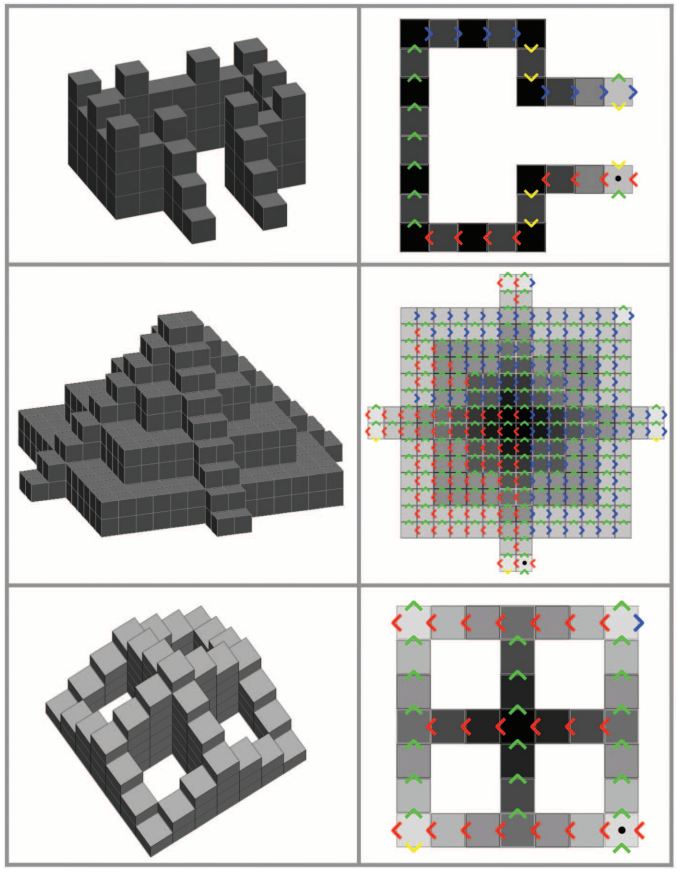


Fig. 3. Target structures and corresponding structpaths. For each predefined target structure at left, the corresponding structpath representation at right is generated by the offline compiler (*19*). From top to bottom: a simple structure with a unique structpath if the seed location is given; the temple of Fig. 2C, showing one of many possible structpaths; a structure enclosing internal courtyards. Sites in the structpath are shaded according to height (darker = higher); a dot marks the seed brick. Directions are color-coded to clarify flows (red, left; blue, right; green, up; yellow, down).

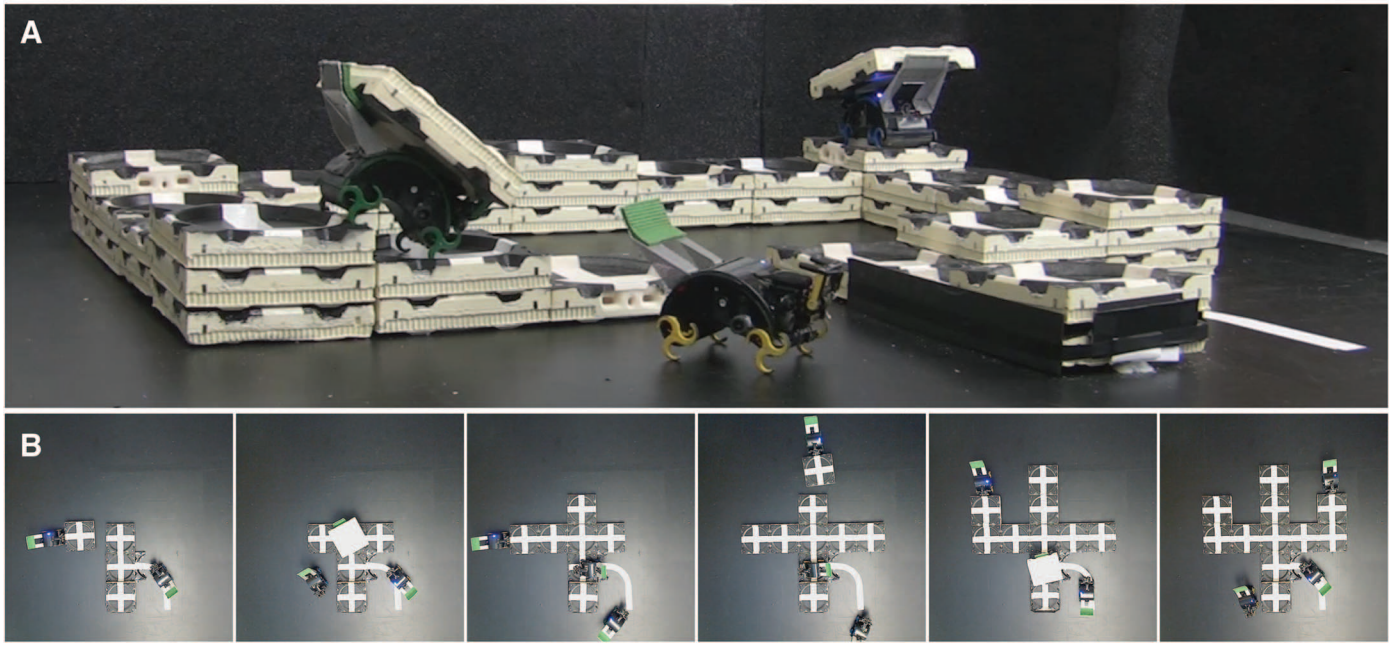


Fig. 4. Hardware demonstration. Independent autonomous robots with purely onboard sensing collectively work on prespecified structures. (**A**) A castle-like structure (movie S3). (**B**) A sequence of overhead snapshots building a branching structure (movie S4).

bricks, with a spring-loaded gripper to hold them securely while carried; and five ultrasound sonar units that let each robot evaluate and maintain its distance from the structure perimeter, as well as detect other sonar-using robots nearby. The robot footprint (17.5 cm × 11.0 cm) is smaller than that of the bricks, facilitating their maneuverability atop a wall one brick wide. Bricks (21.5 cm × 21.5 cm × 4.5 cm) are made from expanded urethane foam, with physical features to achieve self-alignment and neodymium magnets for attachment.

This hardware system demonstrates multiple simultaneously active, independent robots executing the full algorithm with entirely on-board sensing. supplementary movies show fully autonomous robots working on different user-specified structures (Fig. 4 and movies S3 and S4), adding bricks both atop the structure and on the ground, climbing over the structure as they build it, and adapting to one another's presence and actions, without human intervention beyond reloading the brick cache. The reactive nature of the approach can be demonstrated via externally imposed changes made to a structure while robots work on it (movie S5). Many approaches to mobile robotics deal with the continuous, noisy real world by probabilistically modeling its uncertainty (*24*); our system instead uses carefully engineered hardware to effectively discretize robot actions on the structure. Sensor feedback and brick features matched to the robots allow reasonable reliability with simple control (*19*). Minor deviations from ideal behavior are corrected by compensatory routines and/or passive mechanical features; for example, drift in position atop a structure is reduced both by robots checking their pose with respect to the brick markings, and by indentations on brick upper faces that guide the robots to stay within tolerance. Although extending this research prototype to a full production system would require solving many additional engineering challenges (*19*), our work demonstrates that physical hardware can allow the discretized theory to sufficiently represent the continuous reality.

This work provides an example of an engineered complex system, with multiple autonomous robots following simple, local rules and collectively achieving a specific desired result. Tools drawn from the social insects that inspire our approach—the exploitation of regularities that arise from identical programming in multi-agent systems, and the use of the environment as a means of implicit coordination—make these results possible. Future progress in our ability to design complex systems will advance our capacity to engineer systems that work as nature does (*25*–*27*), with large numbers of functionally limited, interchangeable parts, individually unreliable, collectively robust.

References and Notes

1. J. S. Turner, *Nat. Hist.* **111**, 62–67 (2002).
2. S. Camazine *et al.*, *Self-Organization in Biological Systems* (Princeton Univ. Press, Princeton, NJ, 2001).
3. G. Theraulaz, E. Bonabeau, *Science* **269**, 686–688 (1995).
4. J. Werfel, thesis, Massachusetts Institute of Technology (2006).
5. P. W. Anderson, *Science* **177**, 393–396 (1972).
6. Y. Bar-Yam, *Dynamics of Complex Systems* (Westview, Boulder, CO, 1997).
7. M. Mitchell, *Complexity: A Guided Tour* (Oxford Univ. Press, New York, 2009).
8. B. Khoshnevis, *Autom. Constr.* **13**, 5–19 (2004).
9. S. Wismer, G. Hitz, M. Bonani, A. Gribovski, S. Magnenat, Autonomous construction of a roofed structure: Synthesizing planning and stigmergy on a mobile robot. In *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (7 to 12 October 2012, Vilamoura, Algarve, Portugal), pp. 5436–5437.
10. F. Nigl, S. Li, J. E. Blum, H. Lipson, *IEEE Robot. Autom. Mag.* **20**, 60–71 (2013).
11. Q. Lindsey, D. Mellinger, V. Kumar, Construction of cubic structures with quadrotor teams. In *Robotics: Science and Systems VII*, H. Durrant-Whyte, N. Roy, P. Abbeel, Eds. (MIT Press, Cambridge, MA, 2012), pp. 177–184.
12. J. Willmann *et al.*, *Int. J. Archit. Comput.* **10**, 439–460 (2012).
13. A. Bolger *et al.*, Experiments in decentralized robot construction with tool delivery and assembly robots. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems* (18 to 22 October 2010, Taipei), pp. 5085–5092.
14. J. Worcester, J. Rogoff, M. A. Hsieh, Constrained task partitioning for distributed assembly. In *Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems* (25 to 30 September 2011, San Francisco), pp. 4790–4796.
15. R. A. Brooks, A. M. Flynn, *J. Br. Interplanet. Soc.* **42**, 478–485 (1989).
16. P.-P. Grassé, *Insectes Soc.* **6**, 41–81 (1959).
17. Y. Shoham, M. Tennenholtz, *Artif. Intell.* **73**, 231–252 (1995).
18. J. Maynard Smith, *J. Theor. Biol.* **47**, 209–221 (1974).
19. See supplementary materials on *Science* Online.
20. R. Brooks, *IEEE J. Robot. Autom.* **2**, 14–23 (1986).
21. R. Arkin, *Behavior-Based Robotics* (MIT Press, Cambridge, MA, 1998).
22. R. D. Quinn *et al.*, Insect designs for improved robot mobility. In *Proceedings of the 4th International Conference on Climbing and Walking Robots* (24 to 26 September 2001, Karlsruhe, Germany), pp. 69–76.
23. K. Petersen, R. Nagpal, J. Werfel, TERMES: An autonomous robotic system for three-dimensional collective construction. In *Robotics: Science and Systems VII*, H. Durrant-Whyte, N. Roy, P. Abbeel, Eds. (MIT Press, Cambridge, MA, 2012), pp. 257–264
24. S. Thrun, W. Burgard, D. Fox, *Probabilistic Robotics* (MIT Press, Cambridge, MA, 2005).
25. E. Bonabeau, M. Dorigo, G. Théraulaz, *Swarm Intelligence: From Natural to Artificial Systems* (Oxford Univ. Press, New York, 1999).
26. A. J. Ijspeert, A. Martinoli, A. Billard, L. M. Gambardella, *Auton. Robots* **11**, 149–171 (2001).
27. M. Brambilla, E. Ferrante, M. Birattari, M. Dorigo, *Swarm Intell.* **7**, 1–41 (2013).

Acknowledgments: Supported by the Wyss Institute for Biologically Inspired Engineering.

Supplementary Materials

www.sciencemag.org/content/343/6172/754/suppl/DC1
Materials and Methods
Figs. S1 to S12
Table S1
References (*28*–*34*)
Movies S1 to S5

11 September 2013; accepted 30 December 2013
10.1126/science.1245842

High-Energy Surface X-ray Diffraction for Fast Surface Structure Determination

J. Gustafson,[1]* M. Shipilin,[1] C. Zhang,[1] A. Stierle,[2,3] U. Hejral,[2,3] U. Ruett,[2] O. Gutowski,[2] P.-A. Carlsson,[4] M. Skoglundh,[4] E. Lundgren[1]

Understanding the interaction between surfaces and their surroundings is crucial in many materials-science fields, such as catalysis, corrosion, and thin-film electronics, but existing characterization methods have not been capable of fully determining the structure of surfaces during dynamic processes, such as catalytic reactions, in a reasonable time frame. We demonstrate an x-ray-diffraction–based characterization method that uses high-energy photons (85 kiloelectron volts) to provide unexpected gains in data acquisition speed by several orders of magnitude and enables structural determinations of surfaces on time scales suitable for in situ studies. We illustrate the potential of high-energy surface x-ray diffraction by determining the structure of a palladium surface in situ during catalytic carbon monoxide oxidation and follow dynamic restructuring of the surface with subsecond time resolution.

Understanding solid surfaces and their interactions with their surroundings has been a major research field for decades, motivated by important areas such as catalysis, corrosion, nanotechnology, and thin-film electronics (*1*–*6*). Therefore, a large number of experimental techniques with high surface sensitivity have been developed. However, many of these techniques, such as low-energy electron diffraction (LEED), x-ray photoelectron spectroscopy, and low-energy ion scattering (*1*–*3*), gain surface sensitivity from the limited mean-free path of the electrons or ions used as probes. Hence, these methods typically need ultra-high vacuum (UHV) conditions, and the use of these techniques to study surface structures under near-ambient conditions requires several stages of differential pumping.

[1]Synchrotron Radiation Research, Lund University, Box 118, SE-221 00 Lund, Sweden. [2]Deutsches Elektronen-Synchrotron (DESY), D-22603 Hamburg, Germany. [3]Fachbereich Physik Universität Hamburg, Jungiusstrasse 9, 20355 Hamburg, Germany. [4]Competence Centre for Catalysis, Chalmers University of Technology, SE-412 96, Göteborg, Sweden.

*Corresponding author. E-mail: johan.gustafson@sljus.lu.se

In contrast, surface x-ray diffraction (SXRD) (*7*–*9*) is a photon-in–photon-out technique, and because x-rays interact weakly with matter, the method is much less influenced by any gas surrounding the sample. When x-rays interact with a crystalline material, the scattered x-rays form a diffraction pattern, which provides structural information about the crystal's reciprocal lattice—the Fourier transform of the real-space atomic lattice—and specifically the periodicity (and thus indirectly the positions) of the atoms. The reciprocal lattice will be dominated by Bragg reflections from diffraction in the bulk of the crystal. Because of the broken periodicity at the surface, so-called crystal truncation rods (CTRs) connect the Bragg reflections perpendicular to the surface. In case the in-plane periodicity at the surface differs from that of the underlying bulk, additional superstructure rods arise in reciprocal space. The shapes of these CTRs and superstructure rods hold detailed information about the atomic surface structure, but the weak interaction of x-rays with matter requires intense synchrotron radiation to detect the surface signal. Thus, SXRD is one of very few methods available for surface structure determination under ambient conditions.

A serious drawback of conventional SXRD, with x-rays in the range of 10 to 30 keV and a point or small two-dimensional (2D) detector, is the limited amount of data that can be acquired in a reasonable time frame. Exploring 2D maps from a substantial part of reciprocal space is extremely time-consuming, and mapping of the 3D reciprocal space with high resolution is currently impossible even with synchrotron radiation. As a result, the probed surface structure has to be known qualitatively from other measurements, and an unexpected structure may easily be left unnoticed, especially under harsh conditions. Furthermore, obtaining a quantitative data set (from an already qualitatively known structure) takes on the order of 10 hours with traditional use of SXRD.

We demonstrate how the use of high-energy x-rays (85 keV) in combination with a large 2D detector accelerates the data collection by several orders of magnitude and enables full surface-structure determination by 3D mapping of reciprocal space on a time scale suitable for in situ studies. In addition, the small diffraction angles, resulting from the high photon energy, and the large detector result in data that are easily presented in a more intuitive way, because each detector image contains the projection of a full plane in reciprocal space and straight lines in reciprocal space correspond to straight lines on the detector. We provide a proof-of-principle demonstration of high-energy SXRD (HESXRD) by performing a full determination of the surface structure during catalytic CO oxidation over Pd(100). However, our method is general and also applicable to other surface- or interface-related in situ studies.

Figure 1A shows a schematic map of the reciprocal lattice of the Pd(100) crystal surface used in our study, including Bragg reflections (green spots) as well as CTRs (green lines). To probe a certain point in the reciprocal lattice, it is necessary to (i) rotate the sample so that the so-called Ewald sphere—the spherical surface in reciprocal space, which is probed at a given orientation between the sample and the beam—intersects this point and (ii) place the detector so that the correspondingly scattered x-rays are detected. Using traditional photon energies (10 to 30 keV) necessitates moving the detector around the sample, while only being able to probe a small part of reciprocal space simultaneously. For higher photon energies, however, the size of the Ewald sphere increases and the scattering angles decrease. Hence, the use of high-energy x-rays enables the detection of large volumes of reciprocal space with a stationary 2D detector.

The section of the Ewald sphere that is covered by the detector in our experiment is shown schematically in Fig. 1A. The red circles indicate points where a CTR intersects the Ewald sphere and where the intensity of scattered x-rays is recorded. By using an incidence angle close to the critical angle for total external reflection (0.04° in this case), the surface signal is maximized. The corresponding detector image is shown in Fig. 1B, where the two CTRs marked in Fig. 1A are indicated by arrows. Also seen in the detector image are the shadows of tungsten pieces placed in front of the detector to protect it from the high-intensity Bragg reflections (about 7 orders of magnitude more intense than the CTRs). The lack of intensity in the bottom center and the top corners of the detector image is due to a mask that prevented x-rays scattered in the walls of the experimental chamber to reach the detector (*10*).

To map out the reciprocal space, the sample (and hence the reciprocal lattice) is rotated around the surface normal, such that the Ewald sphere scans through the reciprocal space as illustrated in movie S1. The result is a 3D data set of the diffraction from the Pd(100) surface, and the scan takes on the order of 10 min. With HESXRD, we performed a full surface structure determination in situ during catalytic CO oxidation over Pd(100), which has not been possible before. In a specially designed SXRD flow reactor (*11*), the sample was exposed to a total gas pressure of 75 Torr at 600 K. In the gas mixture, a flow of 2 ml/min of O_2 and 4 ml/min of CO was set using Ar as carrier gas, resulting in partial pressures of O_2 and CO of ~3 and 6 Torr, respectively. Under these conditions, the Pd(100) surface is highly catalytically active, converting all of the CO reaching the surface to CO_2. During the reaction, HESXRD data were collected by a sample rotation (as described above) over 90°, and the diffraction intensities were collected every 0.1° with an exposure time of 1 s. The data show a surface with a ($\sqrt{5}$ by $\sqrt{5}$) rotated (R) 27° surface oxide structure (henceforth denoted $\sqrt{5}$), consisting of a single PdO(101) plane on top of the Pd(100) surface (Fig. 2A) (*12*, *13*).

The data can be visualized in different ways. In Fig. 2B, we show the result of combining 900 detector images to render the maximum-intensity result for each pixel. Such a projection onto the rotational plane provides a direct view of all the CTRs (indexed below the image) and superstructure rods (indexed above) within the probed volume. One can immediately identify the different structures present at the surface and draw some qualitative conclusions. For instance, the

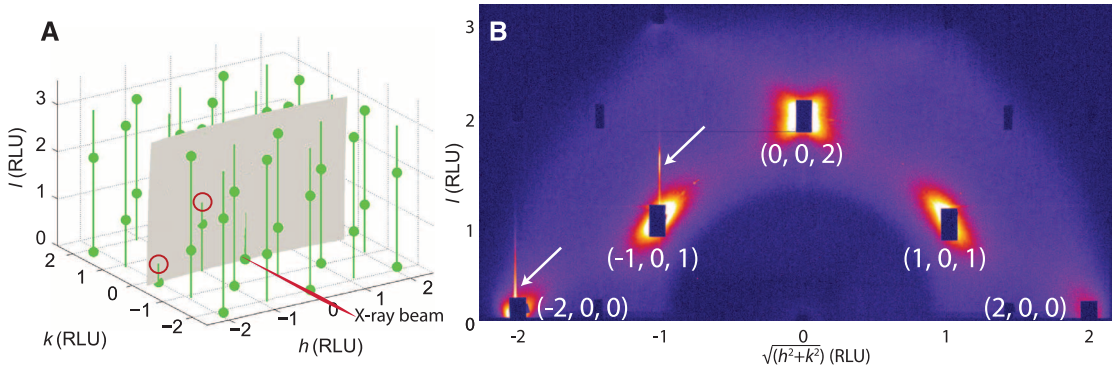


Fig. 1. Interpretation of the HESXRD patterns. (A) Illustration of how CTRs from a clean Pd(100) surface cross the Ewald sphere during sample rotation (see also movie S1). *h*, *k*, and *l* are the reciprocal lattice vectors defined such that *h* and *k* are in the surface plane, *l* is perpendicular to the surface plane, and the Bragg reflections of the substrate appear at integer values. **(B)** The corresponding detector image in which the Bragg reflections from the Pd substrate are indexed, and the points where CTRs cross the Ewald sphere are marked with arrows. The dark rectangles at the Pd Bragg reflections stem from absorbers protecting the detector. The reciprocal lattice vectors are defined as $|\mathbf{a}^*| = |\mathbf{b}^*| = 2\pi/d_{110}$; $|\mathbf{c}^*| = 2\pi/a_0$ with $a_0 = 3.89$ Å (Pd bulk lattice constant); $d_{110} = a_0/\sqrt{2}$. **a*** points along the face-centered cubic bulk [011] direction, **b*** along [0,–1,1], **c*** along [100], perpendicular to the surface. **a***, **b***, and **c*** are the basis vectors spanning the reciprocal lattice, i.e., defining the direction and units of the h, k, and l axes.

x-ray scattering intensity corresponding to the superlattice rods contains no Bragg reflections, indicating that the oxide film is very thin.

To allow comparison of the results obtained by HESXRD with those expected from LEED, which is the most common surface diffraction method used under UHV conditions, we show a slice in the *hk*-plane of the HESXRD data around $l = 0.5$ reciprocal lattice unit (RLU) (Fig. 2C). Projected onto the data is a map of the reciprocal lattice corresponding to the Pd(100) substrate (green squares) and the surface oxide (red dots and circles). A zoom-in (Fig. 2D) shows the spots at $(h, k) = (0.4, -0.8)$, $(0.6, -0.8)$, $(0.8, -0.6)$, and $(0.8, -0.4)$, corresponding to the periodicity of the $\sqrt{5}$ structure. It is, however, apparent that these spots are not perfectly in the $\sqrt{5}$ periodicity but that there is a stress-induced mismatch between the oxide and the substrate, as has previously been reported (*13*). This mismatch is readily visible in HESXRD, but it is not possible to infer its presence with LEED because of resolution limitations.

Atomic positions on the surface can be determined from the intensity variations along the CTRs and superstructure rods. For this, the intensities are translated into structure factors, which, after taking standard correction factors into account (*10*), are proportional to the square root of the diffraction intensity. These are then compared to calculated structure factors. Here, we used the atomic positions determined from quantitative LEED and density functional theory (*13*) and calculated corresponding structure factors using the software package ANA-ROD (*14*). We found good agreement between the experimental and calculated structure factors (Fig. 2, E and F, for one CTR and one superstructure rod and fig. S6 for all available rods). In addition to establishing that our HESXRD method is suitable for structural determination, the good agreement between the experimental and calculated structure factors also shows that the structure found in our study is similar to the one revealed by LEED. However, our data were recorded during steady-state CO oxidation conditions, whereas the structure investigated with LEED was prepared in pure O_2 under UHV conditions (*13*).

In addition to facilitating the collection of data for structural determinations, probing a larger part of reciprocal space simultaneously makes HESXRD an ideal tool for time-resolved measurements. We illustrate these capabilities by following how the surface structure of Pd(100), during catalytic CO oxidation, responds to a change in the gas stoichiometry (*15*). At a constant sample temperature of 575 K, we changed from a gas mixture of 6 Torr CO and 1.5 Torr O_2 (with Ar as carrier gas) to 6 Torr CO and 3 Torr O_2 (i.e., the same mixture as used above) and monitored the partial pressures of the reactant and product gases with mass spectrometry (MS). The time dependence of the O_2, CO, and CO_2 signals is shown in Fig. 3A. The most important features are found at 100 s, when the O_2 pressure was increased, and after about 550 s, when there is a sudden increase in the CO_2 production. Diffraction data were collected continuously while keeping the sample at a fixed position set to reveal the presence and structure of any oxide on the surface. Diffraction images were recorded every 0.5 s, resulting in a movie of 2000 frames (movie S3). In Fig. 3C, we show three snapshots from the movie at the times indicated by the lines I to III in Fig. 3A, chosen such that the images are representative for the diffraction at the time (I) before, (II) during, and

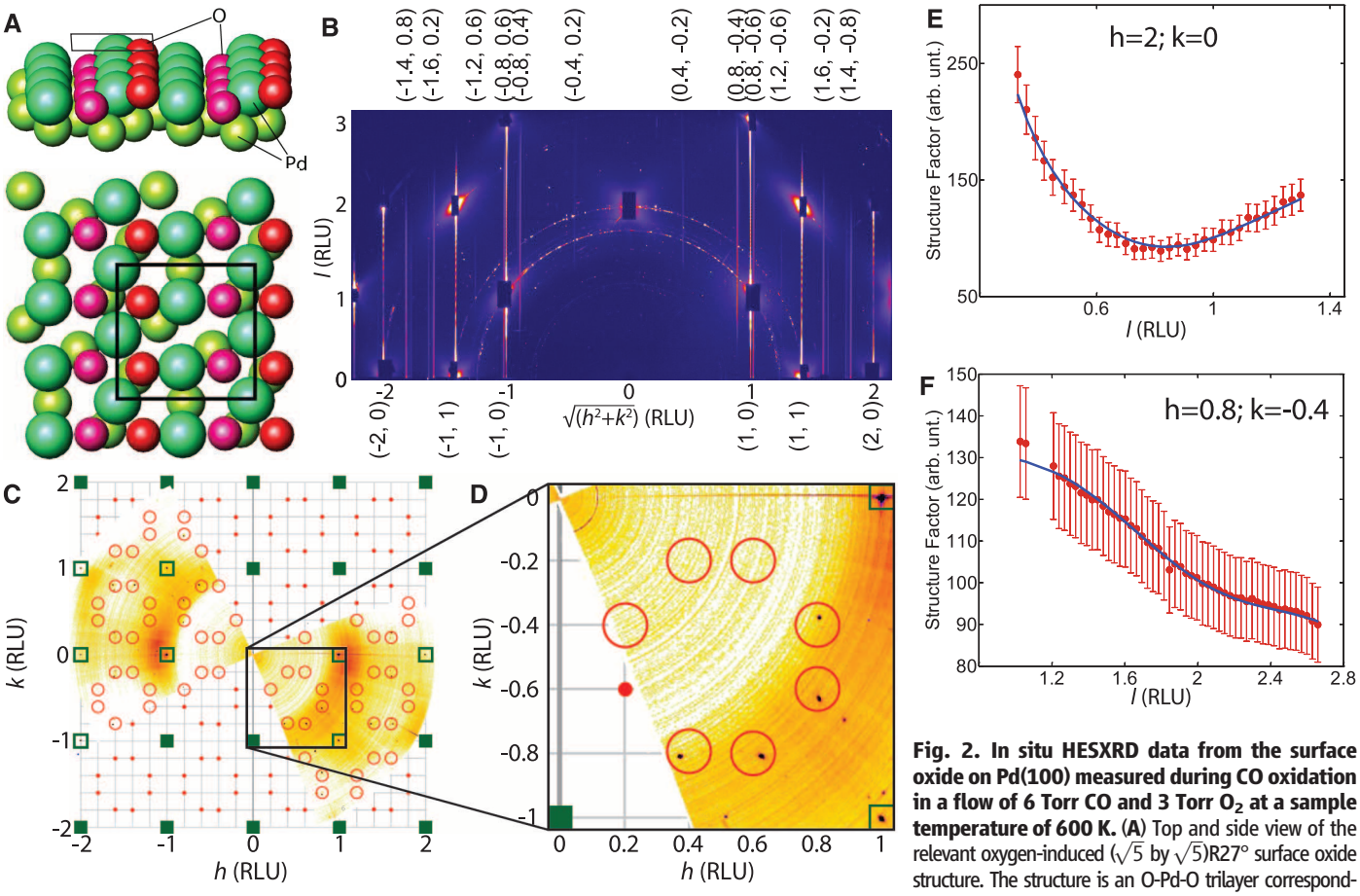


Fig. 2. In situ HESXRD data from the surface oxide on Pd(100) measured during CO oxidation in a flow of 6 Torr CO and 3 Torr O_2 at a sample temperature of 600 K. (**A**) Top and side view of the relevant oxygen-induced ($\sqrt{5}$ by $\sqrt{5}$)R27° surface oxide structure. The structure is an O-Pd-O trilayer corresponding to one PdO(101) plane (*12*, *13*). (**B**) All images collected during the rotational scan combined into a single image, in which the CTRs and superlattice rods are indicated. The way these images are combined enhances background noise, such as the powder diffraction rings originating from polycrystalline defects in the crystal. (**C**) In-plane view (*hk*-plane at $l = 0.5$) of the angular range measured in the current experiment. The Pd CTRs (squares) and surface oxide superlattice rods (circles) are indicated. (**D**) Magnification of reciprocal space showing the superlattice reflections. The reflections are not in the middle of the red circle, directly revealing the mismatch between the PdO(101) and the Pd(100) substrate. (**E** and **F**) Extracted (dots) CTRs and superlattice rods from the rotational images as seen in (B), and calculated structure factors (full lines). All the detector images are shown individually in movie S2.

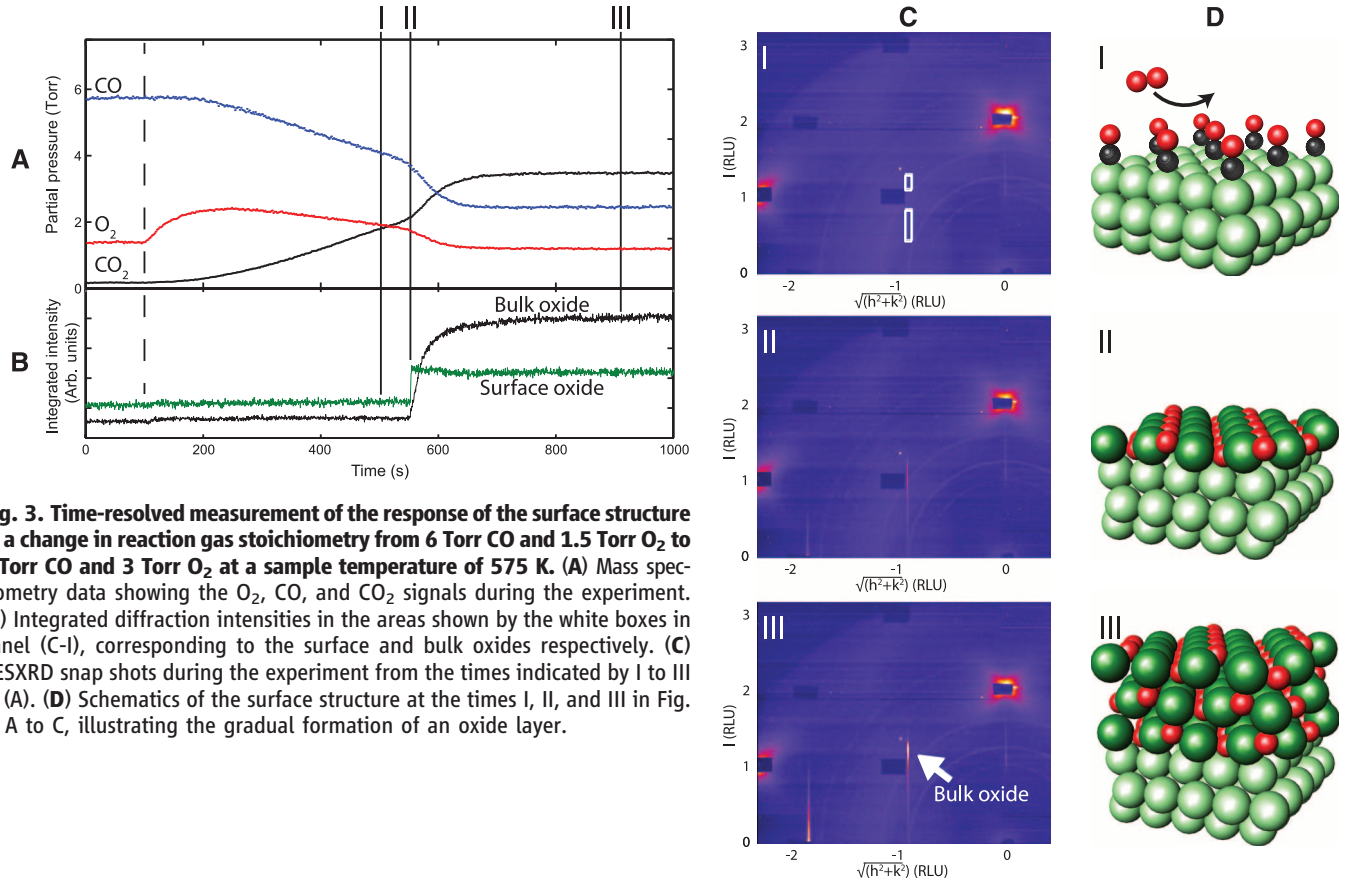


Fig. 3. Time-resolved measurement of the response of the surface structure to a change in reaction gas stoichiometry from 6 Torr CO and 1.5 Torr O_2 to 6 Torr CO and 3 Torr O_2 at a sample temperature of 575 K. (**A**) Mass spectrometry data showing the O_2, CO, and CO_2 signals during the experiment. (**B**) Integrated diffraction intensities in the areas shown by the white boxes in panel (C-I), corresponding to the surface and bulk oxides respectively. (**C**) HESXRD snap shots during the experiment from the times indicated by I to III in (A). (**D**) Schematics of the surface structure at the times I, II, and III in Fig. 3, A to C, illustrating the gradual formation of an oxide layer.

(III) after the increase in catalytic activity at around 550 s. At time I, no oxide can be detected, which means that the surface is metallic. The surface is at this stage predominantly covered by CO hindering the dissociative adsorption of O_2, and the reaction is so-called CO self-poisoned (see Fig. 3D) (*16*). At time II, at which the catalytic activity increases considerably, two rods corresponding to the $\sqrt{5}$ surface oxide appear. As the reaction proceeds, the oxide structure continues to develop, and, when the new steady state is reached at time III, there is an increased intensity at an elevated *l* value (shown by the white arrow in Fig. 3C), revealing the onset of the growth of a several-layer-thick PdO film with a (101)-oriented surface (*17*, *18*). To enable a more direct comparison between the MS and diffraction data, Fig. 3B shows the integrated intensities inside the white boxes shown in panel I of Fig. 3C. There is a strong correlation between the formation of the surface oxide and the increased activity, and the activity remains high during the development of the thicker oxide. Hence, this proof-of-principle study strongly indicates that the activity for CO oxidation is particularly high over the PdO(101) surface. Although our best fit indicates that the PdO(101) film covers close to 100% of the surface, we cannot exclude minor areas, such as defects, affecting the activity, a subject for intense discussion in recent literature (*19*, *20*).

Our results demonstrate how high-energy x-rays open up new opportunities in the use of surface x-ray diffraction. Thanks to the high energy, the scattering angles decrease and the diffraction pattern can be collected on a stationary 2D detector, which simplifies and speeds up the data acquisition process considerably. There are, however, some drawbacks with the high energy. The small critical angle for total reflection makes the sample alignment very sensitive, for instance, to changes of the sample temperature. This can be overcome with an automatic feedback system for sample alignment. Further, the scattered intensity, as well as the sensitivity of the detectors, drops with increasing photon energy. This is, to a large extent, made up for by the improved source and high-energy optics performance at third-generation synchrotron radiation sources. The high penetration ability at high energies also enables the use of complex sample environments, crucial for in situ experiments, and structure determination of buried or electrochemical solid-liquid interfaces, without relying on thin-film geometries.

References and Notes

1. A. Zangwill, *Physics at Surfaces* (Cambridge Univ. Press, Cambridge, 1988).
2. H. Lüth, *Surfaces and Interfaces of Solids* (Springer, Berlin, 1993).
3. G. Ertl, *Reactions at Solid Surfaces* (Wiley, Hoboken, NJ, 2010).
4. G. A. Somorjai, Y. Li, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 917–924 (2011).
5. J. H. Sinfelt, *Surf. Sci.* **500**, 923–946 (2002).
6. J. G. Ekerdt, Y. M. Sun, A. Szabo, G. J. Szulczewski, J. M. White, *Chem. Rev.* **96**, 1499–1518 (1996).
7. I. K. Robinson, *Phys. Rev. Lett.* **50**, 1145–1148 (1983).
8. R. Feidenhans'l, *Surf. Sci. Rep.* **10**, 105–188 (1989).
9. H. Dosch *et al.*, *Phys. Rev. Lett.* **60**, 2382–2385 (1988).
10. Materials and methods are available as supplementary materials on *Science* Online.
11. R. van Rijn *et al.*, *Rev. Sci. Instrum.* **81**, 014101 (2010).
12. M. Todorova *et al.*, *Surf. Sci.* **541**, 101–112 (2003).
13. P. Kostelnik *et al.*, *Surf. Sci.* **601**, 1574–1581 (2007).
14. E. Vlieg, *J. Appl. Cryst.* **33**, 401–405 (2000).
15. R. van Rijn *et al.*, *Phys. Chem. Chem. Phys.* **13**, 13167–13171 (2011).
16. S. Blomberg *et al.*, *Phys. Rev. Lett.* **110**, 117601 (2013).
17. N. Seriani, J. Harl, F. Mittendorfer, G. Kresse, *J. Chem. Phys.* **131**, 054701 (2009).
18. H. H. Kan, J. F. Weaver, *Surf. Sci.* **602**, L53–L57 (2008).
19. F. Gao, Y. Wang, D. W. Goodman, *J. Phys. Chem. C* **114**, 6874 (2010).
20. R. van Rijn *et al.*, *J. Phys. Chem. C* **114**, 6875–6876 (2010).

Acknowledgments: This work is done within the Röntgen-Ångström collaboration "Catalysis on the atomic scale". The authors thank the Swedish Research Council, the Swedish Foundation for Strategic Research, the Crafoord Foundation, the Knut and Alice Wallenberg Foundation, the Anna and Edwin Berger Foundation, and the German Federal Ministry of Education and Research (BMBF) (project 05K10PS1 NanoXcat) for financial support. The raw diffraction data are available in databases S1 to S3.

Supplementary Materials

www.sciencemag.org/content/343/6172/758/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Table S1
Movies S1 to S3
Databases S1 to S3
References (*21*–*24*)

4 October 2013; accepted 6 January 2014
Published online 30 January 2014;
10.1126/science.1246834

The New Madrid Seismic Zone: Not Dead Yet

Morgan T. Page* and Susan E. Hough

The extent to which ongoing seismicity in intraplate regions represents long-lived aftershock activity is unclear. We examined historical and instrumental seismicity in the New Madrid central U.S. region to determine whether present-day seismicity is composed predominantly of aftershocks of the 1811–1812 earthquake sequence. High aftershock productivity is required both to match the observation of multiple mainshocks and to explain the modern level of activity as aftershocks; synthetic sequences consistent with these observations substantially overpredict the number of events of magnitude ≥ 6 that were observed in the past 200 years. Our results imply that ongoing background seismicity in the New Madrid region is driven by ongoing strain accrual processes and that, despite low deformation rates, seismic activity in the zone is not decaying with time.

Seismic hazard is not isolated to tectonic plate boundaries, as evidenced by earthquakes that occur in stable continental regions. Intraplate earthquakes, which are related to the internal deformation of plates rather than motion at plate boundaries, can be large and damaging, as with the 2001 Bhuj earthquake (*1*). In this work, we study the 1811–1812 New Madrid sequence, which is of paramount importance for understanding intraplate seismogenesis and for probabilistic seismic hazard assessment in the central and eastern United States and other midcontinental regions. The sequence included four events that were widely felt throughout the central and eastern United States, conventionally regarded as three primary mainshocks and the large dawn aftershock following the first mainshock. Magnitude estimates for these events have varied widely, from a low of magnitude (M) ≈ 7 for the largest mainshocks (*2*) to values over 8 in magnitude (*3*).

Aftershocks of the 1811–1812 sequence have been considered in two ways. Several studies have used archival accounts of large aftershocks and/or tallies of felt earthquakes to estimate magnitudes for large aftershocks and consider the overall magnitude distribution of early aftershocks [e.g., (*4*, *5*)]. Two studies have considered the long-term rate of seismicity in the New Madrid Seismic Zone (NMSZ) and concluded that it is well characterized as a long-lived aftershock sequence (*6*, *7*). It is important to note, however, that these latter two studies do not show a fit, from 1811 to present, to traditional Omori decay (*8*, *9*). Such direct evidence has been observed for the classic long-lived aftershock sequence following the 1891 Nobi earthquake, for which an Omori decay can be seen for 100 years (*10*). In the New Madrid case, however, a direct fit is not possible given uncertainties in the early New Madrid catalog. In this study, we reconsider the long-lived aftershock hypothesis using rigorous tests assuming an Epidemic Type Aftershock Sequence (ETAS) model (*11*). ETAS modeling allows us to determine probabilities of observing robust features of the New Madrid catalog, should the long-lived aftershock hypothesis be true.

The ETAS model, developed on the premise that all earthquakes potentially trigger their own aftershocks, successfully explains the empirical Omori decay law, which, so far as is known, universally describes the temporal decay of aftershocks. The ETAS model explains observed foreshock rates and multiplets (*12*) and has been shown to accurately characterize seismicity, including both short- and long-term aftershock sequences [e.g., (*13*)], and is now a widely used short-term earthquake clustering model (*14*). The model has been used to characterize and forecast seismicity rates in a wide range of tectonic environments, including intraplate regions and regions characterized by swarmy activity (*15*, *16*). In this work, we use ETAS modeling in an attempt to generate synthetic catalogs that match well-constrained features of the New Madrid earthquake sequence (see materials and methods in the supplementary materials).

To test the long-lived aftershock hypothesis, we identified three robust observational constraints that are not dependent on particular contentious magnitude values. Our first imposed constraint

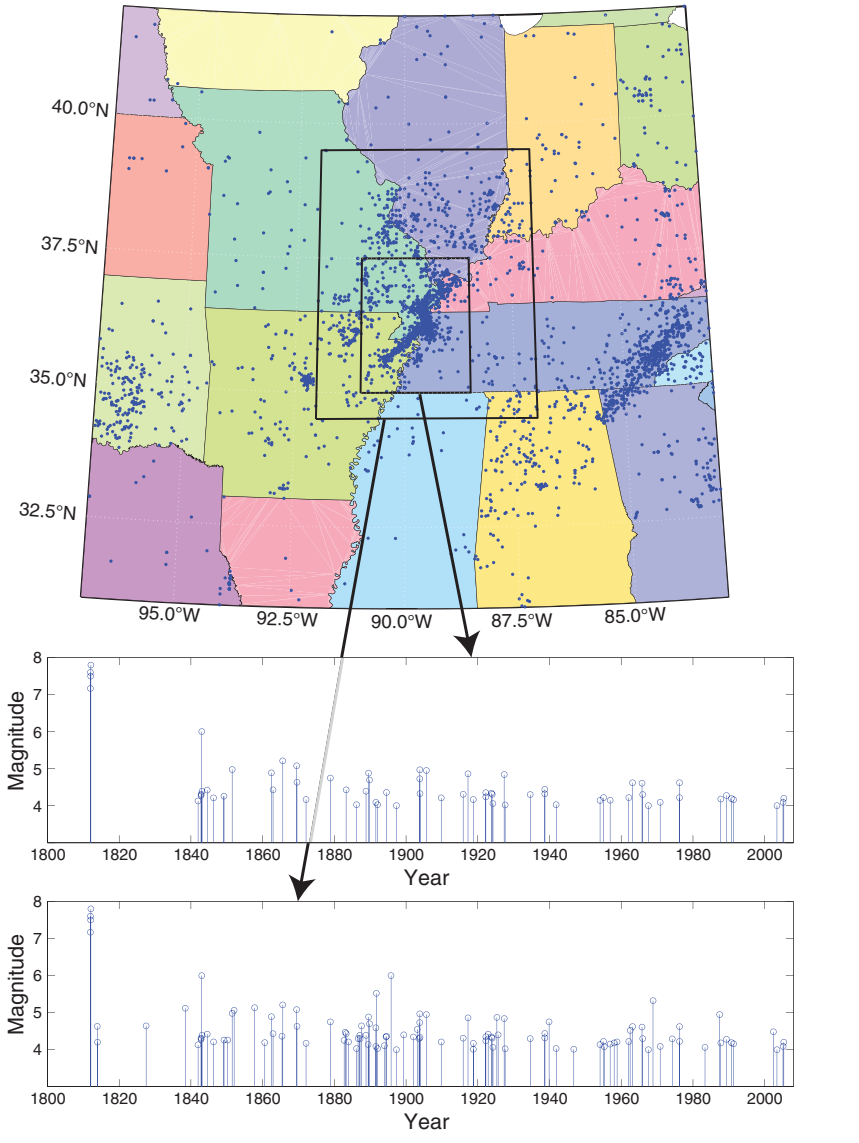


Fig. 1. Seismicity in the New Madrid region (CEUS catalog, 1800–2008, $M \geq 4$). Note that the early catalog is not complete to $M4$.

U.S. Geological Survey, Pasadena, CA, USA.
*Corresponding author. E-mail: pagem@caltech.edu

is that the sequence included four principal events of comparable magnitude, separated by no more than 0.7 magnitude units. This is based on the range in event magnitudes inferred by different studies (*2*, *3*, *17*). Although the absolute magnitudes of these earthquakes remain a subject for debate, the relative magnitudes are much more reliably determined. Analysis of prehistoric sandblows in the NMSZ shows that protracted sequences, with multiple large mainshocks, are apparently the norm for this region (*18*).

The second constraint is on the recent rate of moderate-sized ($M \geq 4$) earthquakes. Because using different catalogs and box sizes produce different estimates, we used the most conservative estimate of three $M \geq 4$ earthquakes over 10 years (Fig. 1), taken from the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) catalog (*19*) (see materials and methods).

The third constraint is the number of moderate ($M \geq 6$) events in the NMSZ after the initial cluster in the first year. The CEUS-SSC catalog (*19*) includes two such events, the 1843 Marked Tree, Arkansas, and 1895 Charleston, Missouri, earthquakes, both with preferred magnitudes of 6.0. Although a recent reinterpretation of macroseismic effects of the 1843 earthquake (*20*) estimates a lower preferred magnitude of 5.4, we assume, for conservatism, that the sequence produced no more than two $M \geq 6$ late events (see materials and methods).

We generated synthetic ETAS catalogs, searching for a single set of subcritical, direct Omori parameters that matched the three robust observational constraints described above. The fraction of stochastic catalogs that are consistent with both early clustering behavior and recent seismicity in the New Madrid region are shown in Fig. 2, A and B, respectively. These two constraints reduce the possible ETAS phase space to a small region (Fig. 2C). Synthetic catalogs produced in this region of the ETAS phase space are very productive both early and late in the sequence. We find that synthetic sequences that are active enough to match observed New Madrid–style early clustering behavior and current seismicity rates contain many more $M \geq 6$ events at intermediate times than have been observed (table S1). At 95% confidence, no set of direct Omori parameters is consistent with all three of our constraints: early

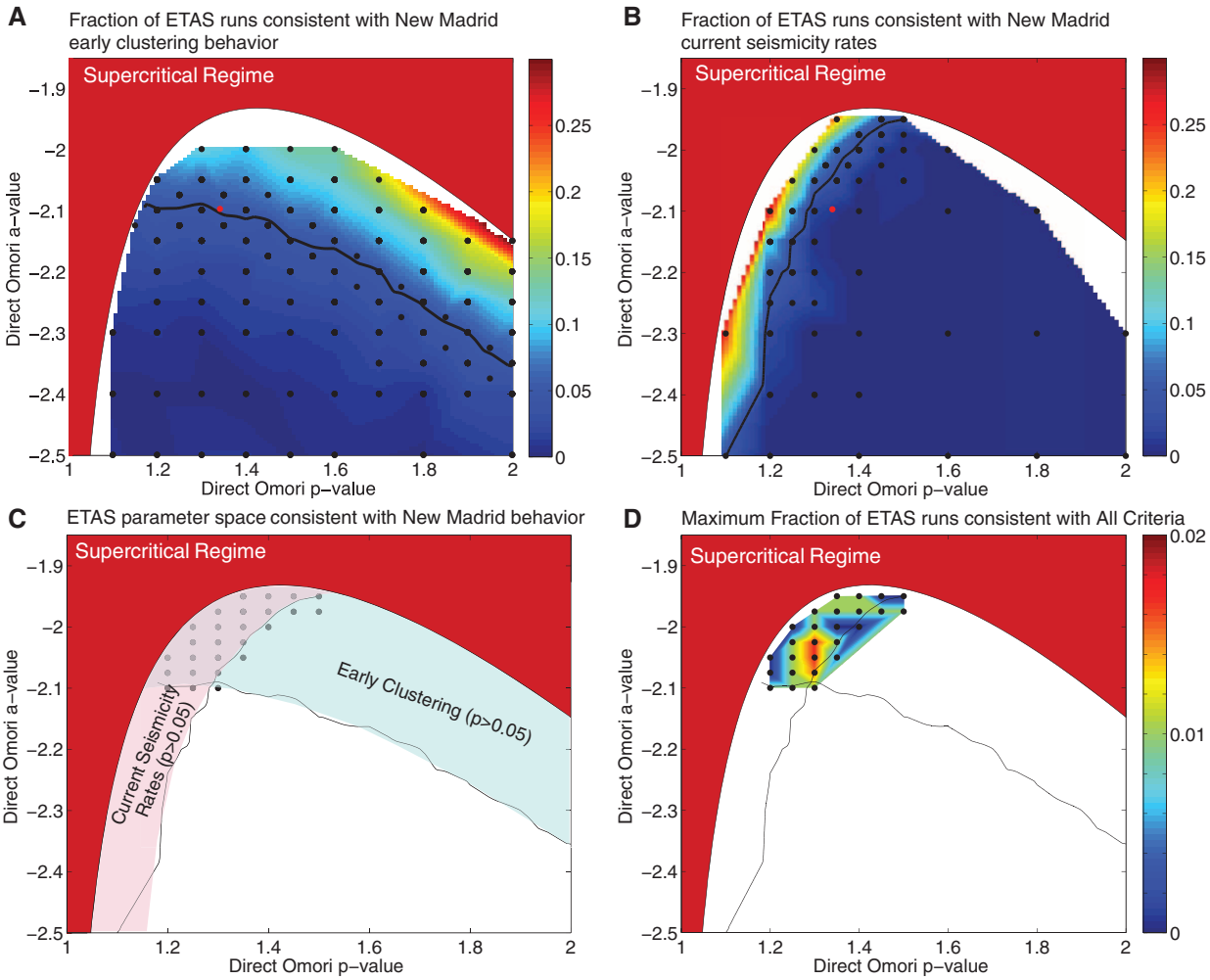


Fig. 2. Regions of ETAS parameter space consistent with New Madrid behavior. The unphysical, supercritical regime (see materials and methods) is shown in red. (**A**) ETAS simulations within the subcritical regime are sampled at the black points; colors show a linear interpolation of the fraction of synthetic sequences for which the four largest shocks in the first 2 months are within 0.7 magnitude units of each other, as was seen in the New Madrid sequence. Above the black line (which theoretically is smooth but has small irregularities due to sampling error), at least 5% of synthetic sequences are consistent with New Madrid clustering behavior; below this line, the early behavior is less productive than observations. The red dot shows average California parameters (*25*) for reference. (**B**) The fraction of synthetic sequences that have a late (200 years post-mainshock) aftershock rate that matches current New Madrid seismicity rates. (**C**) The parameter space consistent with both early clustering and current seismicity rates is confined to a small region; we sample sequences at the points shown and find that sequences with parameters in this region typically produce a much higher rate of *M*6 earthquakes after the first year than that observed. (**D**) The maximum fraction, over all mainshock magnitudes, that is consistent with early clustering, current seismicity rates, and the rate of $M \geq 6$ earthquakes after the first year, linearly interpolated between sampling points. Although some variation in this plot is due to sampling error, all points have been sampled sufficiently to determine that the fraction is less than 5%, at 95% confidence (see table S1).

clustering, current seismicity rates, and the rate of $M \geq 6$ events after the first year (Fig. 2D). Among sequences sampled that were consistent with New Madrid early clustering behavior and current seismicity rates, the mean number of $M \geq 6$ earthquakes from 1 year to 200 years post-mainshock was 135. At best, at some points in ETAS phase space ~1.7% of the sequences are consistent with our criteria. Results using a stricter criteria that includes the observation that no $M \geq 6$ earthquakes occurred in the region in the past 100 years (table S1) show that we can reject the long-lived aftershock hypothesis at even higher confidence.

Based on our statistical analysis, the hypothesis that current seismicity in the New Madrid region is primarily composed of aftershocks from the 1811–1812 sequence fails. This is because a sequence active enough at late times to produce the seismicity rates observed today and active enough at early times to produce the short-term clustering observed in the first few months would be highly likely to produce too many aftershocks in the intermediate times. If current seismicity in the New Madrid region is not composed predominantly of aftershocks, there must be continuing strain accrual. This is in agreement with recent work finding nonzero strain measurements in the region that are consistent with ongoing interseismic slip of about 4 mm/year (*21*), in contrast to earlier studies [e.g., (*22*)]. The spatial distribution of the stress pattern driven by this model would be generally consistent with the stress change caused by an earthquake on the Reelfoot fault. This could explain how ongoing microseismicity is not part of an aftershock sequence but is still consistent with the predicted stress change associated with the 1811–1812 sequence (*23*). If ongoing microseismicity does result from ongoing strain accrual, this suggests that the region, along with the neighboring Wabash Valley where nonzero strain has also been observed (*24*), will continue to be a source of hazard.

References and Notes

1. H. K. Gupta, N. P. Rao, B. K. Rastogi, D. Sarkar, *Science* **291**, 2101–2102 (2001).
2. S. E. Hough, M. Page, *J. Geophys. Res.* **116**, (B3), B03311 (2011).
3. A. C. Johnston, *Geophys. J. Int.* **126**, 314–344 (1996).
4. O. W. Nuttli, *Bull. Seismol. Soc. Am.* **63**, 227 (1973).
5. S. E. Hough, *Seismol. Res. Lett.* **80**, 1045–1053 (2009).
6. J. E. Ebel, K.-P. Bonjer, M. C. Oncescu, *Seismol. Res. Lett.* **71**, 283–294 (2000).
7. S. Stein, M. Liu, *Nature* **462**, 87–89 (2009).
8. F. Omori, *J. Coll. Sci. Imp. Univ. Tokyo* **7**, 111 (1895).
9. T. Utsu, *Geophys. Mag.* **30**, 521 (1961).
10. T. Utsu, Y. Ogata, R. S. Matsu'ura, *J. Phys. Earth* **43**, 1–33 (1995).
11. Y. Ogata, *J. Am. Stat. Assoc.* **83**, 9–27 (1988).
12. K. R. Felzer, R. E. Abercrombie, G. Ekstršm, *Bull. Seismol. Soc. Am.* **94**, 88–98 (2004).
13. Y. Ogata, *J. Geophys. Res.* **97**, (B13), 19845 (1992).
14. M. C. Gerstenberger, D. A. Rhoades, *Pure Appl. Geophys.* **167**, 877–892 (2010).
15. Y. Ogata, J. Zhuang, *Tectonophysics* **413**, 13–23 (2006).
16. A. L. Llenos, J. J. McGuire, Y. Ogata, *Earth Planet. Sci. Lett.* **281**, 59–69 (2009).
17. W. H. Bakun, M. G. Hopper, *Bull. Seismol. Soc. Am.* **94**, 64–75 (2004).
18. M. P. Tuttle, *Bull. Seismol. Soc. Am.* **92**, 2080–2089 (2002).
19. K. J. Coppersmith *et al*., Central and Eastern United States Seismic Source Characterization for Nuclear Facilities Project, Technical Report (Electric Power Research Institute, Palo Alto, CA, 2012).
20. S. E. Hough, *Bull. Seismol. Soc. Am.* **103**, 2767 (2013).
21. A. Frankel, R. Smalley, J. Paul, *Bull. Seismol. Soc. Am.* **102**, 479–489 (2012).
22. E. Calais, J. Y. Han, C. DeMets, J. M. Nocquet, *J. Geophys. Res.* **111**, (B6), 6402 (2006).
23. K. Mueller, S. E. Hough, R. Bilham, *Nature* **429**, 284–288 (2004).
24. G. A. Galgana, M. W. Hamburger, *Seismol. Res. Lett.* **81**, 699–714 (2010).
25. K. R. Felzer, R. E. Abercrombie, G. Ekström, *Bull. Seismol. Soc. Am.* **93**, 1433–1448 (2003).

Acknowledgments: We thank J. Hardebeck, C. Mueller, and two anonymous reviewers for comments on the manuscript. The CEUS-SSC catalog is available at: www.ceus-ssc.com/Report/Downloads.html. Author Contributions: M.T.P. did the ETAS modeling, and S.E.H. provided expertise on the historical catalog. Both authors participated in the writing.

Supplementary Materials

www.sciencemag.org/content/343/6172/762/suppl/DC1
Materials and Methods
Fig. S1
Table S1
References (*26*–*30*)

7 November 2013; accepted 15 January 2014
Published online 23 January 2014;
10.1126/science.1248215

Evolutionarily Dynamic Alternative Splicing of *GPR56* Regulates Regional Cerebral Cortical Patterning

Byoung-Il Bae,[1]* Ian Tietjen,[1]*† Kutay D. Atabay,[1] Gilad D. Evrony,[1] Matthew B. Johnson,[1] Ebenezer Asare,[1] Peter P. Wang,[1] Ayako Y. Murayama,[2] Kiho Im,[3] Steven N. Lisgo,[4] Lynne Overman,[4] Nenad Šestan,[5] Bernard S. Chang,[6] A. James Barkovich,[7] P. Ellen Grant,[3] Meral Topçu,[8] Jeffrey Politsky,[9]‡ Hideyuki Okano,[2] Xianhua Piao,[10] Christopher A. Walsh[1]§

The human neocortex has numerous specialized functional areas whose formation is poorly understood. Here, we describe a 15–base pair deletion mutation in a regulatory element of *GPR56* that selectively disrupts human cortex surrounding the Sylvian fissure bilaterally including "Broca's area," the primary language area, by disrupting regional *GPR56* expression and blocking RFX transcription factor binding. *GPR56* encodes a heterotrimeric guanine nucleotide–binding protein (G protein)–coupled receptor required for normal cortical development and is expressed in cortical progenitor cells. GPR56 expression levels regulate progenitor proliferation. *GPR56* splice forms are highly variable between mice and humans, and the regulatory element of gyrencephalic mammals directs restricted lateral cortical expression. Our data reveal a mechanism by which control of *GPR56* expression pattern by multiple alternative promoters can influence stem cell proliferation, gyral patterning, and, potentially, neocortex evolution.

Although most mammals have elaborate and species-specific patterns of folds ("gyri") in the neocortex, the genetic and evolutionary mechanisms of cortical gyrification are poorly understood (*1*–*3*). Abnormal gyrification, such as polymicrogyria (too many small gyri), invariably signals abnormal cortical development, so regional disorders of gyrification are of particular interest, because they highlight mechanisms specific to cortical regions. The human cortex contains dozens of cortical regions specialized for distinct functions—such as language, hearing, and sensation—yet it is unsolved how these cortical regions form and how human cortical regions evolved from those of prehuman ancestors.

Examination of >1000 individuals with gyral abnormalities identified five individuals from three

[1]Division of Genetics and Genomics, Manton Center for Orphan Disease, and Howard Hughes Medical Institute, Boston Children's Hospital, Broad Institute of MIT and Harvard, and Departments of Pediatrics and Neurology, Harvard Medical School, Boston, MA 02115, USA. [2]Department of Physiology, Keio University School of Medicine, Tokyo 160-8582, Japan. [3]Division of Newborn Medicine, Center for Fetal Neonatal Neuroimaging and Developmental Science, Department of Radiology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA. [4]The MRC-Wellcome Trust Human Developmental Biology Resource (HDBR), Newcastle, Institute of Genetic Medicine, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK. [5]Department of Neurobiology and Kavli Institute of Neuroscience, Yale University School of Medicine, New Haven, CT 06520, USA. [6]Beth Israel Deaconess Medical Center, Comprehensive Epilepsy Center, Boston, MA 02215, USA. [7]Departments of Radiology, Pediatrics, Neurology, and Neurological Surgery, University of California San Francisco, San Francisco, CA 94143, USA. [8]Department of Pediatrics, Hacettepe University Faculty of Medicine, Ankara, Turkey. [9]Department of Neurology, Medical College of Georgia, Augusta, GA 30912, USA. [10]Division of Newborn Medicine, Boston Children's Hospital and Harvard Medical School, Boston, MA 02115, USA.

*The authors contributed equally to this work.
†Present address: Department of Anesthesiology, Pharmacology and Therapeutics, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.
‡Present address: Northeast Regional Epilepsy Group, Atlantic Neuroscience Institute Epilepsy Center, Summit, NJ 07901, USA.
§Corresponding author. E-mail: christopher.walsh@childrens.harvard.edu

families (one Turkish and two Irish-American) with strikingly restricted polymicrogyria limited to the cortex surrounding the Sylvian fissure (Fig. 1, A and B; fig. S1; and movies S1 and S2), which suggests a rare, but genetically distinctive, condition. Affected individuals suffered intellectual and language difficulty, as well as refractory seizures (onset 7 months to 10 years), but had no motor disability (table S1). Magnetic resonance imaging (MRI) and quantitative gyral analysis showed abnormal inferior and middle gyri in prefrontal and motor cortex, with mildly affected temporal lobes. Broca's area—the "motor center for speech" (*4*)—in the left hemisphere and the corresponding areas of the right hemisphere were most severely affected. Affected neocortical surface showed abnormally numerous, small gyral-like folds that fused in coarse, irregular patterns, with abnormal and highly irregular white matter protrusions, consistent with polymicrogyria (*5*, *6*), along with widening of the Sylvian fissure (Fig. 1A and fig. S1B).

Genome-wide analysis identified a single linked locus on chromosome 16q12.2-21 (Fig. 1C) containing the *GPR56* gene, which, when mutated in its coding region, leads to polymicrogyria of the entire neocortex, as well as cerebellar and white matter abnormalities (*7*–*9*). As we found no mutations in the exons of *GPR56*, we sequenced 38 conserved non-exonic elements (table S2), in one of which we identified a small deletion in all five individuals. The mutated element normally contains two copies of a 15–base pair (bp) tandem repeat, but all affected individuals have a homozygous deletion of one 15-bp repeat (Fig. 1, E and F). The deletion is heterozygous in parents of the affected individuals, who manifest no obvious clinical signs, and is absent from thousands of control chromosomes in the Single-Nucleotide Polymorphism Database and 1000 Genomes database. The two Irish-American families carry the mutation on the same chromosomal haplotype, which reflects a common founder. It is noteworthy that the Turkish family carries the same deletion on a distinct haplotype, which indicates that the mutation arose independently (Fig. 1D). The element is located <150 bp upstream of the transcriptional start site of noncoding exon 1m (e1m) of *GPR56*, which suggests that it may regulate e1m expression as a cis-regulatory element. *GPR56* has at least 17 alternative transcription start sites, each beginning from a different noncoding first exon; all of the start sites are predicted to drive transcription of mRNAs whose coding sequence starts from exon 3 (Fig. 2A and fig. S2A) and all of which encode the same GPR56 protein (*10*, *11*). The diverse noncoding first exons have distinct expression profiles, with e1m being the most robustly transcribed first exon in fetal human brain but with several other alternative transcripts also expressed in fetal and adult brain (Fig. 2A and fig. S2, B to D).

To confirm that the 15-bp deletion disrupts perisylvian *GPR56* expression, we generated transgenic mice with the 23-kb human *GPR56* upstream region driving green fluorescent protein (GFP) expression. The 23-kb region encompasses 16 of the 17 transcription start sites containing e1m and ends before the translation start codon (Fig. 2A). This construct drives GFP expression in the entire central nervous system, including neocortex, and recapitulates the location and relative amount of expression of endogenous mouse GPR56 protein (Fig. 2B and fig. S3). In contrast, the 23-kb construct containing the 15-bp deletion drives expression in medial, but not lateral, cortex or lateral ganglionic eminence (Fig. 2B). These data suggest that the cis-regulatory element upstream of e1m drives *GPR56* expression in the perisylvian and lateral cortex, whereas disruption of the element, with consequent impairment of e1m expression, causes the perisylvian malformation.

To elucidate how the 15-bp deletion in the cis-regulatory element disrupts e1m expression,

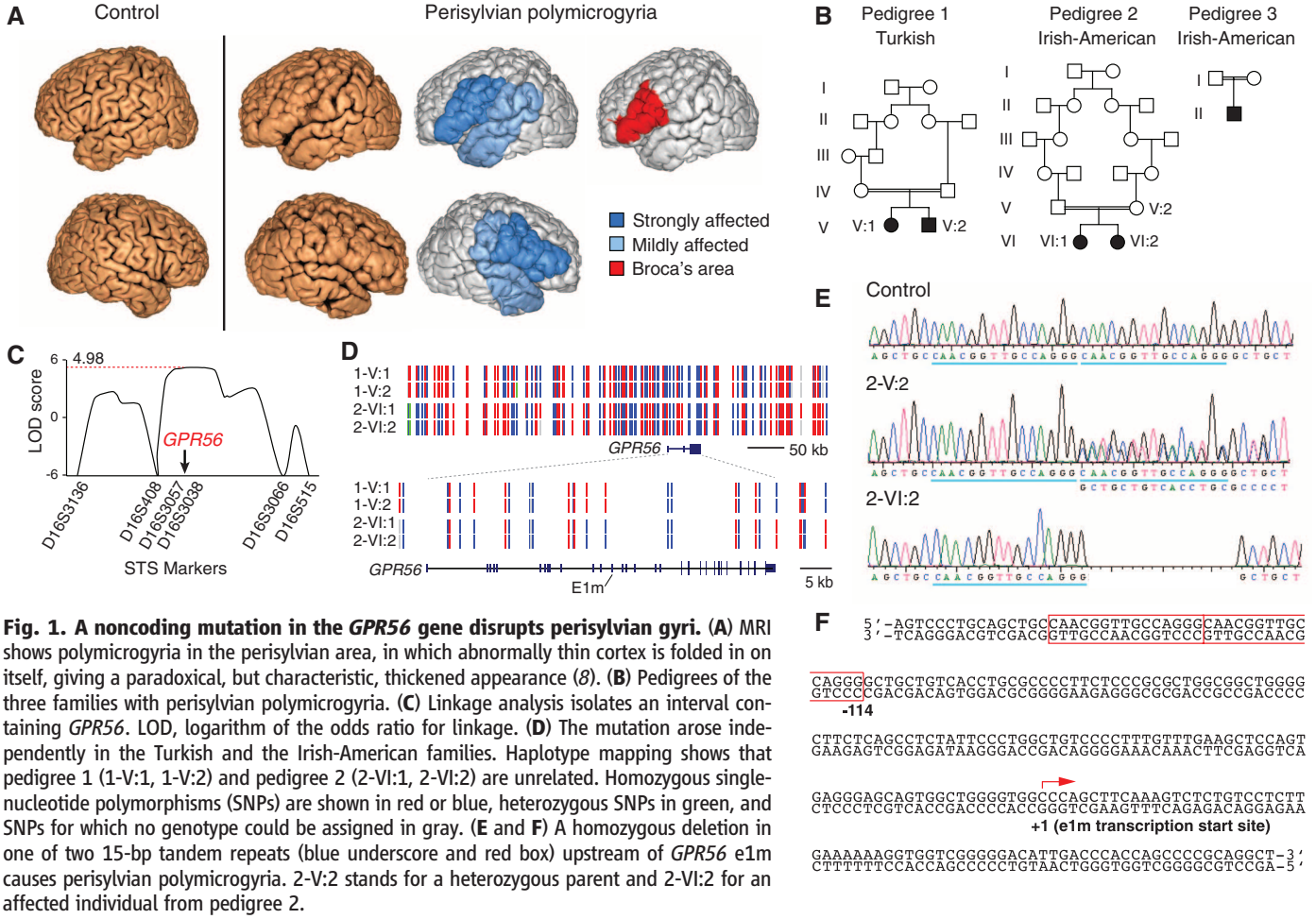


Fig. 1. A noncoding mutation in the *GPR56* gene disrupts perisylvian gyri. (**A**) MRI shows polymicrogyria in the perisylvian area, in which abnormally thin cortex is folded in on itself, giving a paradoxical, but characteristic, thickened appearance (*8*). (**B**) Pedigrees of the three families with perisylvian polymicrogyria. (**C**) Linkage analysis isolates an interval containing *GPR56*. LOD, logarithm of the odds ratio for linkage. (**D**) The mutation arose independently in the Turkish and the Irish-American families. Haplotype mapping shows that pedigree 1 (1-V:1, 1-V:2) and pedigree 2 (2-VI:1, 2-VI:2) are unrelated. Homozygous single-nucleotide polymorphisms (SNPs) are shown in red or blue, heterozygous SNPs in green, and SNPs for which no genotype could be assigned in gray. (**E** and **F**) A homozygous deletion in one of two 15-bp tandem repeats (blue underscore and red box) upstream of *GPR56* e1m causes perisylvian polymicrogyria. 2-V:2 stands for a heterozygous parent and 2-VI:2 for an affected individual from pedigree 2.

we performed yeast one-hybrid (Y1H) screening of a mouse forebrain cDNA library with the human cis-regulatory element as bait and obtained multiple yeast colonies encoding members of the *regulatory factor X (Rfx)* transcription factor family (Fig. 2C) (*12*). RFX1 and RFX3 bind the normal element in vitro, with binding decreased 60 to 70% by the 15-bp deletion (Fig. 2D). Chromatin immunoprecipitation sequencing confirmed RFX3 binding to the element (fig. S4) (*13*). RFX1 and GPR56 colocalize in germinal zones of fetal human brain (Fig. 2E). Dominant-negative RFX abrogates normal, but not mutant, e1m promoter activity on embryonic day 13.5 (E13.5) in mouse cortical cultures (Fig. 2F). Furthermore, genetic ablation of *Rfx4* decreases *Gpr56* expression in developing mouse brain (*14*). *RFX* and *GPR56* expression patterns are correlated (fig. S5, A and B) (*15*), with *RFX3* and *RFX7* most prominent in human ventrolateral prefrontal cortex, the region affected in perisylvian polymicrogyria (Fig. 2G), which suggests that multiple RFX proteins regulate the element.

GPR56 encodes an adhesion heterotrimeric guanine nucleotide–binding protein (G protein)–coupled receptor that is highly expressed in cortical progenitors (*7*, *16*) and binds extracellular matrix proteins (*17*). Loss of *GPR56* disrupts radial glia and causes breaches in the pial basement membrane, through which some neurons overmigrate (*9*, *16*). However, even where the pia is intact, we found that neocortical thickness and organization are irregular, with occasional thin regions in *Gpr56* knockout mice (Fig. 3A). Postmortem analysis of a human with biallelic *GPR56* coding mutations showed a very thin cortex, which suggested potential roles of *GPR56* in neurogenesis as well (*9*). GPR56 protein is most highly expressed in progenitors in the ventricular and subventricular zones during neurogenesis in mice (*16*, *18*). GPR56 expression in developing human and marmoset neocortex is highest in the ventricular zone, as well as in the outer subventricular zone, which is expanded in mammals with larger brains (*2*) (Fig. 3B and fig. S5, C and D).

Impairment and overexpression of *GPR56* show that its expression regulates proliferation. *Gpr56* knockout mice show fewer phosphohistone H3 (PH3)–positive mitotic progenitor cells and TBR2-positive intermediate progenitors than wild-type mice in the neocortex at E14.5. Conversely, mice carrying a transgene that directs overexpression of human *GPR56* show increased

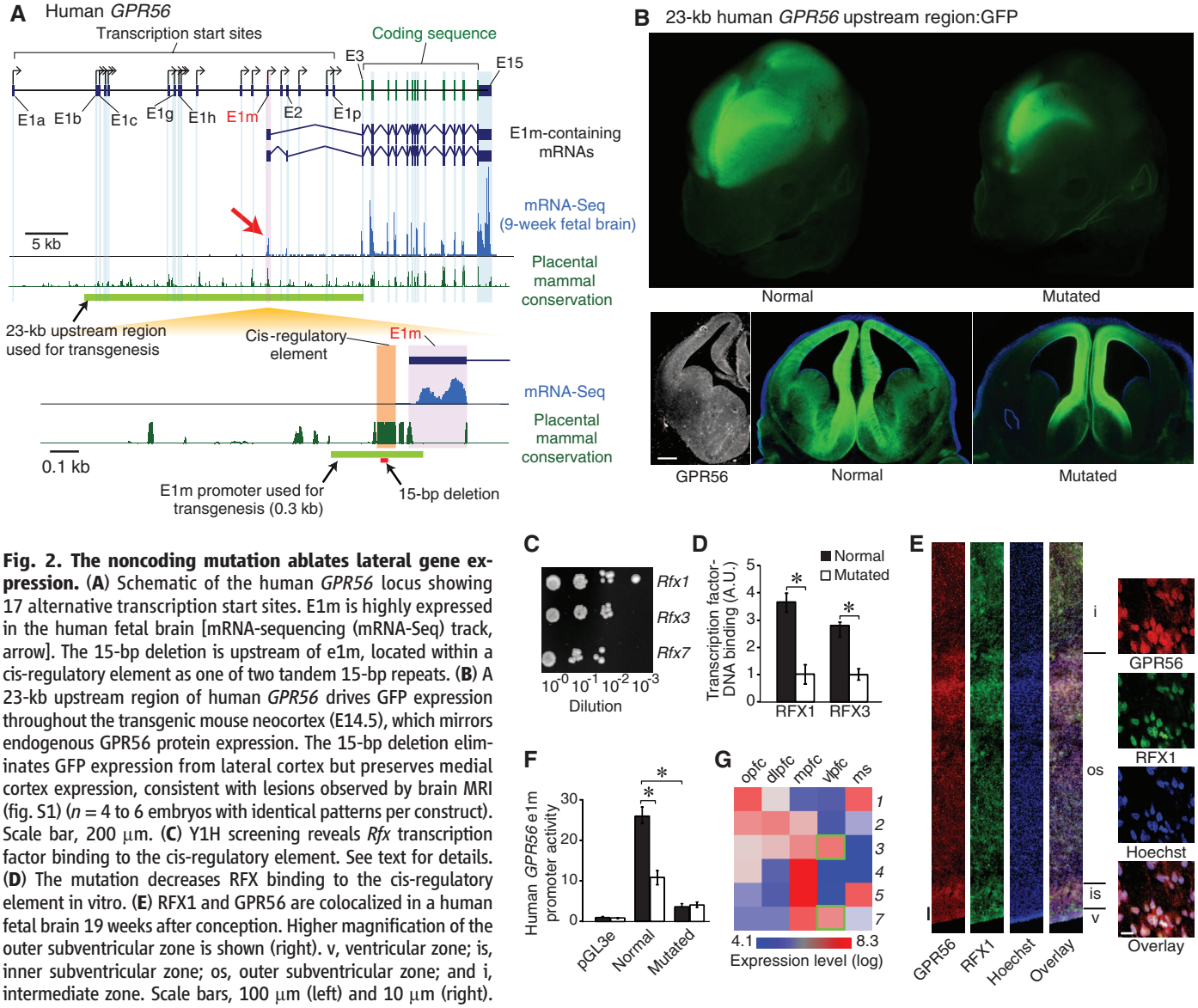


Fig. 2. The noncoding mutation ablates lateral gene expression. (**A**) Schematic of the human *GPR56* locus showing 17 alternative transcription start sites. E1m is highly expressed in the human fetal brain [mRNA-sequencing (mRNA-Seq) track, arrow]. The 15-bp deletion is upstream of e1m, located within a cis-regulatory element as one of two tandem 15-bp repeats. (**B**) A 23-kb upstream region of human *GPR56* drives GFP expression throughout the transgenic mouse neocortex (E14.5), which mirrors endogenous GPR56 protein expression. The 15-bp deletion eliminates GFP expression from lateral cortex but preserves medial cortex expression, consistent with lesions observed by brain MRI (fig. S1) ($n = 4$ to 6 embryos with identical patterns per construct). Scale bar, 200 μm. (**C**) Y1H screening reveals *Rfx* transcription factor binding to the cis-regulatory element. See text for details. (**D**) The mutation decreases RFX binding to the cis-regulatory element in vitro. (**E**) RFX1 and GPR56 are colocalized in a human fetal brain 19 weeks after conception. Higher magnification of the outer subventricular zone is shown (right). v, ventricular zone; is, inner subventricular zone; os, outer subventricular zone; and i, intermediate zone. Scale bars, 100 μm (left) and 10 μm (right). (**F**) Dominant-negative RFX (white bars) abrogates normal e1m promoter activity. Black bars, GFP control. (**G**) Each *RFX* gene has distinct expression patterns in the fetal human brain. Each number means the corresponding *RFX* isoform. *RFX3* and *RFX7* are enriched in regions affected by perisylvian polymicrogyria (green boxes). pfc, prefrontal cortex; opfc, orbital pfc; dlpfc, dorsolateral pfc, mpfc, medial pfc; vlpfc, ventrolateral pfc; ms, motor-sensory cortex. *$P < 0.001$, *t* test.

mitotic progenitor cells and intermediate progenitor cells (Fig. 3, C and D). In utero electroporation (at E13.5 with analysis 48 hours later at E15.5) of a plasmid encoding *GPR56* (as well as GFP, to mark the cells) caused cells to persist in proliferative zones compared with cells expressing GFP alone (Fig. 3E). Changes in the number of intermediate progenitors in transgenic and knockout mice may be secondary to changes in the radial progenitor cells that generate them or might reflect a direct role of GPR56 in intermediate progenitors but is consistent with a report that loss of *TBR2* (*EOMES*) also causes polymicrogyria in humans (*19*).

The cis-regulatory element upstream of *GPR56* e1m is found in genomes of all placental mammals, but not monotremes, marsupials, or nonmammals, which suggests that it emerged after placental and nonplacental mammals diverged 85 to 100 million years ago (fig. S7B). The cis-regulatory element sequence is only found at the e1m locus in *GPR56* but not elsewhere in the genome. E1m itself shows homology at its 3′ end to a long interspersed nuclear element (LINE)–4/RTE, a family of retrotransposons active in early mammals after divergence from marsupials (*20*). Another noncoding *GPR56* exon (exon 2), present only in primates, derives from a primate-specific *Alu* insertion (fig. S7B). In contrast to the >17 alternative first exons in humans, we found only five noncoding first exons in the mouse *Gpr56* gene (Fig. 4A and fig. S7A) (*10*, *11*). Thus, *GPR56* acquired many noncoding upstream exons and generated alternative splice forms with distinct expression patterns (fig. S2, B and D), in the lineage leading to humans. Transposable element insertion played a role in generating this diversity.

To test directly whether evolutionary changes in *GPR56*'s alternative splice forms have functional effects, we generated transgenic mice in which the β-galactosidase (β-gal) gene is driven by a minimal 300-bp e1m promoter from human, mouse, marmoset, dolphin, and cat (Fig. 2A and fig. S6A). The mouse e1m promoter drives β-gal expression broadly in the nervous system in diverse cell types, which suggests that this simple 300-bp e1m promoter is sufficient to recapitulate major features of the endogenous mouse GPR56 expression (*16*–*18*) (Fig. 4B and fig. S6B). In contrast, the corresponding human e1m promoter has a variety of deletions and single-nucleotide variants, relative to the mouse sequence (fig. S6A), and drives much more limited expression in rostral-lateral cortex (Fig. 4B and fig. S6B). Weak lateral cortical expression is visible in embryos carrying the mouse e1m promoter:β-gal transgene,

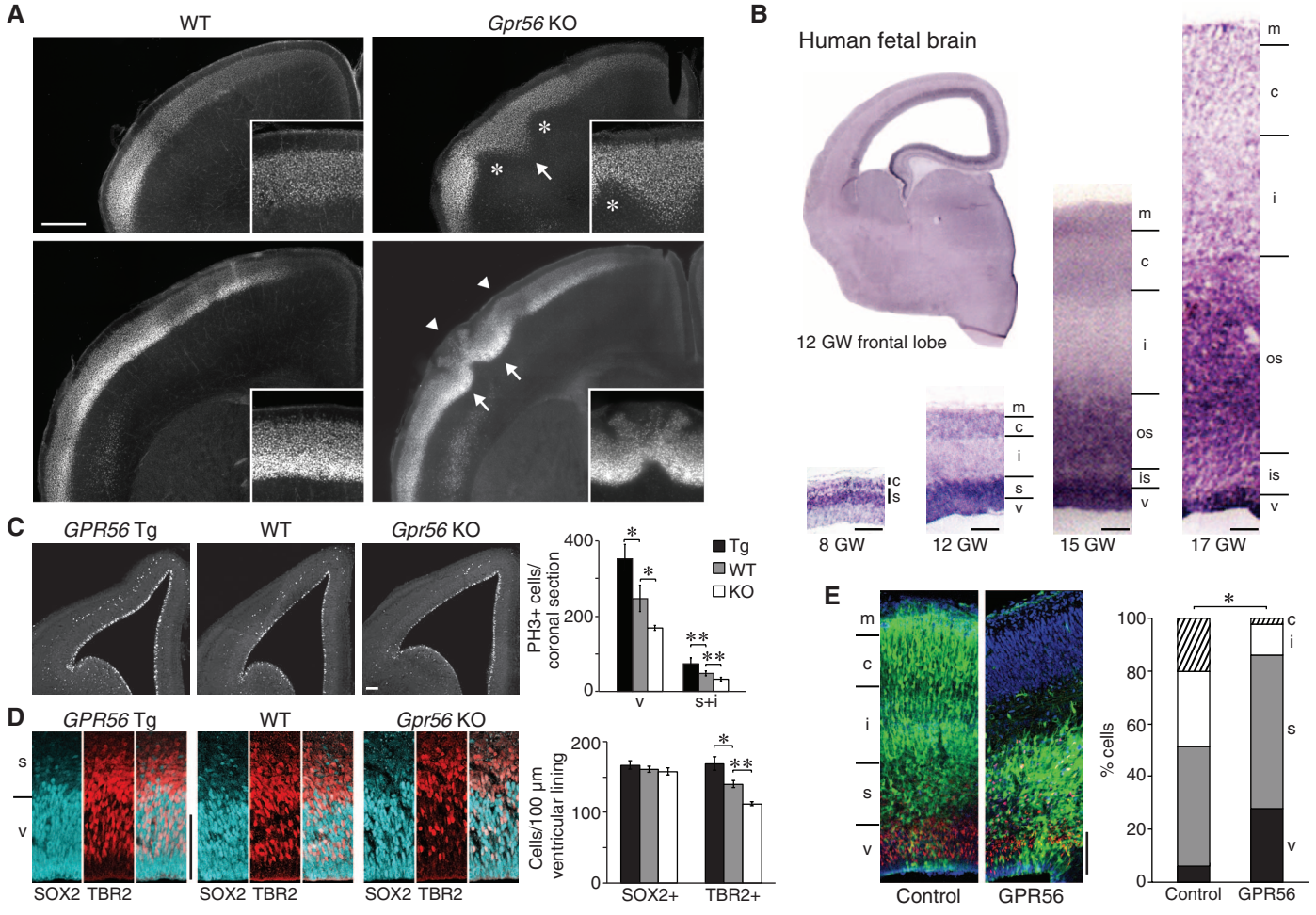


Fig. 3. GPR56 regulates neuroprogenitor proliferation. (**A**) In *Gpr56* knockout mice, neurons overmigrate through breached pial basement membrane (arrowheads) or undermigrate (arrows) forming irregular cortical layers, as shown by immunostaining of Cux1, an upper layer (II to IV) marker (p9). Thin cortex is occasionally observed (asterisks). (**B**) *GPR56* is highly expressed in human ventricular zone and outer subventricular zone at 12 to 17 weeks of gestation (GW), which suggests roles in neuroprogenitors. v, ventricular zone; s, subventricular zone; is, inner subventricular zone; os, outer subventricular zone; i, intermediate zone; c, cortical plate; and m, marginal zone. (**C** to **D**) Human *GPR56* transgenic (Tg) mice have significantly more mitotic (PH3+) neuroprogenitor cells and intermediate progenitor (TBR2+) cells than wild-type (WT). In contrast, *Gpr56* knockout (KO) mice have significantly fewer mitotic cells and intermediate progenitors than WT (E13.5 to E14.5). (n = 7 mice per genotype; *P < 0.005; **P < 0.001; paired t test). (**E**) The cells that are in utero electroporated (from E13.5 to E15.5) with human *GPR56*-IRES-GFP [either side of the internal ribosome entry site (IRES), GFP expressing] persist in the germinal zones longer that the GFP control cells. Red, TBR2; blue, Hoechst. (n = 11 mouse embryos per construct; *P < 0.0001; chi-squared test). Scale bars, 500 μm (A) and 100 μm (B) to (E).

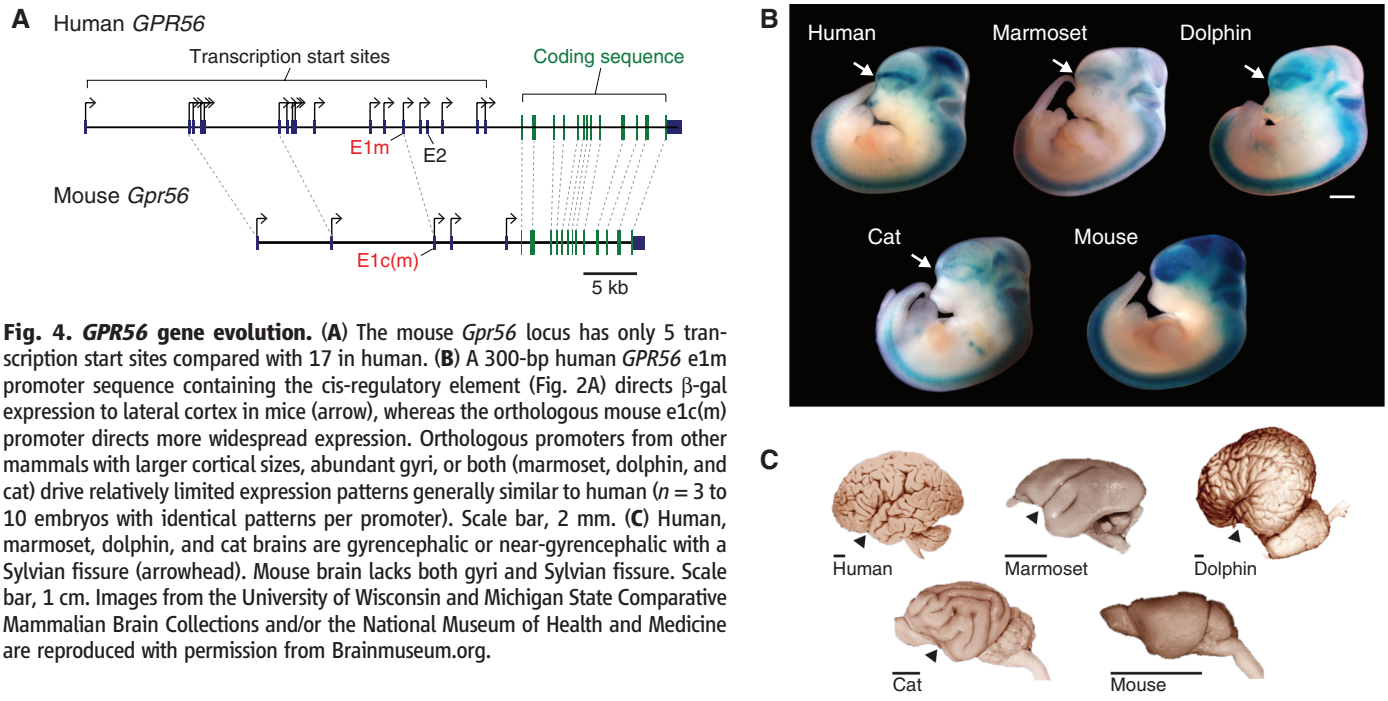


Fig. 4. *GPR56* gene evolution. (**A**) The mouse *Gpr56* locus has only 5 transcription start sites compared with 17 in human. (**B**) A 300-bp human *GPR56* e1m promoter sequence containing the cis-regulatory element (Fig. 2A) directs β-gal expression to lateral cortex in mice (arrow), whereas the orthologous mouse e1c(m) promoter directs more widespread expression. Orthologous promoters from other mammals with larger cortical sizes, abundant gyri, or both (marmoset, dolphin, and cat) drive relatively limited expression patterns generally similar to human ($n = 3$ to 10 embryos with identical patterns per promoter). Scale bar, 2 mm. (**C**) Human, marmoset, dolphin, and cat brains are gyrencephalic or near-gyrencephalic with a Sylvian fissure (arrowhead). Mouse brain lacks both gyri and Sylvian fissure. Scale bar, 1 cm. Images from the University of Wisconsin and Michigan State Comparative Mammalian Brain Collections and/or the National Museum of Health and Medicine are reproduced with permission from Brainmuseum.org.

which suggests that, in humans, additional elements besides the 300-bp e1m promoter region are required to drive the full complement of *GPR56* expression. E1m promoters from marmoset, dolphin, and cat drive expression patterns generally similar to human. The shared expression patterns in the four mammals that have a Sylvian fissure (Fig. 4C) suggest that elaboration of *GPR56* noncoding regulation is consistent with the larger number of noncoding first exons in gyrencephalic mammals and humans. Elaboration of additional alternative splice forms provides a mechanism for potentially independent evolution of these multiple forms.

Our studies show that levels of *GPR56* control proliferation of progenitors in the neocortex. Loss of *GPR56* expression impairs neurogenesis, and overexpression enhances proliferation and progenitor number. Selective *GPR56* loss causes strikingly regional defects of cortical development. *GPR56* likely influences progenitor proliferation by stabilizing the basal process of radial neuroepithelial progenitors, because (i) GPR56 protein localizes to the basal processes of radial neuroepithelial cells (*16*); (ii) GPR56 binds extracellular matrix proteins in the pial basement membrane, such as collagen type III (*17*) and tetraspanins, which bind integrins as well (*21*); (iii) GPR56 is required for normal attachment of basal processes to the pial basement membrane in mice (*16*); and (iv) basal processes regulate progenitor proliferation via integrin signaling (*22*, *23*), and *GPR56* interacts genetically with $\alpha_3\beta_1$ integrin (*24*). The elaboration of the *GPR56* locus in gyrencephalic mammals, and especially humans, to produce many alternative splice forms with diverse expression patterns presents *GPR56* as a key target that could influence the dramatic changes in shape and folding that characterize the forebrain of many mammals. Elaboration and specialization of alternative transcripts with distinct transcription start sites is an evolutionary mechanism that has been difficult to study because of the lack of comprehensive catalogs of RNA splice forms, but continued RNA sequencing studies may soon provide the opportunity to assess its importance systematically.

References and Notes

1. P. Rakic, *Nat. Rev. Neurosci.* **10**, 724–735 (2009).
2. J. H. Lui, D. V. Hansen, A. R. Kriegstein, *Cell* **146**, 18–36 (2011).
3. K. Zilles, N. Palomero-Gallagher, K. Amunts, *Trends Neurosci.* **36**, 275–284 (2013).
4. K. Amunts *et al.*, *PLOS Biol.* **8**, e1000489 (2010).
5. J. A. Golden, B. N. Harding, *Nat. Rev. Neurol* **6**, 471–472 (2010).
6. A. J. Barkovich, *Neuroradiology* **52**, 479–487 (2010).
7. X. Piao *et al.*, *Science* **303**, 2033–2036 (2004).
8. X. Piao *et al.*, *Ann. Neurol.* **58**, 680–687 (2005).
9. N. Bahi-Buisson *et al.*, *Brain* **133**, 3194–3209 (2010).
10. D. Thierry-Mieg, J. Thierry-Mieg, *Genome Biol.* **7** (suppl. 1), S12–S14 (2006).
11. Y. Suzuki, R. Yamashita, K. Nakai, S. Sugano, *Nucleic Acids Res.* **30**, 328–331 (2002).
12. S. J. Ansley *et al.*, *Nature* **425**, 628–633 (2003).
13. A. Jolma *et al.*, *Genome Res.* **20**, 861–873 (2010).
14. D. Zhang *et al.*, *J. Neurochem.* **98**, 860–875 (2006).
15. H. J. Kang *et al.*, *Nature* **478**, 483–489 (2011).
16. S. Li *et al.*, *J. Neurosci.* **28**, 5817–5826 (2008).
17. R. Luo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12925–12930 (2011).
18. S. J. Jeong, R. Luo, S. Li, N. Strokes, X. Piao, *J. Comp. Neurol.* **520**, 2930–2940 (2012).
19. L. Baala *et al.*, *Nat. Genet.* **39**, 454–456 (2007).
20. T. S. Mikkelsen *et al.*, *Nature* **447**, 167–177 (2007).
21. L. Xu, R. O. Hynes, *Cell Cycle* **6**, 160–165 (2007).
22. R. Radakovits, C. S. Barros, R. Belvindrah, B. Patton, U. Müller, *J. Neurosci.* **29**, 7694–7705 (2009).
23. S. A. Fietz *et al.*, *Nat. Neurosci.* **13**, 690–699 (2010).
24. S. J. Jeong *et al.*, *PLOS ONE* **8**, e68781 (2013).

Acknowledgments: Research performed on samples of human origin was conducted according to protocols approved by participating institutions, including Boston Children's Hospital and Beth Israel Deaconess Medical Center. The human embryonic and fetal material was provided by the Joint Medical Research Council (grant no. G0700089)–Wellcome Trust (grant no. GR082557) Human Developmental Biology Resource (www.hdbr.org) and the National Institute of Child Health and Human Development, NIH, Brain and Tissue Bank at the University of Maryland (contract no. HHSN275200900011C, reference no. NO1-HD-9-0011). *Gpr56* knockout mice are from Genentech. This work was supported by the Strategic Research Program for Brain Sciences and from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Japan (H.O.); Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program) (H.O.); U01MH081896 from National Institute of Mental Health, NIH (N.S.); 2R01NS035129 from National Institute of Neurological Disorders and Stroke, NIH (C.A.W.); and The Paul G. Allen Family Foundation (C.A.W.). Additional funding support listed in supplementary materials. C.A.W. is an investigator of the Howard Hughes Medical Institute. *Gpr56* knockout mice are available from Genentech subject to a Material Transfer Agreement.

Supplementary Materials

www.sciencemag.org/content/343/6172/764/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
Tables S1 and S2
Movies S1 to S2
References (*25*–*40*)

7 August 2013; accepted 17 December 2013
10.1126/science.1244392

Origin and Spread of de Novo Genes in *Drosophila melanogaster* Populations

Li Zhao,[1]* Perot Saelao,[1] Corbin D. Jones,[2] David J. Begun[1]*

Comparative genomic analyses have revealed that genes may arise from ancestrally nongenic sequence. However, the origin and spread of these de novo genes within populations remain obscure. We identified 142 segregating and 106 fixed testis-expressed de novo genes in a population sample of *Drosophila melanogaster*. These genes appear to derive primarily from ancestral intergenic, unexpressed open reading frames, with natural selection playing a significant role in their spread. These results reveal a heretofore unappreciated dynamism of gene content.

Although the vast majority of genes present in any species descend from a gene present in an ancestor, recent analyses suggest that some genes originate from ancestrally nongenic sequences (*1*–*3*). Evidence for these "de novo" genes has generally derived from a combination of phylogenetic and genomic/transcriptomic analyses that reveal evidence of lineage- or species-specific transcripts associated with nongenic orthologous sequences in sister species. De novo genes, which were first identified in *Drosophila* (*1*–*3*), have also been identified in humans, rodents, rice, and yeast (*4*–*9*). In *Drosophila*, de novo genes tend to be specifically expressed in tissues associated with male reproduction (*2*, *10*), which suggests that sexual or gametic selection may be important (*1*–*3*, *9*), although other functional roles may evolve (*10*, *11*). Because previous studies of de novo gene evolution used comparative rather than population genetic approaches, the earliest steps in de novo gene origination remain mysterious. Here, we used population genomic and transcriptomic data from *Drosophila melanogaster* and its close relatives to investigate the origin and spread of de novo genes within populations.

Illumina paired-end RNA sequencing (RNA-seq) and de novo and reference-guided assembly and alignment were used to characterize the testis transcriptome of six previously sequenced inbred Raleigh (RAL) *D. melanogaster* strains (*12*); an average of 65 million paired-end reads were produced for each strain (table S1). We inferred (*13*) the presence of 142 polymorphic de novo candidate genes that are expressed in at least one RAL strain but are not known on the basis of publicly available data from *D. melanogaster*. The median number of segregating de novo genes carried per strain was 49. Reverse transcription polymerase chain reaction (RT-PCR) and 5′ and 3′ rapid amplification of cDNA ends (RACE) in a subset of genes supported inferences from RNA-seq analysis (table S2). These candidate polymorphic genes correspond to unique, intergenic sequence in the *D. melanogaster* reference sequence (table S3), are alignable to unique orthologous regions in the *D. simulans* and *D. yakuba* reference sequences, and show no significant BLASTP hits to the NCBI nr (nonredundant) protein database. The candidate genes exhibited expression neither in testis RNA-seq data from three *D. simulans* and two *D. yakuba* strains (table S1 and fig. S1) nor in whole male and female RNA-seq data from 59 *D. simulans* strains (*13*). None of the candidates showed significant expression in whole females from the same *D. melanogaster* strains used for testis RNA-seq (table S4). These data support the hypothesis that the 142 candidates are new, male-specific, de novo genes still segregating in *D. melanogaster*. Expression levels of the candidate genes greatly exceed levels of background transcription in intergenic sequence (fig. S2) (*13*); several additional attributes of these genes, as described below, support the hypothesis that the observed transcripts are biologically meaningful.

Segregating de novo genes were moderately expressed (Fig. 1A and Table 1), but their expression was significantly lower than that of annotated male-biased genes (*13*) (Table 1) or annotated genes (table S6). We observed no enrichment of polymorphic de novo genes near annotated male-biased genes and no significant correlation between the strand (+/−) of polymorphic de novo genes and that of their immediate annotated neighbors [χ^2 test, $P > 0.1$ (table S5 and fig. S3), supported by simulations (*13*)]. There was a marginally significant underrepresentation of X chromosome segregating de novo genes relative to annotated male-biased genes (10 genes are X-linked; *t* exact test, $P = 0.01$; Fig. 1B). This result stands in contrast to speculation based on a small sample of older, fixed de novo genes (*2*, *3*) that de novo male-biased genes are overrepresented on the X chromosome.

As expected, de novo genes were significantly shorter and simpler than annotated genes and annotated male-biased genes (Table 1 and table S6). This pattern is likely due mostly to the larger proportion of polymorphic de novo genes that are single-exon (57.0%) compared to the proportion of annotated single-exon (table S6) or single-exon male-biased genes (Table 1) (*13*). Among the 61 multi-exon de novo genes, the majority of splice events (98%) were associated with canonical sites; rare noncanonical splice sites were found in four genes as minor isoform splice events, which were similar to those previously observed in *D. melanogaster* (*14*). Alternative splicing was observed in 20 of the 61 multi-exon segregating de novo genes (table S7), with conserved reading frames across alternative isoforms. Genes associated with alternative splicing generally exhibited

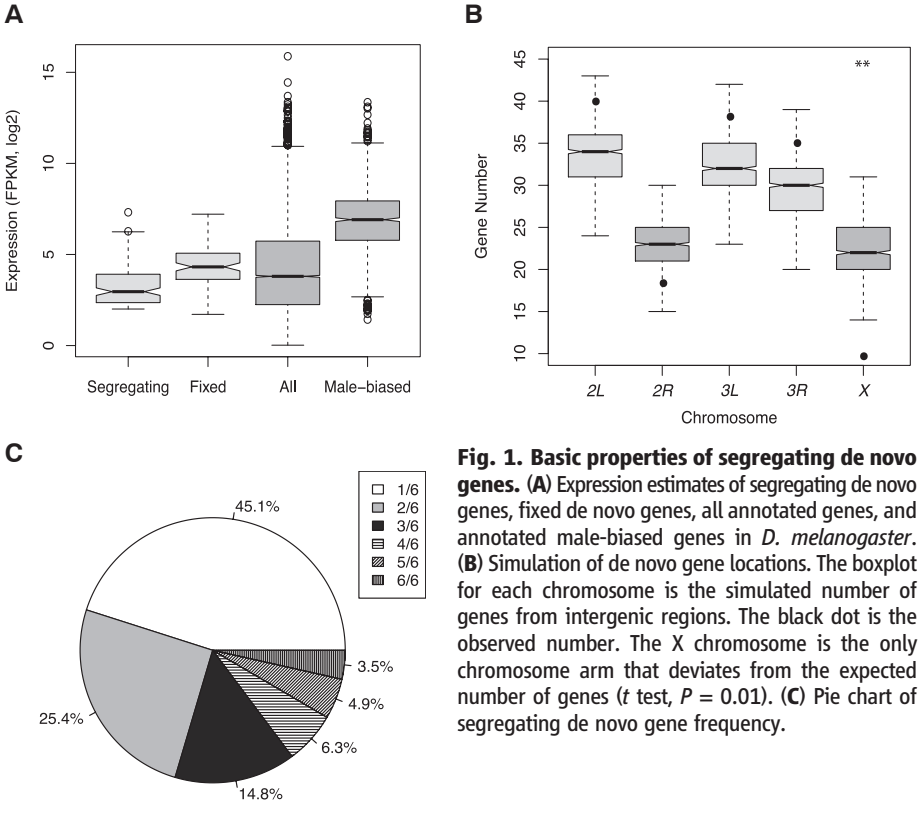


Fig. 1. Basic properties of segregating de novo genes. (**A**) Expression estimates of segregating de novo genes, fixed de novo genes, all annotated genes, and annotated male-biased genes in *D. melanogaster*. (**B**) Simulation of de novo gene locations. The boxplot for each chromosome is the simulated number of genes from intergenic regions. The black dot is the observed number. The X chromosome is the only chromosome arm that deviates from the expected number of genes (*t* test, $P = 0.01$). (**C**) Pie chart of segregating de novo gene frequency.

[1]Department of Evolution and Ecology, University of California, Davis, CA 95616, USA. [2]Department of Biology and Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC 27599, USA.

*Corresponding author. E-mail: lizzhao@ucdavis.edu (L.Z.); djbegun@ucdavis.edu (D.J.B.)

multiple isoforms across strains that expressed the corresponding gene, with no evidence of genetic variation for alternative splice use.

Of 142 polymorphic genes, 134 (94%) had a minimum open reading frame (ORF) of at least 150 base pairs (bp) and were classified as potentially coding. To determine the likelihood that the high proportion of genes harboring long ORFs occurred by chance, we investigated the coding potential of intergenic regions in the reference sequence, focusing on single-exon ORFs. We observed that 59.9% of random 800-bp intergenic sequences were associated with a ≥150-bp single-exon ORF, whereas 97.5% of the observed single-exon de novo genes were associated with such an ORF ($P < 0.01$). Moreover, the mean length of single-exon de novo gene ORFs was substantially greater than that expected in random intergenic sequence ($P < 0.05$). These observations further support the idea that the observed transcripts are unlikely to be explained simply as random noise. The eight polymorphic de novo genes that did not satisfy our arbitrary minimum ORF criterion were autosomal and slightly smaller (mean transcript length = 743 bp) than ORF-containing polymorphic genes. Because orthologous sequences from expressing and non-expressing *D. melanogaster* lines have similar coding potential, most segregating de novo genes are likely to have resulted from the recruitment of small, preexisting, unexpressed ORFs (*1*). For *D. simulans* and *D. yakuba* orthologous sequences, 70% and 45%, respectively, contained ORFs similar to those observed for segregating genes in *D. melanogaster*. Of the 134 predicted de novo proteins, 41.8% may be intrinsically unfolded (fig. S4, A to D) and 50% of these have predicted binding regions (fig. S4E); both observations are consistent with potential biological function (*15*). For putative protein-coding genes, the average 5′ and 3′ untranslated region (5′UTR and 3′UTR) lengths—248 bp and 364 bp, respectively—were slightly shorter than the average lengths for annotated *D. melanogaster* genes but were slightly longer than the averages for annotated male-biased genes (Table 1). The incidence of the two major polyadenylation signals (AAUAAA and AUUAAA) in or near the putative 3′UTRs of segregating de novo genes was similar to, but slightly lower than, the incidence in the whole genome (table S8). Overall, polymorphic de novo genes have structural organization consistent with small protein-coding genes in the species.

Segregating de novo genes either were expressed at a relatively high level in expressing strains or showed almost no evidence of expression in other strains. Hartigan's dip test on transcript abundance estimates rejected unimodality for 134 of 142 genes and was consistent with bimodal expression across lines for most genes. We used a cutoff of two fragments per kilobase of exon per million fragments mapped (FPKM > 2) for inferring expression of a transcript in a line (*16*) to determine the proportion of strains, from 0.17 (1/6) to 1.0 (6/6) expressing each transcript. Because no candidates show expression in the reference sequence strain, the genes expressed in all six RAL strains are considered to be polymorphic in the species. More than half the genes (55%) were not rare in the Raleigh sample, as they were expressed in at least two of the six RAL strains (Fig. 1C); 29.5% were definitely common, being expressed in three or more strains, which is inconsistent with mutation-selection balance. We observed 106 unannotated male-specific transcripts expressed in all six strains and in the reference strain (table S9) but not in the outgroup strains. The corresponding "fixed" de novo genes were not included in downstream analyses relating to segregating genes.

We extracted the 100 bp upstream and 50 bp downstream of the inferred transcription start site (TSS) from the genome sequences of the expressing strains for each of the 61 multi-exon genes. MDscan identified and clustered motifs in these flanking sequences; sequence logos were then generated. We observed four common consensus sequence motifs (8 or 10 bp; Fig. 2A), each of which was found associated with roughly half the segregating de novo genes (*13*) (table S10). In total, 371 annotated male-biased genes (23.3%) were also associated with at least one of these motifs, which suggests that the de novo genes share regulatory features with known male-biased genes. We identified 67 annotated male-biased genes (table S11) that have two or more motifs in the 5′ regions. However, GO (Gene Ontology) enrichment analysis (fig. S5) provided no insight into the possible functions of de novo genes.

These data support the hypothesis that de novo gene expression is influenced by cis-acting variants in the regions corresponding to the 5′ flanking regions of expressing chromosomes. In the simplest case that de novo gene expression is due to a single noncoding nucleotide change, one would predict an excess of fixed differences between expressing and non-expressing chromosomes in flanking regions compared to random samples of intergenic sequences. We focused on the 32 genes expressed in more than two strains and for which our genetic analysis (*13*) supported cis-acting variation driving de novo gene expression. Of these genes, 31.2% exhibited a fixed, derived single-nucleotide polymorphism (SNP) within 500 bp upstream of the TSS, whereas only 8.43% of simulated "genes" (intergenic regions defined by harboring derived SNPs with the same frequency distribution as the 32 observed genes) exhibited a fixed SNP in the comparable 5′ region ($P < 0.01$). More generally, divergence between expressing and non-expressing chromosomes for these 500-bp regions was significantly greater than divergence in simulated data ($P = 0.048$); this finding supports the hypothesis that cis-regulatory changes play a role in de novo gene origination.

Under this hypothesis, segregating genes should be associated with allele-specific expression. We thus measured allelic imbalance (*17*, *18*) in the testis in a set of three unique F_1 genotypes created by crossing the six RAL strains (table S1) (*13*). For the 59 autosomal genes for which one parent expressed the gene and the other did not, expression patterns in the heterozygote for 51 genes were explained completely by cis-acting variation (i.e., allelic imbalance was complete); 7 genes showed evidence of regulation by both cis-acting and trans-acting factors. Only 1 of the 59 genes showed no evidence of allelic imbalance, consistent with expression driven solely by trans-acting variation (table S12). More generally, for genes expressed in both parents, the expression of alleles in F_1 was consistent with expression levels in each parental line (table S13), further supporting the importance of cis-acting expression variants. The roughly bimodal expression patterns and the dominant role of cis effects support the idea that the proportion of lines expressing a gene provides an estimate of its population frequency.

One population genetic explanation for polymorphic de novo genes is that singleton genes (45% of genes) are primarily deleterious and that higher-frequency genes are primarily neutral. If the deleterious nature of de novo genes were due to the cost of transcription or translation, or from toxic interactions of the resulting RNAs or proteins with other molecules, then lower-frequency genes should be more abundantly expressed and

Table 1. Properties of segregating and fixed de novo genes and comparison with annotated male-biased genes in *D. melanogaster*. Wilcoxon test, ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$; ns, not significant. For segregating de novo genes, *P* values are comparisons of segregating versus fixed genes and segregating versus male-biased genes. For fixed de novo genes, *P* values are comparisons of fixed de novo genes versus male-biased genes. Male-biased genes are as defined in (*13*). All estimates are medians, except for exon number (mean).

	Segregating de novo genes	Fixed de novo genes	Male-biased genes
Number	142	106	1595
Transcript length (bp)	801 ns/***	1013**	1184
Exon length (bp)	518*/***	512***	355
Exon number	1.47*/***	1.79*	2.37
Intron length (bp)	91*/***	70.5***	77
5′UTR length (bp)	248*/***	267.5***	170
3′UTR length (bp)	364 ns/***	337***	267
Single-exon gene (%)	57*/***	48.1***	35.8
Expression (FPKM)	7.78***/***	19.96***	66.54

longer than higher-frequency genes. However, contrary to this expectation, lower-frequency genes were expressed at a lower level, were shorter, and were less complex than higher-frequency genes (table S6) (*13*). The different properties of rare versus common de novo genes (Table 2) (*13*) supports the idea that de novo genes having certain properties (e.g., greater expression, longer transcripts, more exons) are more likely to spread under selection.

We investigated the role of directional selection on polymorphic de novo genes by determining whether they are associated with reduced nucleotide diversity (*19*, *20*). For each de novo gene expressed in at least two strains, we compared the nucleotide diversity (π) for expressed sequence (strains) versus non-expressed orthologous sequence (non-expressing strains) and compared the observed differences to a frequency-corrected expected value from resampling of intergenic sequence from the six RAL strains (*13*). For 46 of 65 genes, π was lower in the expressed lines (mean = 0.0060) than in the non-expressed lines (mean = 0.0092) and exhibited a roughly 38% reduction relative to non-expressed orthologous sequence over the 65 genes (Wilcoxon test, $P = 0.003$). For 30 genes, π was significantly lower in the expressed lines (Wilcoxon test, $P < 0.05$). The region of reduced heterozygosity near expressed sequences is on the scale of 5 kb or less (Fig. 2B and fig. S6), which is contrary to the expectation of strong selection on new mutations (*19*) but consistent with weaker selection (*20*) or soft selective sweeps (*21*) (Fig. 2, C and D). Polymorphic de novo genes were significantly (Wilcoxon test, $P < 0.001$) (*13*) more likely to be differentially expressed between populations (29 of 142, or 17%) relative to annotated genes (4.5%) and male-biased genes (6.3%), which also supports the idea that selection may play a role in their spread.

We used the Hudson-Kreitman-Aguade–like (HKAl) test statistic (*22*, *23*) to compare the heterozygosity/divergence ratio for genomic regions associated with fixed de novo genes to that observed for appropriately sampled intergenic regions (*13*, *20*). The HKAl for fixed regions (mean = –0.48) was significantly smaller than that expected for comparable random intergenic regions (mean = 0.12; Wilcoxon test, $P < 0.001$). Moreover, regions corresponding to fixed genes associated with higher expression (FPKM > 10) exhibited a smaller HKAl statistic relative to regions associated with fixed genes having lower (FPKM ≤ 10) expression (HKAl = –0.33 versus –0.86; Wilcoxon test, $P < 0.001$). These observations also support the hypothesis that de novo genes have been influenced by directional selection.

Our analyses suggest that there are many polymorphic de novo male-specific genes in *D. melanogaster* populations, likely recruited by selection primarily from ancestral, unexpressed ORFs (fig. S7). Given the small number of genotypes investigated for a single tissue and our strict filtering criteria, we have likely substantially underestimated the number of polymorphic de novo genes. Our results also suggest the existence of many more fixed de novo *D. melanogaster* genes than previously inferred (*2*, *4*, *10*), which supports the idea that a substantial genetic component of male reproductive biology in this species remains completely unexplored.

More generally, our results suggest that important attributes of an organism's biology cannot be accurately represented or investigated without knowledge of de novo gene variation within species. In the absence of gene loss, de novo gene gain would lead to a long-term increase in gene number. Although our analyses are consistent with substantial numbers of polymorphic gene losses, we observed no population genetic evidence that losses result from directional selection (*13*). Thus, de novo genes may often spread under selection, while gene loss may occur primarily as a result of drift associated with loss of ancestral gene function. However, important details of such processes remain obscure, and much additional work is required to clarify the dynamics, biochemical and genetic properties, and phenotypic effects of young de novo genes and the processes underlying gene loss in natural populations.

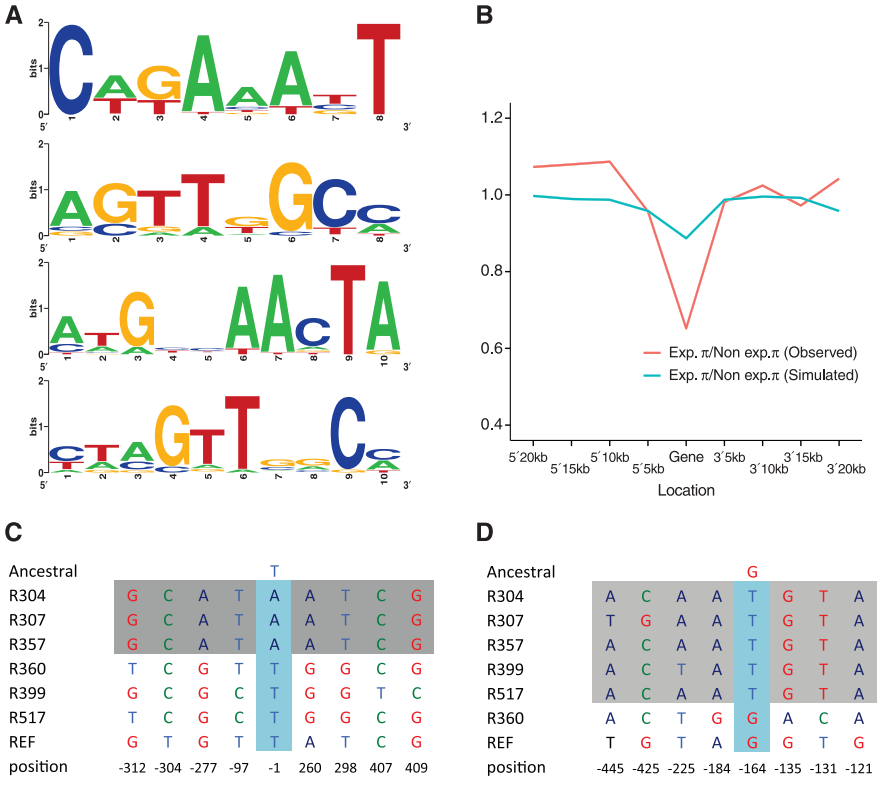


Fig. 2. Regulation and population genetics of segregating de novo genes. (**A**) Potential cis-regulatory elements. The most common shared 8- and 10-bp consensus motifs in 5′ flanking regions are listed. From top to bottom, 34, 29, 25, and 30 multiple-exon genes show these motifs. (**B**) Nucleotide diversity (π) for de novo genes and flanking regions. The red line is the ratio of π-expressing lines to π–non-expressing lines; the green line shows expected values from resampling of intergenic DNA conditional on the same derived allele frequency distribution as the observed de novo genes. π estimates for 5′ and 3′ flanking regions of genes were incremented in 5-kb windows. (**C**) A gene (Gene_X_141) that may have experienced a hard selective sweep. Gray box denotes expressing lines. The TSS region contains a derived allele fixed in expressing strains and absent in non-expressing strains; flanking regions are homozygous in expressing strains. (**D**) A gene (Gene_3L_079) showing no evidence of hard sweep. Gray box denotes expressing lines. The TSS region includes a derived allele fixed in expressing lines, but the flanking regions of expressing chromosomes retain nucleotide variation.

Table 2. Properties of segregating genes differ across frequency classes. Wilcoxon test, ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$; P values are comparison of singleton versus nonsingleton genes and singleton versus high-frequency (≥ 3/6) genes. FPKM and transcript length estimates are medians; exon numbers are means.

	Singleton	Nonsingleton	High-frequency
FPKM	5.76***/***	9.91	12.31
Transcript length (bp)	723**/***	869	1312
Exon number	1.38*/**	1.53	1.81

References and Notes

1. D. J. Begun, H. A. Lindfors, M. E. Thompson, A. K. Holloway, *Genetics* **172**, 1675–1681 (2006).
2. M. T. Levine, C. D. Jones, A. D. Kern, H. A. Lindfors, D. J. Begun, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9935–9939 (2006).
3. D. J. Begun, H. A. Lindfors, A. D. Kern, C. D. Jones, *Genetics* **176**, 1131–1137 (2007).
4. Q. Zhou *et al.*, *Genome Res.* **18**, 1446–1455 (2008).
5. D. G. Knowles, A. McLysaght, *Genome Res.* **19**, 1752–1759 (2009).
6. W. Xiao *et al.*, *PLOS ONE* **4**, e4603 (2009).
7. J. Cai, R. Zhao, H. Jiang, W. Wang, *Genetics* **179**, 487–496 (2008).
8. A. R. Carvunis *et al.*, *Nature* **487**, 370–374 (2012).
9. T. J. A. J. Heinen, F. Staubach, D. Häming, D. Tautz, *Curr. Biol.* **19**, 1527–1531 (2009).
10. S. Chen, Y. E. Zhang, M. Long, *Science* **330**, 1682–1685 (2010).
11. J. A. Reinhardt *et al.*, *PLOS Genet.* **9**, e1003860 (2013).
12. T. F. C. Mackay *et al.*, *Nature* **482**, 173–178 (2012).
13. See supplementary materials on *Science* Online.
14. N. Sheth *et al.*, *Nucleic Acids Res.* **34**, 3955–3967 (2006).
15. V. N. Uversky, *Protein Sci.* **11**, 739–756 (2002).
16. D. Hebenstreit *et al.*, *Mol. Syst. Biol.* **7**, 497 (2011).
17. D. Schwartz, *Genetics* **67**, 411–425 (1971).
18. R. M. Graze *et al.*, *Mol. Biol. Evol.* **29**, 1521–1532 (2012).
19. J. M. Smith, J. Haigh, *Genet. Res.* **23**, 23–35 (1974).
20. N. L. Kaplan, R. R. Hudson, C. H. Langley, *Genetics* **123**, 887–899 (1989).
21. J. Hermisson, P. S. Pennings, *Genetics* **169**, 2335–2352 (2005).
22. D. J. Begun *et al.*, *PLOS Biol.* **5**, e310 (2007).
23. C. H. Langley *et al.*, *Genetics* **192**, 533–598 (2012).

Acknowledgments: We thank the Begun lab for valuable comments, especially N. Svetec, J. Cridland, A. Sedghifar, T. Seher, and Y. Brandvain. We also thank three anonymous reviewers and M. W. Hahn, C. H. Langley, J. Anderson, and A. Kopp for comments and suggestions. We acknowledge Phyllis Wescott (1950–2011) for more than 30 years of service in the *Drosophila* media kitchen at UC Davis. Funded by NIH grant GM084056 (D.J.B.) and NSF grant 0920090 (C.D.J and D.J.B.). Author contributions: L.Z. and D.J.B. conceived and designed the experiments; L.Z., P.S., and C.D.J. generated data; L.Z. performed data analysis; P.S. and L.Z. did molecular experiments; L.Z. and D.J.B. wrote the paper; and L.Z., D.J.B., and C.D.J. revised the paper. Illumina reads produced in this study are deposited at NCBI BioProject under accession number PRJNA210329.

Supplementary Materials

www.sciencemag.org/content/343/6172/769/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S17
References (*24*–*58*)

11 November 2013; accepted 7 January 2014
Published online 23 January 2014;
10.1126/science.1248286

Crude Oil Impairs Cardiac Excitation-Contraction Coupling in Fish

Fabien Brette,[1] Ben Machado,[1] Caroline Cros,[1] John P. Incardona,[2] Nathaniel L. Scholz,[2] Barbara A. Block[1]*

Crude oil is known to disrupt cardiac function in fish embryos. Large oil spills, such as the Deepwater Horizon (DWH) disaster that occurred in 2010 in the Gulf of Mexico, could severely affect fish at impacted spawning sites. The physiological mechanisms underlying such potential cardiotoxic effects remain unclear. Here, we show that crude oil samples collected from the DWH spill prolonged the action potential of isolated cardiomyocytes from juvenile bluefin and yellowfin tunas, through the blocking of the delayed rectifier potassium current (I_{Kr}). Crude oil exposure also decreased calcium current (I_{Ca}) and calcium cycling, which disrupted excitation-contraction coupling in cardiomyocytes. Our findings demonstrate a cardiotoxic mechanism by which crude oil affects the regulation of cellular excitability, with implications for life-threatening arrhythmias in vertebrates.

Crude oil is a complex chemical mixture containing hydrocarbons (aliphatic and aromatic) and other dissolved-phase organic compounds. Toxicity research on crude oil constituents has focused mainly on polycyclic aromatic hydrocarbons (PAHs) (*1*, *2*), pervasive environmental contaminants that are also found in coal tar, creosote, air pollution, and land-based runoff. In the aftermath of oil spills, PAHs can persist for many years in marine habitats and thereby create pathways for lingering biological exposure and associated adverse effects.

PAH toxicity is structure-dependent, and the carcinogenic, mutagenic, and teratogenic properties of many individual PAHs are known (*3*, *4*). Developing fish are particularly vulnerable to dissolved PAHs in the range of ~100 parts per billion (ppb or μg/liter) down to ≤10 μg/liter. Consequently, PAH toxicity to fish early life stages is an important contributor to both acute and long-term impacts of environmental disasters (*2*, *5*). Numerous studies on crude oils and PAHs, particularly in the aftermath of the Exxon Valdez spill, have described embryonic heart failure, bradycardia, arrhythmias, reduction of contractility, and a syndrome of cardiogenic fluid accumulation (edema) in exposed fish embryos (*6*, *7*). These severe effects are lethal to embryos and larval fishes (*8*–*10*) and could be due to atrioventricular conduction block (*11*).

Despite recent progress using zebrafish and other experimental models to study PAH cardiotoxicity (*12*), the mechanisms that underpin the physiological effects on cardiac function and changes in cardiac morphology during development are not known. The Deepwater Horizon (DWH) oil spill released >4 million barrels of crude oil during the peak spawning months for Atlantic bluefin tuna (*Thunnus thynnus*) in 2010. This large and long-lived species reaches a mass of 650 kg over a life span of 35 years or more (*13*), and the Gulf of Mexico population of bluefin tuna is severely depleted (*14*). Electronic-tagging data confirm that bluefin tuna spawn in the vicinity of the DWH spill, which indicates that bluefin tuna embryos, larvae, juveniles, and adults were likely exposed to crude oil–derived PAHs (*14*). Many other Gulf of Mexico pelagics may have spawned in oiled habitats, including yellowfin tuna, dolphin fish, blue marlin, and swordfish (*15*).

To more precisely define the mechanisms of crude oil cardiotoxicity and to evaluate the potential vulnerability of eggs, larvae, and juveniles in the vicinity of the DWH spill, we assessed the impact of field-collected DWH oil samples on in vitro cardiomyocyte preparations dissociated from the hearts of bluefin tuna (*T. orientalis)* and yellowfin tuna (*T. albacares*). Juvenile tunas were caught at sea and held in captivity at the Tuna Research Conservation Center and the Monterey Bay Aquarium (*16*).

The cardiotoxic effects of four distinct environmental samples of MC252 crude oil were assessed as water-accommodated fractions (WAFs) prepared in Ringer solution for marine fish (*16*). Oil samples were collected under chain of custody during the DWH spill response effort. The samples included riser "source" oil (sample 072610-03), riser oil that was "artificially weathered" by heating at 90° to 105°C (sample 072610-W-A), and two skimmed oil samples: "slick A" (sample CTC02404-02), collected 29 July 2010, and "slick B" (sample GU2888-A0719-OE701), collected 19 July 2010 by the U.S. Coast Guard cutter *Juniper*. High-energy WAFs were prepared in a commercial blender that dispersed oil droplets to mimic release conditions at the MC252 well head (*16*). As expected from previous studies (*11*, *12*), the total sum (∑) of PAHs declined in WAFs from source oil to the surface-weathered samples, owing to loss of naphthalenes, whereas the total concentrations of three-ringed PAHs (e.g., phenanthrenes) increased proportionately (fig. S1 and table S1). PAH concentrations were in a range expected to cause cardiotoxicity in intact embryos and consistent with the ∑PAHs measured in some surface water samples during the DWH oil spill (up to 84 μg/liter) (*16*, *17*). WAFs in Ringer solution were perfused over freshly dissociated, isolated tuna cardiomyocytes, and we assessed the effects of these oil-containing solutions on excitation-contraction (EC) coupling using electrophysiological and Ca^{2+}-imaging techniques.

[1]Department of Biology, Stanford University, Hopkins Marine Station, Pacific Grove, CA 93950, USA. [2]Northwest Fisheries Science Center, National Oceanic and Atmospheric Administration, Seattle, WA 98112, USA.

*Corresponding author. E-mail: bblock@stanford.edu

Patch-clamp recordings revealed a strong effect of DWH source oil and weathered oil on bluefin and yellowfin tuna cardiomyocytes' action potential duration (APD) (Fig. 1 and fig. S3). A concentration-dependent lengthening of the APD waveform was observed in both tuna species. APD at 90% repolarization (i.e., equivalent to the QT interval on an electrocardiogram) was significantly increased across all four oil samples at ∑PAH concentrations ranging from 4 to 61 μg/liter (table S1). The source and weathered oils significantly decreased the APD at 10% repolarization (APD_{10}) (Fig. 1). WAF exposures did not influence other action potential parameters, such as resting membrane potential and action potential amplitude (figs. S2 and S3). This suggests that I_{K1}, the background current responsible for resting membrane potential, and I_{Na}, the current responsible for the upstroke of the action potential, are not modified by crude oil. All four oil samples significantly increased the time for repolarization from APD_{30} to APD_{90}. This increase in triangulation (Fig. 1, I to L, and fig. S3, E and F) is a strong predictor of fatal cardiac arrhythmia (*18*). Pharmacological agents that cause a cardiac repolarization disorder by lengthening cardiomyocyte APD, as well as congenital mutations of hERG (human *ether-à-go-go–related gene* or *KCNH2*) channels—the mammalian homolog to the fish delayed rectifier potassium current (I_{Kr}) (*19*)—are known to cause or aggravate ventricular arrhythmias, which can result in torsade de pointes and/or sudden death (*20*).

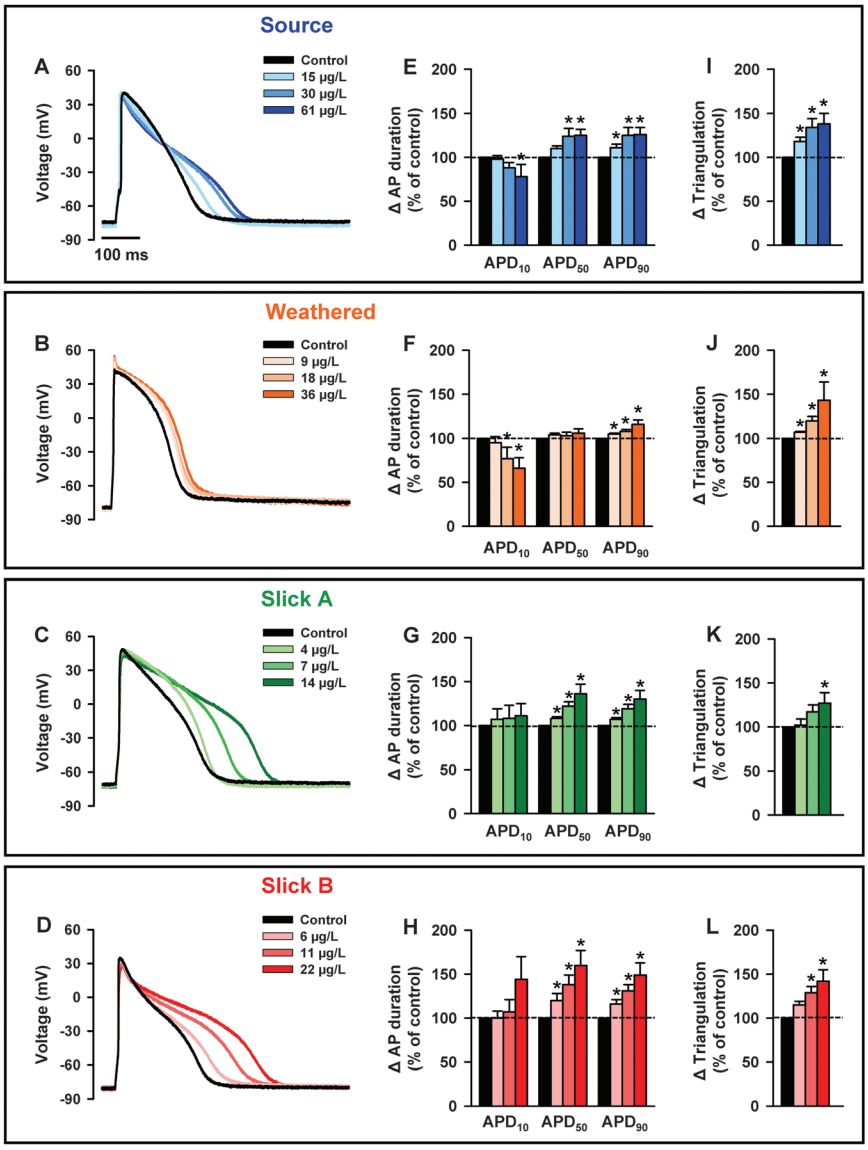


Fig. 1. Effect of oil WAFs on action potential characteristics from bluefin tuna ventricular cardiomyocytes. (**A** to **D**) Action potentials in controls (black) and with ascending concentrations of source oil (blue traces), artificially weathered (orange traces), slick A (green traces), and slick B (red traces) WAFs. (**E** to **H**) APD (expressed as a percentage of control) at 10, 50, and 90% repolarization in control (black bars) and with ascending concentrations of source ($n = 9$), artificially weathered ($n = 8$), slick A ($n = 7$), and slick B ($n = 7$). (**I** to **L**) Action potential triangulation (expressed as a percentage of control; calculated as $APD_{90} - APD_{30}$) in control (black bars) and with ascending concentrations of source, artificially weathered, slick A, and slick B. (E) to (L): Means ± SEM. *$P < 0.05$.

Overall, the effects of MC252 oil WAFs on cardiomyocyte action potentials in bluefin and yellowfin tunas were similar. However, the cardiotoxic potency of each oil sample correlated closely with the concentrations of three-ringed PAHs rather than total ∑PAHs (fig. S4), as evidenced in particular by the extensively weathered slick B sample (fig. S1 and table S1), which increased both APD and triangulation without affecting resting membrane potential or amplitude (Fig. 1 and figs. S2 and S3). In some cardiomyocytes, WAFs caused unstable action potentials after depolarizations (fig. S5B). Such arrhythmias were not observed among ventricular cells in Ringer solution over an equivalent recording duration (fig. S5A).

The functional effects of PAHs on fish cardiac rhythmicity suggest that components of crude oil interfere with EC coupling, which links electrical excitation to contraction in cardiomyocytes (*21*, *22*). Depolarization of the cardiac sarcolemmal membrane opens voltage-gated ion channels, including L-type Ca^{2+} channels, which results in Ca^{2+} entry into the cytosol. This Ca^{2+} transient triggers the release of additional Ca^{2+} from internal stores [sarcoplasmic reticulum (SR)] by means of a Ca^{2+}-induced Ca^{2+} release mechanism (CICR) (*23*–*25*). The rise in intracellular Ca^{2+} activates the contractile machinery within the cardiomyocyte. Critical for action potential repolarization are the opening and closing of voltage-gated Na^{+}, Ca^{2+}, and K^{+} channels, which renew the EC coupling process at every heartbeat. The repolarization of the tuna cardiomyocyte action potential involves a delicate balance of inward and outward ionic currents. Thus, cardiac action potential prolongation could be due to a decrease in outward current, an increase in inward current, or both. To distinguish between these possibilities, we used electrophysiological analyses (voltage clamp) to investigate the influence of slick B (as a representative oil sample of all four WAFs) on the major outward currents (I_K) and inward calcium current (I_{Ca}) in isolated cardiomyocytes.

We characterized the rapid component of the delayed potassium current (I_{Kr}) in the bluefin tuna using electrophysiological and pharmacological techniques as previously described (i.e., E-4031–sensitive current (*26*)]. In the bluefin tuna ventricular cardiomyocyte, I_{Kr} amplitude and tail current were reduced in a concentration-dependent manner in response to exposures to slick B WAF (Fig. 2, A to C) with a half maximum inhibitory concentration (IC_{50}) of 51 ± 6 μg ∑PAHs per liter and a Hill coefficient of 1.19 ± 0.11. Perfusion with surface oil (slick A) also decreased I_{Kr} in bluefin tuna ventricular cardiomyocytes (fig. S6) with a similar IC_{50} (53 ± 31 μg/liter) and Hill coefficient (1.16 ± 0.43). In yellowfin tuna, exposure of ventricular cardiomyocytes to slick B WAF also significantly decreased I_{Kr} tail currents IC_{50} = 61 ± 12 μg/liter, Hill coefficient = 0.84 ±

0.11) (fig. S7). Taken together, these data show that DWH crude oils significantly decrease I_{Kr} currents in both species. The effect of slick B (22 μg ∑PAHs per liter) on the I_{Kr} current-voltage (*I-V*) relation is shown in Fig. 2D. WAF perfusion reduced I_{Kr} amplitudes across all voltages without affecting the shape of the *I-V* curve (Fig. 2E). In addition, I_{Kr} tail currents were decreased at all voltages without shifting the curve (Fig. 2F). Bluefin tuna ventricular cardiomyocytes exposed to source oil (61 μg/liter) and yellowfin tuna ventricular cardiomyocytes exposed to slick B (22 μg/liter) showed comparable blockade of I_{Kr} (figs. S8 and S9). Thus, dissolved constituents of MC252 crude oil do not affect the voltage-dependent (gating) properties of the K^+ channel but rather inhibit outward conductance in the open state, most likely by blocking the K^+ channel pore. This mechanism would be consistent with the observed prolongation of the cardiomyocyte action potential. To confirm this, we perfused tuna ventricular cardiomyocytes with the specific I_{Kr} blocker, E-4031 (2 μM in Ringer solution). As anticipated, E4031 significantly prolonged APD_{90}, consistent with I_{Kr} shaping the repolarization of bluefin and yellowfin tuna cardiomyocytes (fig. S10).

I_{Ca} also plays a critical role in cardiomyocyte APD (*27*, *28*). Exposure to the weathered slick B surface sample significantly decreased the amplitude of I_{Ca} (Fig. 3, A to C) in a concentration-dependent manner, with an IC_{50} of 36 ± 7 μg ∑PAHs per liter and a Hill coefficient of 0.76 ± 0.13 for bluefin tuna cardiomyocytes. Note that slick B WAF also slowed the inactivation decay of I_{Ca} (Fig. 3, D and E) and thereby allowed more Ca^{2+} entry during depolarization (*27*). As indicated by quantification of Ca^{2+} entry, there was a small, but not significant, decrease in charge passing through the channel during the square pulse (Fig. 3, F and G). *I-V* relations (Fig. 3H) revealed an inhibitory effect of slick B WAF on I_{Ca} across all voltages, with a slight influence on the shape of the *I-V* curve (Fig. 3I, top), which suggested a change in the voltage-dependent properties of Ca^{2+} channels. Perfusion with the slick B WAF shifted the activation curve toward more hyperpolarized potentials (by ~7 mV) (Fig. 3I, bottom), which allowed more Ca^{2+} entry at negative potentials. As with bluefin tuna, slick B WAF (22 μg ∑PAHs per liter) also inhibited I_{Ca} in ventricular cardiomyocytes of yellowfin tuna (IC_{50} = 46 ± 5 μg/liter, Hill coefficient = 1.01 ± 0.09) (fig. S11).

To further explore the influence of DWH oil on the voltage-dependent properties of cardiac Ca^{2+} channels, I_{Ca} was measured in bluefin cardiomyocytes with Ba^{2+} as a charge carrier. In the absence of Ca^{2+}-dependent inactivation, the channel inactivates primarily via voltage-dependent processes (*27*). Similar to the effects on I_{Ca}, slick B WAF significantly decreased the amplitude of I_{Ba} but did not slow the inactivation of the current (fig. S12). This suggests that the observed change in I_{Ca} inactivation rate in response to oil is Ca^{2+}-dependent and not voltage-dependent. The decrease in I_{Ca} amplitude and slowing of inactivation might have countervailing effects on Ca^{2+} entry during the plateau phase of the action potential, as measured from action potential waveforms in response to physiological pulses (*29*).

The entry of Ca^{2+} via I_{Ca} during action potentials was similar among controls and ventricular cardiomyocytes perfused with slick B WAF (22 μg ∑PAHs per liter) for bluefin (fig. S13) and yellowfin tunas (fig. S14). Overall, the absence of an effect of crude oil on Ca^{2+} entry during a physiological pulse is attributable to (i) an increase in APD, allowing more time for Ca^{2+} entry; (ii) a leftward shift in the activation properties of Ca^{2+} channels; and (iii) a slowing of I_{Ca} inactivation. Although our findings are not sufficient to explain action potential prolongation, they show that DWH crude oil significantly decreases I_{Ca} amplitude in cardiomyocytes of tunas. L-type Ca^{2+} channels play a key role in initiating the critical CICR from SR internal stores; thus, the next series of experiments were designed to measure whole-cell Ca^{2+} cycling in isolated cardiomyocytes exposed to DWH oils.

Intracellular Ca^{2+} transients in bluefin tuna cardiomyocytes were recorded using confocal microscopy and Ca^{2+}-sensitive dye (Fluo-4). Exposures to each oil sample (source, artificially weathered source, slick A, and slick B at 30, 18, 7, and 11 μg ∑PAHs per liter, respectively) significantly decreased the Ca^{2+} transient amplitudes and slowed the decay of the Ca^{2+} transients in bluefin tuna ventricular cardiomyocytes (Fig. 4). This reduction in Ca^{2+} transient amplitudes would decrease contractility and would reduce cardiac output at the scale of the whole heart. A diminished cytosolic Ca^{2+} transient could be a consequence of reduced extracellular Ca^{2+} influx, a smaller Ca^{2+} release from SR internal stores, or both (*24*, *30*).

The direct measurements of Ca^{2+} transients in cardiomyocytes indicate there may be inhibitory effects of oil on SR Ca^{2+} release and/or reuptake. To investigate the possible SR sites of interaction, cardiomyocytes were exposed to pharmacological inhibitors of SR Ca^{2+} release channels (5 μM ryanodine) and Ca^{2+} adenosine triphosphatase (ATPase) pumps (2 μM thapsigargin) for at least 30 min before exposures to WAFs. Under pharmacological blockade, the four dis-

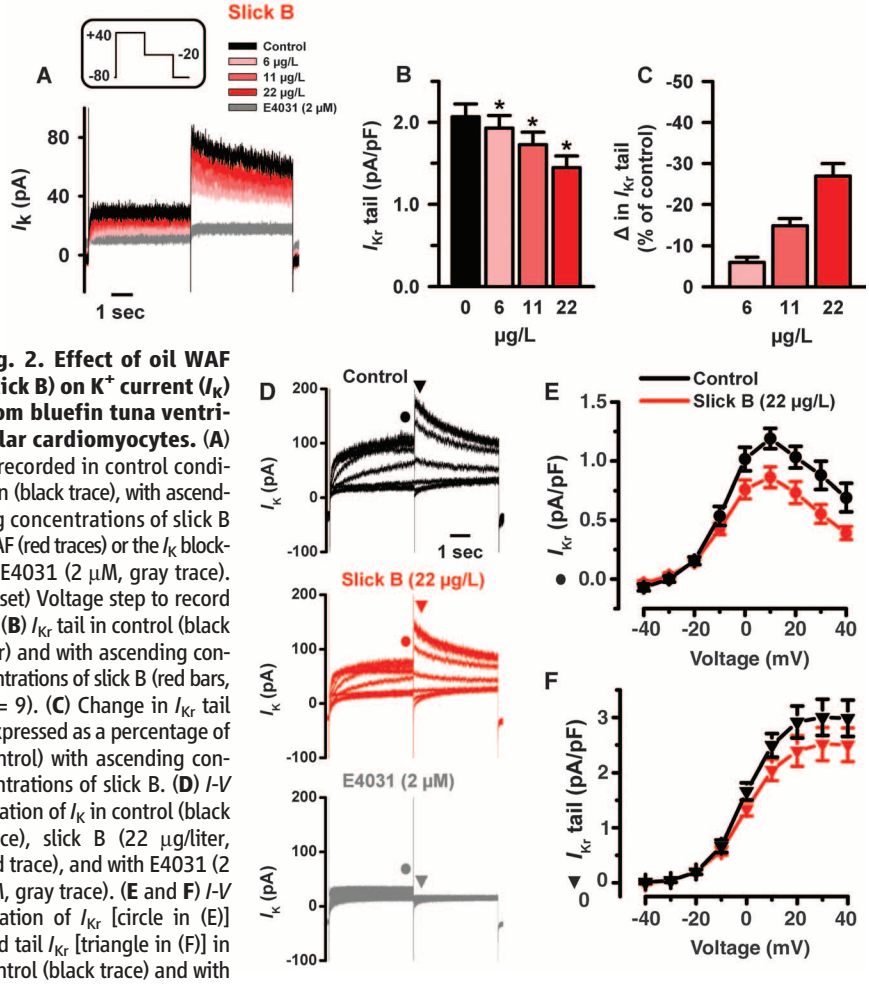


Fig. 2. Effect of oil WAF (slick B) on K^+ current (I_K) from bluefin tuna ventricular cardiomyocytes. (**A**) I_K recorded in control condition (black trace), with ascending concentrations of slick B WAF (red traces) or the I_K blocker E4031 (2 μM, gray trace). (Inset) Voltage step to record I_K. (**B**) I_{Kr} tail in control (black bar) and with ascending concentrations of slick B (red bars, n = 9). (**C**) Change in I_{Kr} tail (expressed as a percentage of control) with ascending concentrations of slick B. (**D**) *I-V* relation of I_K in control (black trace), slick B (22 μg/liter, red trace), and with E4031 (2 μM, gray trace). (**E** and **F**) *I-V* relation of I_{Kr} [circle in (E)] and tail I_{Kr} [triangle in (F)] in control (black trace) and with slick B (22 μg/liter, red trace, n = 9). (B), (C), (E), and (F): Means ± SEM. *$P < 0.05$.

tinct crude oil samples had no significant effect on the amplitude of the cytosolic Ca^{2+} transient (Fig. 4D). This indicates that the oil-induced decrease in Ca^{2+} transient amplitude observed in the absence of blockers is due to a disruption of SR Ca^{2+} release and/or reuptake from internal stores. However, these toxic effects of oil on intracellular Ca^{2+} cycling were partially offset by an additional influx of I_{Ca} via L-type Ca^{2+} channels during action potential prolongation, consistent with the I_{Ca} results in Fig. 3.

Our experimental findings provide a mechanistic underpinning for cardiac-specific physiological defects previously reported and reinforce the findings that crude oil has deleterious physiological impacts on fish hearts (*10*). Our results with crude oil are similar to the physiological effects of the antimalarial drug halofantrine, a chemical with structural similarities to three-ringed PAHs that causes K^+ channel inhibition and cardiac arrhythmias (*31*). Our results indicate compounds in DWH oil produce a cardiotoxic mechanism that have direct effects on ion channels involved in the EC coupling and cardiac

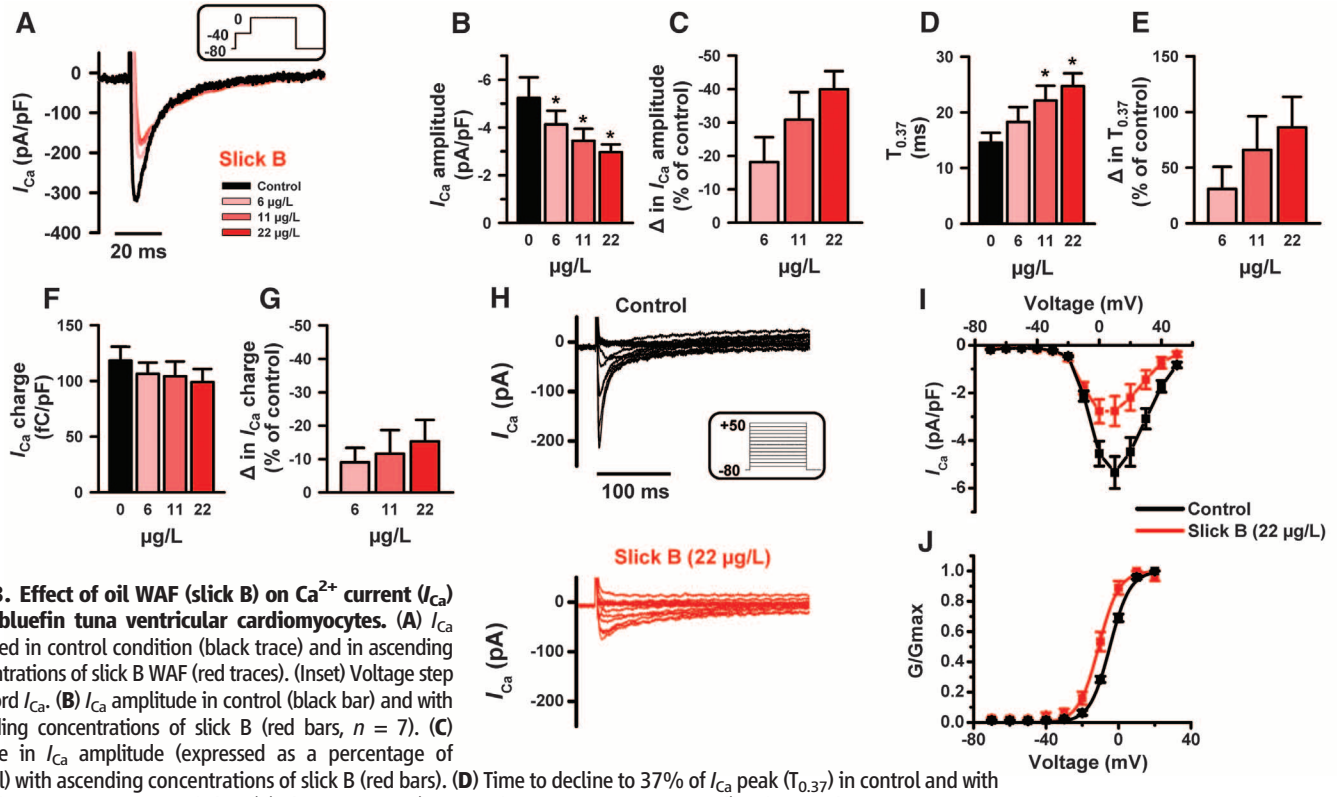


Fig. 3. Effect of oil WAF (slick B) on Ca^{2+} current (I_{Ca}) from bluefin tuna ventricular cardiomyocytes. (**A**) I_{Ca} recorded in control condition (black trace) and in ascending concentrations of slick B WAF (red traces). (Inset) Voltage step to record I_{Ca}. (**B**) I_{Ca} amplitude in control (black bar) and with ascending concentrations of slick B (red bars, $n = 7$). (**C**) Change in I_{Ca} amplitude (expressed as a percentage of control) with ascending concentrations of slick B (red bars). (**D**) Time to decline to 37% of I_{Ca} peak ($T_{0.37}$) in control and with ascending concentrations of slick B. (**E**) Change in $T_{0.37}$ (expressed as a percentage of control) with ascending concentrations of slick B (red bars). (**F**) I_{Ca} charge in control and with ascending concentrations of slick B. (**G**) Change in I_{Ca} charge (expressed as a percentage of control) with ascending concentrations of slick B WAF. (**H**) *I-V* relations of I_{Ca} in control condition (black trace) and with slick B (22 µg/liter, red trace). (Inset) Voltage step to record I_{Ca}. (**I**) *I-V* relation of I_{Ca} in control (black trace) and with slick B (22 µg/liter, red trace, $n = 7$). (**J**) Availability-voltage relation (normalized conductance G/G_{max}) of I_{Ca} in control and with slick B (22 µg/liter, $n = 7$). (B) to (G), (I), and (J): Means ± SEM. *$P < 0.05$.

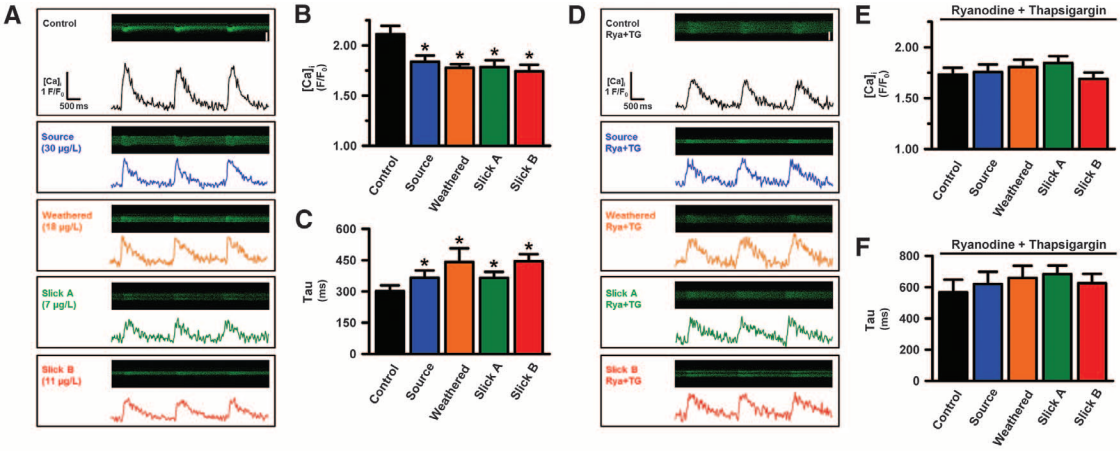


Fig. 4. Effect of oil WAFs on Ca^{2+} transients from bluefin tuna ventricular cardiomyocytes. (**A**) Ca^{2+} transients recorded in control (black trace) and with source (30 µg/liter, blue trace); artificially weathered (18 µg/liter, orange trace); slick A (7 µg/liter, green trace); and slick B (11 µg/liter, red trace) WAFs, respectively. (**B**) Ca^{2+} transients amplitude as fluorescence divided by baseline fluorescence (F/F_0) and (**C**) tau, the decay time constant of Ca^{2+} transients, in control ($n = 37$) and source ($n = 29$), artificially weathered ($n = 27$), slick A ($n = 25$), and slick B ($n = 21$), respectively. (**D**) Ca^{2+} transients recorded in ryanodine (Rya) and thapsigargin (Tg) (black trace) and with source (30 µg/liter, blue trace); artificially weathered (18 µg/liter, orange trace); slick A (7 µg/liter, green trace); and slick B (11 µg/liter, red trace), respectively. (**E**) Ca^{2+} transients amplitude (F/F_0) and (**F**) tau of decay of Ca^{2+} transients in ryanodine and thapsigargin ($n = 22$) and source ($n = 17$), artificially weathered ($n = 18$), slick A ($n = 17$), and slick B ($n = 16$), respectively. (B), (C), (E), and (F): Means ± SEM. *$P < 0.05$.

contractility of cardiomyocytes. These pathways in cardiac muscle cells are highly conserved across all vertebrates, which explains the common, canonical crude oil toxicity syndrome observed in a diversity of fish species from habitats that range from tropical freshwater (zebrafish) to boreal marine (herring).

In conclusion, the oil-induced disruption of cardiomyocyte repolarization via K^+ channel blockade and sarcolemmal and SR Ca^{2+} cycling should call attention to a previously underappreciated risk to wildlife and humans, particularly from exposure to cardioactive PAHs that are also relatively enriched in air pollution. I_{Kr} inhibition by DWH crude oil from the MC252 well was robust, and its properties are consistent with direct channel pore block. These K^+ channel targets and their unique gating properties play a critical role in cardiac action potential repolarization and are highly conserved across the animal kingdom (*20*). These results lead us to believe that PAH cardiotoxicity was potentially a common form of injury among a broad range of species during and after the DWH oil spill. The early life stages of fish and other vertebrates may have been particularly vulnerable, given that even a transient and sublethal effect of PAHs on the embryonic heartbeat can cause permanent secondary changes in heart shape and cardiac output (*32*). Moreover, the underlying ion channel currents that drive the electrical properties of cardiomyocytes in tunas and mammals (such as heart rates), are similar (*26*, *33*). Thus, we suggest the extension of our current oil toxicity results to mammalian, cardiomyocytes may be warranted to better understand PAH threats to human health.

References and Notes

1. M. G. Carls, J. P. Meador, *Hum. Ecol. Risk Assess. Int. J.* **15**, 1084–1098 (2009).
2. C. H. Peterson *et al.*, *Science* **302**, 2082–2086 (2003).
3. C. E. Boström *et al.*, *Environ. Health Perspect.* **110** (suppl. 3), 451–488 (2002).
4. J. J. Stegeman, J. J. Lech, *Environ. Health Perspect.* **90**, 101–109 (1991).
5. J. P. Incardona *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E51–E58 (2012).
6. M. G. Carls, S. D. Rice, J. E. Hose, *Environ. Toxicol. Chem.* **18**, 481 (1999).
7. R. A. Heintz, J. W. Short, S. D. Rice, *Environ. Toxicol. Chem.* **18**, 494 (1999).
8. J. P. Incardona, T. K. Collier, N. L. Scholz, *Toxicol. Appl. Pharmacol.* **196**, 191–205 (2004).
9. J. P. Incardona *et al.*, *Environ. Health Perspect.* **113**, 1755–1762 (2005).
10. J. P. Incardona *et al.*, *Environ. Sci. Technol.* **43**, 201–207 (2009).
11. J. H. Jung *et al.*, *Chemosphere* **91**, 1146–1155 (2013).
12. J. P. Incardona, T. K. Collier, N. L. Scholz, *J. Expo. Sci. Environ. Epidemiol.* **21**, 3–4 (2011).
13. J. D. Neilson, S. E. Campana, *Can. J. Fish. Aquat. Sci.* **65**, 1523–1527 (2008).
14. B. A. Block *et al.*, *Nature* **434**, 1121–1127 (2005).
15. J. R. Rooker *et al.*, *PLOS ONE* **7**, e34180 (2012).
16. Materials and methods are available as supplementary materials on *Science* Online.
17. A. R. Diercks *et al.*, *Geophys. Res. Lett.* **37**, L20602 (2010).
18. L. M. Hondeghem, L. Carlsson, G. Duker, *Circulation* **103**, 2004–2013 (2001).
19. J. M. Nerbonne, R. S. Kass, *Physiol. Rev.* **85**, 1205–1253 (2005).
20. P. Kannankeril, D. M. Roden, D. Darbar, *Pharmacol. Rev.* **62**, 760–781 (2010).
21. D. M. Bers, *Nature* **415**, 198–205 (2002).
22. M. B. Cannell, H. Cheng, W. J. Lederer, *Science* **268**, 1045–1049 (1995).
23. H. A. Shiels, E. V. Freund, A. P. Farrell, B. A. Block, *J. Exp. Biol.* **202**, 881–890 (1999).
24. H. A. Shiels, A. Di Maio, S. Thompson, B. A. Block, *Proc. Biol. Sci.* **278**, 18–27 (2011).
25. M. Vornanen, H. A. Shiels, A. P. Farrell, *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **132**, 827–846 (2002).
26. G. L. Galli, M. S. Lipnick, B. A. Block, *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **297**, R502–R509 (2009).
27. F. Brette, J. Leroy, J. Y. Le Guennec, L. Sallé, *Prog. Biophys. Mol. Biol.* **91**, 1–82 (2006).
28. F. Brette *et al.*, *Biochem. Biophys. Res. Commun.* **374**, 143–146 (2008).
29. F. Brette, L. Sallé, C. H. Orchard, *Biophys. J.* **90**, 381–389 (2006).
30. H. A. Shiels, E. White, *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **288**, R1756–R1766 (2005).
31. H. Tie *et al.*, *Br. J. Pharmacol.* **130**, 1967–1975 (2000).
32. C. E. Hicken *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7086–7090 (2011).
33. J. M. Blank *et al.*, *J. Exp. Biol.* **207**, 881–890 (2004).

Acknowledgments: We thank D. Bers and K. Ginsburg of University of California, Davis; R. Kochevar of Stanford University; and R. Ricker of NOAA for constructive advice on an earlier draft of the manuscript. We are also grateful to C. Farwell, A. Norton, and E. Estess of the Monterey Bay Aquarium, along with husbandry teams for maintenance of bluefin and yellowfin tunas and handling of oil experiments within Tuna Research and Conservation Center; the Government of Mexico for permitting the collection of tunas in their waters; T. Dunn and the crew of the F/V Shogun; and H. Shiels, G. Galli, and S. Thompson for experimental help during preliminary studies of PAH compounds. This work was funded as a contributing study to the DWH–MC252 incident Natural Resource Damage Assessment. Additional funds were provided by the Monterey Bay Aquarium Foundation and Stanford University. All experiments with crude oil were approved by Stanford University Institutional Animal Care and Use Committee procedures. All data necessary to understand this manuscript are presented in the main text or supplementary materials.

Supplementary Materials

www.sciencemag.org/content/343/6172/772/suppl/DC1
Materials and Methods
Figs. S1 to S14
Table S1
References (*34*–*41*)

3 July 2013; accepted 11 December 2013
10.1126/science.1242747

Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types

Diego Adhemar Jaitin,[1]* Ephraim Kenigsberg,[2,3]* Hadas Keren-Shaul,[1]* Naama Elefant,[1] Franziska Paul,[1] Irina Zaretsky,[1] Alexander Mildner,[1] Nadav Cohen,[2,3] Steffen Jung,[1] Amos Tanay,[2,3]†‡ Ido Amit[1]†‡

In multicellular organisms, biological function emerges when heterogeneous cell types form complex organs. Nevertheless, dissection of tissues into mixtures of cellular subpopulations is currently challenging. We introduce an automated massively parallel single-cell RNA sequencing (RNA-seq) approach for analyzing in vivo transcriptional states in thousands of single cells. Combined with unsupervised classification algorithms, this facilitates ab initio cell-type characterization of splenic tissues. Modeling single-cell transcriptional states in dendritic cells and additional hematopoietic cell types uncovers rich cell-type heterogeneity and gene-modules activity in steady state and after pathogen activation. Cellular diversity is thereby approached through inference of variable and dynamic pathway activity rather than a fixed preprogrammed cell-type hierarchy. These data demonstrate single-cell RNA-seq as an effective tool for comprehensive cellular decomposition of complex tissues.

Understanding the heterogeneous and stochastic nature of multicellular tissues is currently approached through a priori defined cell types that are used to dissect cell populations along developmental and functional hierarchies (*1*–*3*). This methodology heavily relies on enumeration of cell types and their precise definition, which can be controversial (*4*–*7*) and is based in many cases on indirect association of function with cell-surface markers (*5*–*8*). Perhaps the best understood model for cellular differentiation and diversification is the hematopoietic system. The developmental tree branching from hematopoietic stem cells toward distinct immunological functions was carefully worked out through many years of study, and effective cell-surface markers are available to quantify and sort the major hematopoietic cell types. Even in this well-explored system, however, it is becoming increasingly difficult to explain modern genome-wide and in vivo data with refined cell types' hierarchy and functions that extend beyond the classical myeloid and lymphoid cell types. For example, dendritic cells (DCs) are antigen-presenting

[1]Department of Immunology, Weizmann Institute, Rehovot 76100, Israel. [2]Department of Computer Science and Applied Mathematics, Weizmann Institute, Rehovot 76100, Israel. [3]Department of Biological Regulation, Weizmann Institute, Rehovot 76100, Israel.

*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. E-mail: amos.tanay@weizmann.ac.il (AT); ido.amit@weizmann.ac.il (IA)

cells that were originally characterized through their morphology (*9*) but are now understood to represent a highly heterogeneous group (*10*) with multiple functions, regulatory circuits, and phenotypes (*6*, *7*, *9*). Despite considerable efforts and progress by use of the marker-based approach, much of the known functional heterogeneity within the DC group is not truly compatible with any of the DC subclassification schemes (*6*, *7*, *11*). Such lack of definitive models for cell types and states is common in many fields of biology.

An attractive alternative to marker-based cellular dissection of complex tissues is to characterize in vivo cell-type compositions through unsupervised sampling and modeling of transcriptional states in single cells. This natural approach was so far difficult to implement because of many technical limitations that are being progressively alleviated with the advent of single-cell RNA sequencing (RNA-seq) (*12*–*20*). Sampling and sequencing RNA from dozens of single cells was recently used to estimate stochastic transcriptional variation in stationary cultured cells (*14*) or during a dynamic process (*12*–*14*, *16*, *19*). An unsupervised framework for dissecting transcriptional heterogeneity within complex tissues may therefore be envisioned, provided that many thousands of cells can be assayed routinely by using single-cell RNA-seq and that data from such experiments can be normalized and modeled effectively even when cells represent highly diverse cell types and states.

We developed an automated massively parallel RNA single-cell sequencing framework

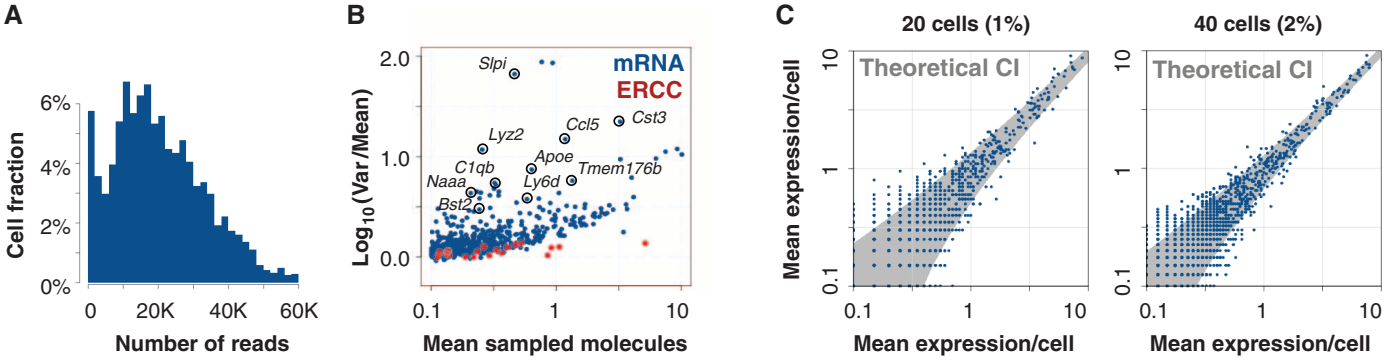
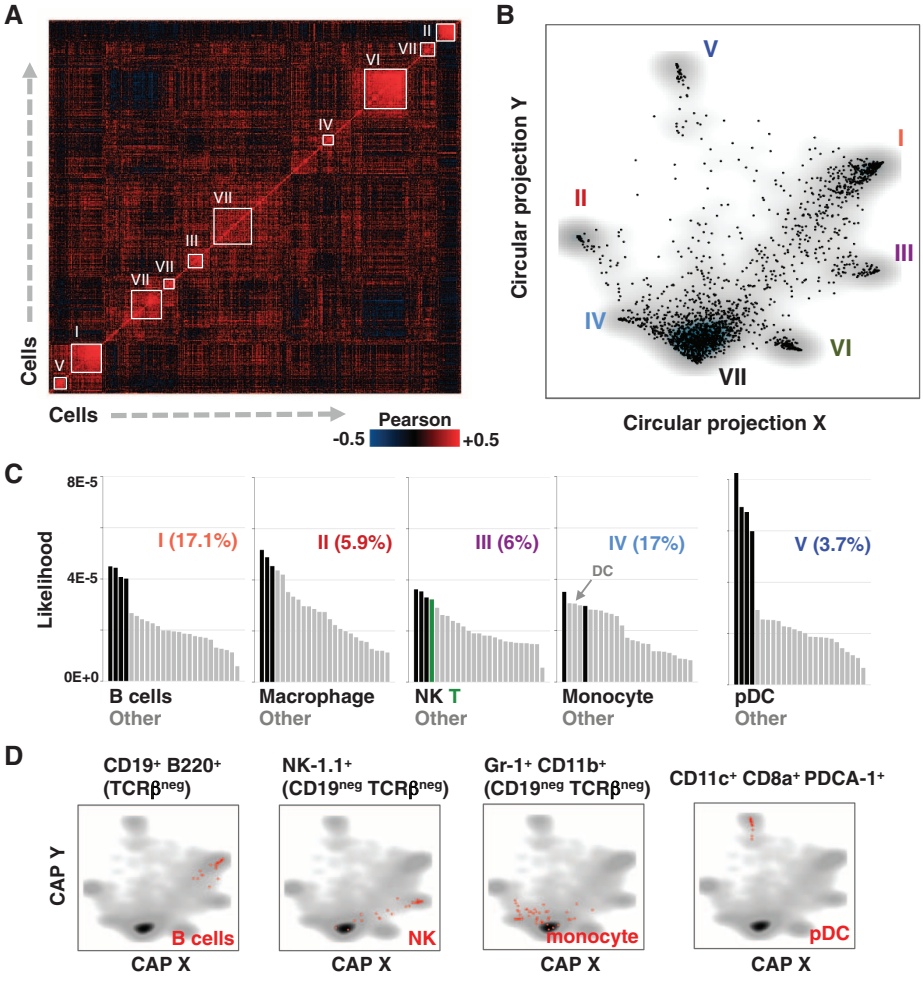


Fig. 1. Massively parallel single-cell RNA-seq. (**A**) Distribution of mapped reads per cell in a multiplexed 1536-cell experiment. (**B**) Mean and variance in mRNA (blue) and spike-in controls (red). (**C**) Mean mRNA counts in replicated pooled population of homogeneous (FACS-sorted) pDCs.

Fig. 2. Single-cell dissection of immune cell types. (**A**) Color-coded correlation matrix of single-cell mRNA profiles. Groups of strongly correlated cells that are used to initialize a probabilistic mixture model are numbered and marked with white frames. (**B**) Circular a posteriori projection (CAP) plot summarizing the predictions of the probabilistic mixture model for the CD11c$^+$ cells. Each cell is projected onto the two-dimensional sphere according to the posterior probability of its association with the model's classes. The dimensions of the CAP plot should not be interpreted linearly or as principal components. (**C**) Bar plots depicting similarities of mean RNA counts in inferred types and Immgen expression profiles. The most correlated group of Immgen profiles is colored specifically as indicated for each type. (**D**) Shown are CAP plots depicting single-cell RNA-seq data sets acquired from marker-based FACS sorting for single pDCs, B cells, NK cells, and monocytes. Sorted cells are shown in red; density of the CD11c$^+$ pool is shown in gray.



(MARS-Seq) (figs. S1 to S6) (*21*) that is designed for in vivo sampling of thousands of cells by multiplexing RNA-seq while maintaining tight control over amplification biases and labeling errors. The method is based on fluorescence-activated cell sorting (FACS) of single cells into 384-well plates and subsequent automated processing that is done mostly on pooled and labeled material, leading to a dramatic increase in throughput and reproducibility. To explore the new technique, we sequenced RNA from more than 4000 mouse spleen single cells (table S1), focusing initially on a heterogeneous cell population enriched for expression of the CD11c surface marker. We hypothesized that this strategy for cell acquisition will sample a diverse collection of splenic cell types while focusing on the challenging DC populations (*6*, *7*).

Our methodology uses three levels of barcoding (molecular-, cellular-, and plate-level tags) to facilitate molecule counting with a high degree of multiplexing. The strategy is to characterize cell subpopulations first by classifying single cells on the basis of low-depth RNA sampling and then studying transcriptional profiles at high resolution by integrating data from dozens to hundreds of cells within each unsupervised class. As shown in Fig. 1A, multiplexing 1536 cells in one sequencing lane provided an average of 22,000 aligned reads per cell, and after extensive normalization, these can be used to unambiguously define 200 to 1500 distinct RNA molecules from each cell. Our labeling and filtering scheme ensures that spiked-in technical controls show cell-to-cell variance that is compatible with the theoretical (binomial) sampling noise, comparing favorably with previously reported techniques (Fig. 1B) (*18*). This technical stability substantially increases the information content of the sampled transcriptional states, which can be directly modeled as unbiased samples of the cells' mRNA pool. In contrast to technical spike-in controls or the bulk of detected genes, we observed high cellular variance for a large number of genes, many of which are well known cell type–specific markers, suggesting that this attests for the high degree of heterogeneity within the splenic cell population (Fig. 1B) and promoting the idea of classifying cells into subpopulations on the basis of covariation of such heterogeneous markers.

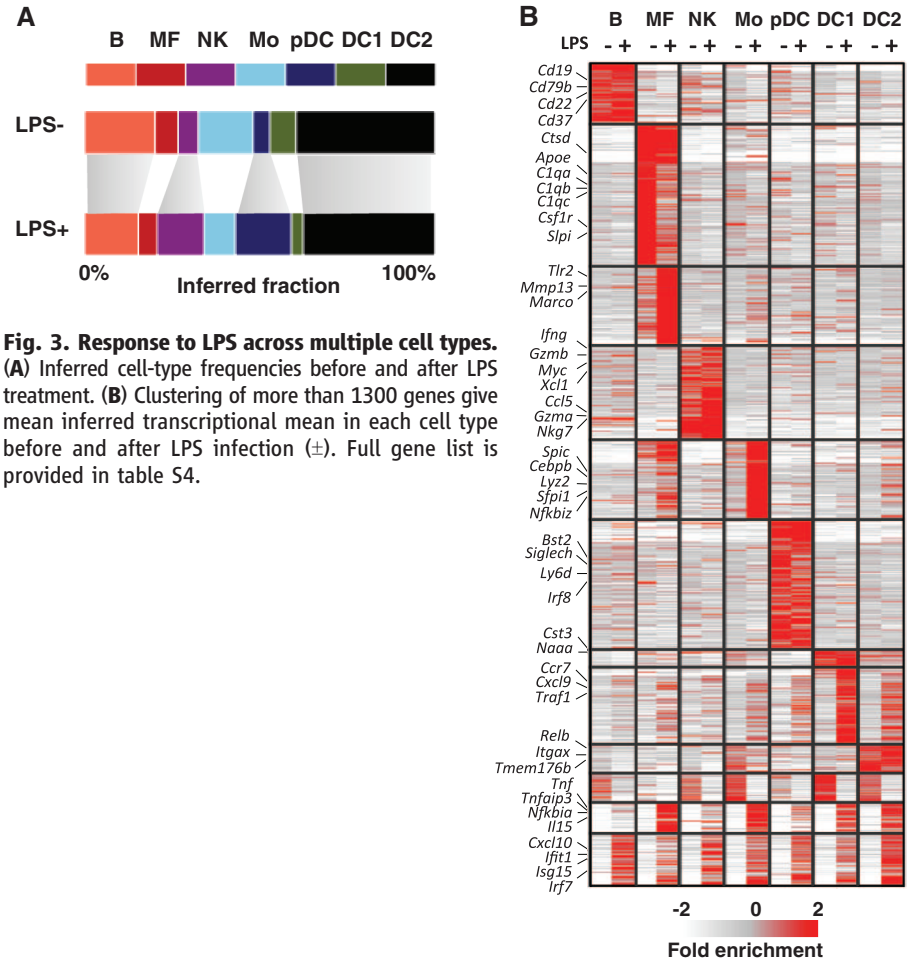


Fig. 3. Response to LPS across multiple cell types. (**A**) Inferred cell-type frequencies before and after LPS treatment. (**B**) Clustering of more than 1300 genes give mean inferred transcriptional mean in each cell type before and after LPS infection (±). Full gene list is provided in table S4.

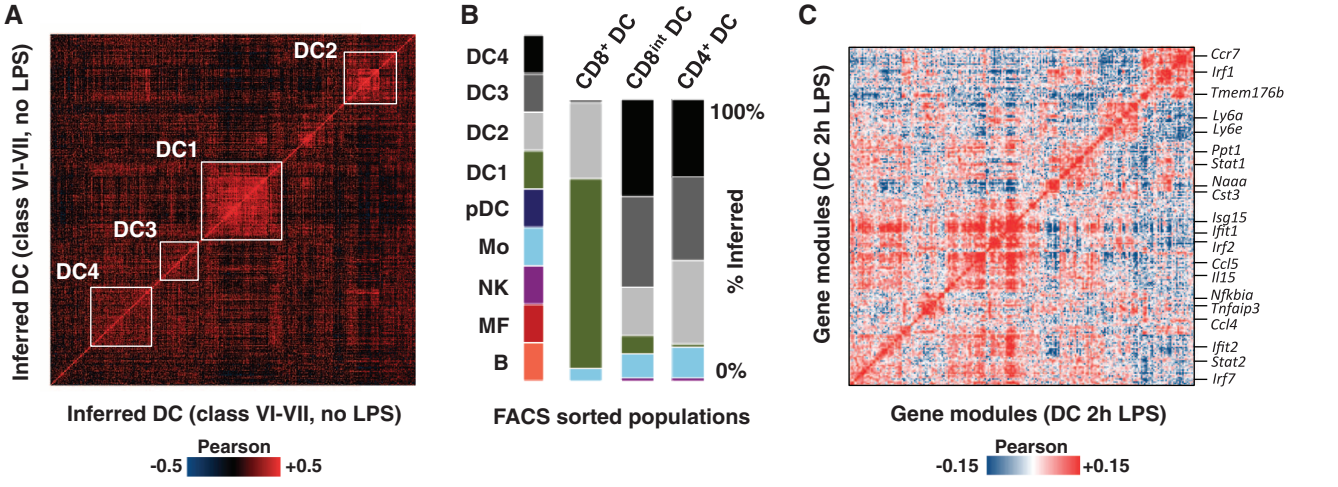


Fig. 4. Gene modules and the distribution and redistribution of DC states. (**A**) Single-cell correlation matrix for cells classified as DCs, showing detected subclasses using white frames. (**B**) Type/class distributions of single-cell RNA-seq data from three different FACS-sorted DC (CD11c enriched) populations: CD8a^{+} CD86^{+}; CD8a intermediate (int) CD86^{-}; and CD8a^{-} CD4^{+} ESAM^{+} (fig. S13A). (**C**) Gene correlation matrix is depicting potential LPS-dependent interactions between 225 genes. Key genes are indicated, with the complete list available in fig. S15.

To test how sensitive our strategy can be for characterizing the transcriptional state of subpopulations in the sample, we estimated coverage and mean mRNA molecule count reproducibility for groups of 10 to 40 single-cell profiles, representing 0.6 to 2.4% of the cells on one sequencing lane. Analysis of single-cell mRNA profiles from FACS-sorted plasmacytoid DCs (pDCs) (Fig. 1C and fig. S6) confirmed that pooling of homogeneous cell populations provides rich and highly reproducible transcriptional profiles. For a subpopulation at a frequency of 2.5%, the assay report on 1255 genes with a standard deviation of less than 35% of the mean, and on 324 genes with a standard deviation of 20% of the mean. Together, the availability of high-variance marker genes and the dynamic range provided by pooled single-cell transcriptional profiles enable unsupervised dissection and characterization of heterogeneous cell populations, opening the way for ab initio cell-type decomposition of splenic populations at a high level of detail.

We have implemented a probabilistic strategy for unsupervised classification of cells into "idealized types." Hierarchical clustering (Fig. 2A) defined seeds of highly correlated cells, leading to the initialization of a probabilistic mixture model and classification of single cells into types or families of homogeneous states. Visualization of the multiclass data by using a new circular a posteriori projection technique (Fig. 2B) represented the splenic cell population as a combination of several molecular behaviors, five of which (classes I to V) being distinctively separated from a group of more loosely defined classes (classes VI and VII). The frequencies of classes I to V range between 3.7 and 17%, allowing in all cases to infer rich transcriptional states through in silico pooling of single-cell mRNA profiles within each class. Analysis of gene enrichment (table S2 and figs. S7 and S8) and comparison of these profiles with existing transcriptional profiles of classical hematopoietic populations (www.immgen.org), unambiguously linked classes I to V to B cells, natural killer (NK) cells, macrophages, monocytes, and pDCs (Fig. 2C). The remaining classes were all linked to DCs. FACS analysis using classical surface markers confirmed our in silico estimations of the frequency of B cells and pDCs within the CD11c-enriched splenic cell population (fig. S9). Further analysis and additional single-cell quantitative polymerase chain reaction experiments confirmed that "marker" genes are robustly enriched in their relevant subpopulations (figs. S10 and S11). Using classical marker-based sorting, we further validated our approach with additional single-cell RNA-seq data from FACS-sorted B cells, NK cells, pDCs, and monocytes. Projection of the new data onto the model we generated from the splenic population showed remarkable compatibility between the traditional marker-based cell-type definition and the marker-free single-cell RNA-seq technique (Fig. 2D). Analysis of splenic cell populations therefore showcased single-cell RNA-seq as a direct and unsupervised way for identifying and characterizing subpopulations within heterogeneous tissues.

We profiled additional 1536 single cells from spleens that were exposed to lipopolysaccharide (LPS) for 2 hours (*22*), aiming to test how an immediate response to an infection-mimicking stimulus can be deciphered across the heterogeneous splenic cell population. We found that the LPS-treated cells are broadly classified into similar cell types to those observed in untreated cells, with some changes in the relative representation of different types (Fig. 3A). Using the non-LPS mixture model, we classified the non-LPS and LPS-exposed cells into classes and inferred a rich transcriptional profile within each class before and after treatment. Clustering 1575 variable genes identified groups of cell type–specific response genes (such as *Tnf* and *Marco* in macrophages and *Xcl1* and *Gzmb* in NK cells), and a large group of type I interferon response genes (*Irf7*, *Stat2*, *Ifit1*, *Cxcl10*, and hundreds more) activated pervasively in all or almost all cell types (Fig. 3B, fig. S12, and tables S3 and S4).

With thousands of samples readily available, single-cell RNA-seq is poised to go beyond the classical cell-types hierarchies that are outlined by current marker-based approaches, examining complex relations between cell subpopulations or continuous spectra of types. Analysis of 1031 single cells that were associated with DC-related classes (VI and VII) in our unsupervised $CD11c^+$ model (Fig. 4A) indicated that although 15% of these cells (class DC1) are strongly linked together, the remaining bulk of DCs could not be organized along a clear clustering hierarchy (*11*). Nevertheless, we found strong support for substantial internal organization within the remaining DC population (DC2 to DC4) (table S5), including a group of cells coexpressing *Relb*, *Nfkbia*, and additional associated genes (DC2) (fig. S13). More generally, we have identified several gene modules that represent combinatorial pathway activity within the DC bulk (fig. S14), indicating that despite the lack of a clear hierarchy, the DC cell population is governed by a high degree of transcriptional organization. Additional single-cell sequencing of $CD8^+$ $CD86^+$, $CD8^{int}$ $CD86^-$, and $CD4^+$ FACS-sorted populations (Fig. 4B) showed that this organization can be approached to a limited extent with existing marker-based classification. Remarkably, exposure to LPS reorganizes the DC population substantially, with a large number of gene modules being activated in a highly heterogeneous fashion (Fig. 4C and fig. S15). According to our analysis, certain specific $CD4^+$ DC subpopulations are activating the Irf4, tumor necrosis factor, and transforming growth factor–β pathways (fig. S16 and table S6), whereas other pathways (such as Irf7) are activated pervasively (table S5). This combinatorial activity of pathways within the LPS-exposed DC pool is not represented in preexisting DC subtypes according to our data. Committed and developmentally stable myeloid and lymphoid cell types maintain their identity during immediate response to infection while responding through generic and cell type–specific pathways. These pathways create substantial cell-to-cell variance and define new cell subpopulations within each of these cell types (fig. S17), forming diversity that may have functional implications. Observation of transcriptional subpopulations, however, does not necessarily imply the existence of further committed and preprogrammed cell subtype hierarchy.

We presented this framework for broad sampling of single-cell transcriptional states from tissues and demonstrated how it can be used to dissect complex functions in a bottom-up fashion. MARS-seq can be readily applied to tissues and organs in normal and disease states to redefine their cell-type and cell-state compositions and link it to detailed genome-wide transcriptional profiling. Given the inherent stochasticity and heterogeneity of multicellular tissues, this approach can prove essential for understanding how in vivo biological function emerges from complex cell ensembles.

References and Notes

1. M. Acar, J. T. Mettetal, A. van Oudenaarden, *Nat. Genet.* **40**, 471–475 (2008).
2. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, *Science* **297**, 1183–1186 (2002).
3. R. N. Germain, *Nat. Immunol.* **13**, 902–906 (2012).
4. S. C. Bendall *et al.*, *Science* **332**, 687–696 (2011).
5. B. M. Bradford, D. P. Sester, D. A. Hume, N. A. Mabbott, *Immunobiology* **216**, 1228–1237 (2011).
6. F. Geissmann, S. Gordon, D. A. Hume, A. M. Mowat, G. J. Randolph, *Nat. Rev. Immunol.* **10**, 453–460 (2010).
7. D. A. Hume, *J. Leukoc. Biol.* **89**, 525–538 (2011).
8. M. C. Nussenzweig, R. M. Steinman, M. D. Witmer, B. Gutchinov, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 161–165 (1982).
9. R. M. Steinman, Z. A. Cohn, *J. Exp. Med.* **137**, 1142–1162 (1973).
10. L. Bar-On *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14745–14750 (2010).
11. D. Hashimoto, J. Miller, M. Merad, *Immunity* **35**, 323–335 (2011).
12. T. Hashimshony, F. Wagner, N. Sher, I. Yanai, *Cell Rep.* **2**, 666–673 (2012).
13. S. Islam *et al.*, *Nat. Protoc.* **7**, 813–828 (2012).
14. D. Ramsköld *et al.*, *Nat. Biotechnol.* **30**, 777–782 (2012).
15. Y. Sasagawa *et al.*, *Genome Biol.* **14**, R31 (2013).
16. A. K. Shalek *et al.*, *Nature* **498**, 236–240 (2013).
17. F. Tang, K. Lao, M. A. Surani, *Nat. Methods* **8**, (Suppl), S6–S11 (2011).
18. A. R. Wu *et al.*, *Nat. Methods* **11**, 41–46 (2014).
19. Q. Deng, D. Ramsköld, B. Reinius, R. Sandberg, *Science* **343**, 193–196 (2014).
20. S. Islam *et al.*, *Nat. Methods*, published online 22 December 2013 (10.1038/nmeth.2772).
21. Materials and methods are available as supplementary materials on *Science* Online.
22. I. Amit *et al.*, *Science* **326**, 257–263 (2009).

Acknowledgments: Research was supported by the European Research Council and Israel Science Foundation (1782/11, 1050/12) grants to I.A. and A.T. RNA-seq data are deposited in Gene Expression Omnibus, accession no. GSE54006.

Supplementary Materials

www.sciencemag.org/content/343/6172/776/suppl/DC1
Materials and Methods
Figs. S1 to S17
Tables S7 to S9
References (*23–25*)
Tables S1 to S6

25 October 2013; accepted 22 January 2014
10.1126/science.1247651

Leaf Shape Evolution Through Duplication, Regulatory Diversification, and Loss of a Homeobox Gene

Daniela Vlad,[1] Daniel Kierzkowski,[2] Madlen I. Rast,[2] Francesco Vuolo,[2] Raffaele Dello Ioio,[2] Carla Galinha,[1,3] Xiangchao Gan,[2] Mohsen Hajheidari,[2] Angela Hay,[2] Richard S. Smith,[2] Peter Huijser,[2] C. Donovan Bailey,[4] Miltos Tsiantis[1,2]*

In this work, we investigate morphological differences between *Arabidopsis thaliana*, which has simple leaves, and its relative *Cardamine hirsuta*, which has dissected leaves comprising distinct leaflets. With the use of genetics, interspecific gene transfers, and time-lapse imaging, we show that leaflet development requires the REDUCED COMPLEXITY (RCO) homeodomain protein. RCO functions specifically in leaves, where it sculpts developing leaflets by repressing growth at their flanks. *RCO* evolved in the Brassicaceae family through gene duplication and was lost in *A. thaliana*, contributing to leaf simplification in this species. Species-specific *RCO* action with respect to its paralog results from its distinct gene expression pattern in the leaf base. Thus, regulatory evolution coupled with gene duplication and loss generated leaf shape diversity by modifying local growth patterns during organogenesis.

Understanding how form evolves requires identifying the genetic changes underlying morphological variation between species and elucidating how those changes influence morphogenesis. In this work, we investigate this problem in the case of angiosperm leaves. Both simple and dissected leaves initiate as entire structures at the flanks of the pluripotent shoot apical meristem (*1*). However, in dissected leaves, elaboration of lateral growth axes after leaf initiation generates leaflets (*2*, *3*). So far, no gene has been identified that expresses specifically at developing leaflets and is sufficient to convert a simple leaf into a more complex one. Rather, a key factor in leaflet formation is the reactivation of meristem genes in leaves, which suggests that evolutionary differences in leaf complexity arose through modification of activity of genes that influence meristem function (*4*). This view is also consistent with the evolutionary origin of leaves from branched shoots (*5*). To determine whether leaflet-specific factors exist, we conducted a genetic screen in *Cardamine hirsuta*, a dissected-leaf relative of the simple-leaf reference plant *Arabidopsis thaliana* (Fig. 1, A and B) (*6*). If such genes exist, then loss of their function might prevent leaflet formation without perturbing meristem function.

The recessive mutant, *rco* (*reduced complexity*), converts the *C. hirsuta* adult leaf from dissected into a simple lobed leaf (Fig. 1, B and C) without affecting the number and positioning of leaves. Thus, *RCO* is required for leaflet development but not leaf initiation (fig. S1, A to G). *RCO* is a homeobox gene present in the genome

[1]Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK. [2]Department of Comparative Development and Genetics, Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Cologne, Germany. [3]Department of Biological and Medical Sciences, Faculty of Health and Life Sciences, Oxford Brookes University, Headington, Oxford OX3 0BP, UK. [4]Department of Biology, New Mexico State University, Las Cruces, NM 88003, USA.

*Corresponding author. E-mail: tsiantis@mpipz.mpg.de

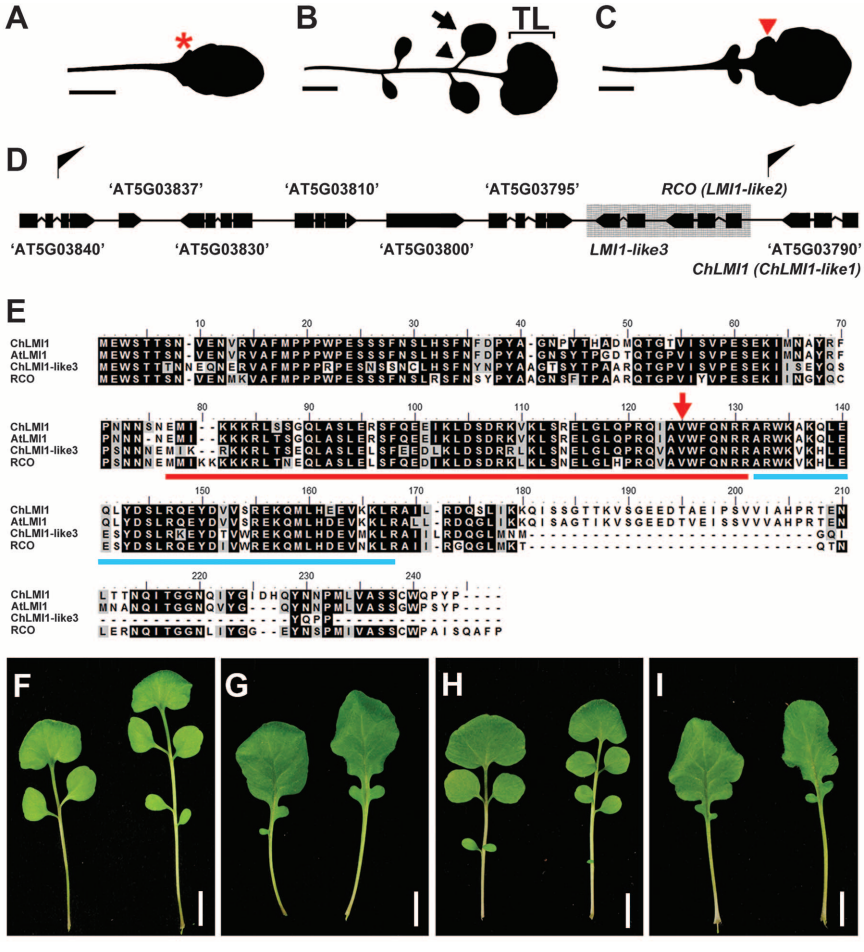


Fig. 1. Mapping of *RCO* and complementation of the *rco* mutant. (**A** to **C**) Silhouettes of an *A. thaliana* simple leaf (A) with small marginal protrusions called serrations (red asterisk); a *C. hirsuta* dissected leaf (B) with lateral leaflets (black arrow) borne by petiolules (black arrowhead) and a terminal leaflet (TL); and a *C. hirsuta rco* mutant leaf (C), in which leaflets are converted to lobes (red arrowhead). (**D**) Schematic representation of genes in the *RCO* genetic interval predicted by sequence similarity to the *A. thaliana* genome. *C. hirsuta* orthologs are indicated with inverted commas; interval borders are marked with black flags; and shading indicates genes absent in *A. thaliana.* (**E**) Alignment of proteins encoded by *LMI1* and *LMI1-like* genes in *A. thaliana* and *C. hirsuta*, respectively. The red arrow marks the last amino acid residue (Val^{124}) of the truncated protein encoded by the *rco* mutant transcript; horizontal lines indicate the homeobox (red line) and homeobox-associated leucine zipper (blue line) domains. Amino acid residues are shown as single-letter abbreviations (*25*). (**F** to **I**) (F). Complementation to WT morphology (F) of the *rco* mutant phenotype (G) by transgenic expression of *RCO* (*rco; RCOg*) (H) but not *ChLMI1* (*rco; ChLMI1g*) (I) genomic fragments. Fourth and fifth leaves are shown. Scale bars in (A to C) and (F to I), 1 cm.

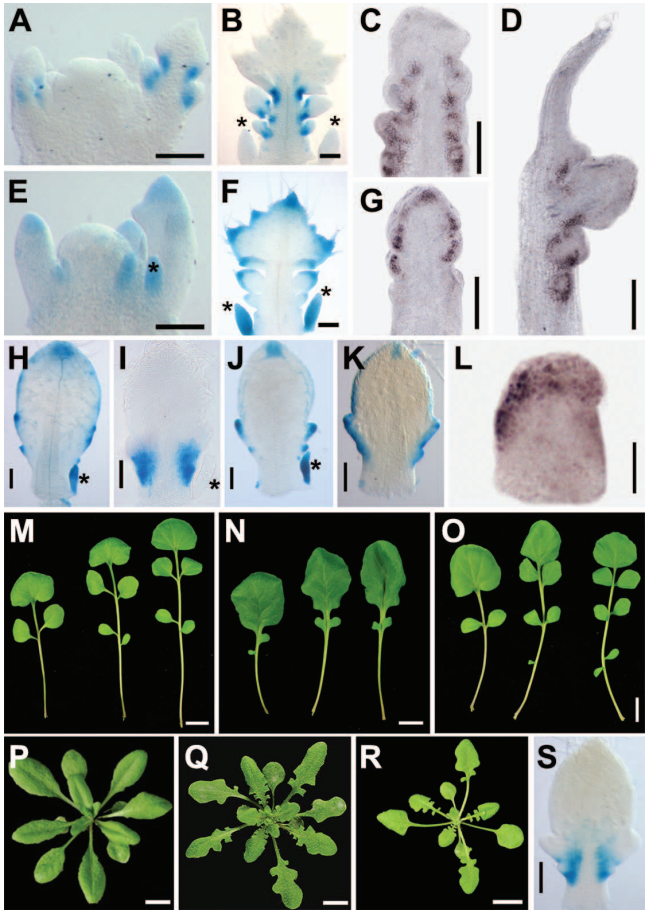


Fig. 2. The *RCO* expression pattern underlies its ability to promote leaf complexity. (**A** to **G**) Complementary expression patterns of *RCO* and *ChLMI1* shown by *RCO::GUS* (A and B) and *ChLMI1::GUS* (E and F) reporter gene analysis and by RNA localization of *RCO* (C and D) and *ChLMI1* (G) at the shoot apex and fifth leaf of *C. hirsuta*. (**H** to **K**) GUS staining in *AtLMI1::GUS* (H), *RCO::GUS* (I), *ChLMI1::GUS* (J), and *AaLMI1::GUS* (K) *A. thaliana* leaves. (**L**) *AaLMI1* RNA localization in vegetative *Aethionema arabicum* leaf. (**M** to **O**) Phenotype of leaves four to six of *C. hirsuta* wild type (M), *rco* mutant (N), and *rco* mutant complemented with an *RCO::ChLMI1* transgene (O). (**P** to **R**) Rosettes of *A. thaliana* wild type (P) and transgenic *RCOg* (Q) and *AlRCOg* (R) plants. (**S**) GUS staining in *AlRCO::GUS A. thaliana* leaf. Asterisks indicate stipules. RNA localization images are minimal projections. Scale bars, 100 μm in (A) to (L) and (S); 1 cm in (M) to (R).

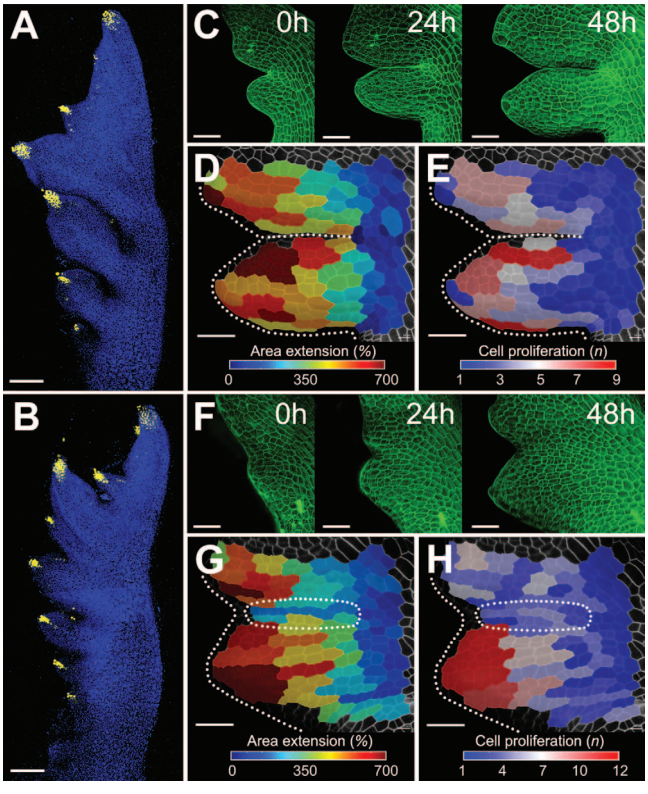


Fig. 3. *RCO* represses growth at the boundary between leaflets and does not influence auxin-based patterning. (**A** and **B**) *DR5::VENUS* expression (yellow) and chlorophyll autofluorescence signal (blue) in the seventh leaf of WT (A) and *rco* (B) *C. hirsuta*. (**C** to **H**) Time-lapse of developing lateral leaflets in WT (C to E) and *rco* (F to H) fifth leaves. Propidium iodide–stained leaf cells (green) are shown in (C) and (F) for each time point. Heat maps of relative surface area increase over 48 hours of growth (color bar: percentage increase) for lateral leaflets are shown in (D) and (G). Heat maps of cell proliferation over 48 hours [color bar: number of cells (*n*) originating from one initial cell] for lateral leaflets are shown in (E) and (H). White dotted lines denote leaf margins; white dotted rings indicate areas with excess growth and cell proliferation in the *rco* mutant. Scale bars, 100 μm in (A) and (B); 30 μm in (C) to (H).

of *C. hirsuta* but absent in *A. thaliana*, and the *rco* phenotype is caused by reduced gene function (Fig. 1, D to H, and fig. S1, H and I). Moreover, *RCO* is part of a tandem gene triplication, but the *A. thaliana* genome has only one of these genes, *LMI1* (*LATE MERISTEM IDENTITY 1*), previously identified as a floral regulator (Fig. 1D) (*7*). Phylogenetic analysis revealed that *RCO* arose from the duplication of *LMI1*-type sequences within the Brassicaceae after the divergence of *Aethionema* and before the last common ancestor of *Arabidopsis* and *Brassica* (fig. S2). *RCO* duplicated further, yielding a cluster of three genes in *C. hirsuta*, *A. lyrata*, and *Capsella rubella*, though *LMI1-like3* is not expressed in *C. hirsuta*, and concerted evolution may have influenced this gene cluster (fig. S1, I; fig. S2; and fig. S3). Secondary loss of *RCO*-type sequences left *LMI1* as a singleton in the *A. thaliana* lineage (fig. S3). We found that simply increasing the dose of *LMI1*-type protein is not sufficient to suppress *rco* or increase leaf complexity in *A. thaliana* (Fig. 1I and fig. S4), suggesting that RCO function in leaflet development is borne out of its specific gene expression properties and/or protein function. Thus, *RCO* is a taxonomically restricted gene underlying a species-specific trait that distinguishes two species that diverged relatively recently (*8*–*10*).

We used reporter gene assays and RNA in situ hybridization to determine the expression pattern of *RCO* (Fig. 2, A to D, and fig. S5, A to F). *RCO* expression is restricted to developing leaves, in two small regions at the base of terminal and lateral leaflets, and is absent from the meristem-leaf boundary (Fig. 2, A to D; fig. S5, A to F; and fig. S6, A to C). By contrast, *ChLMI1* is expressed in a near-complementary pattern to *RCO* in terminal and lateral leaflet margins and also in stipules and flowers, similar to its *A. thaliana* ortholog (Fig. 2, E to H; fig. S5G; and fig. S6, D to F) (*7*). Thus, *RCO* activity at the base of leaflets is required for leaflet formation, and its domain of expression distinguishes *RCO* from its paralog *ChLMI1*. The orthologous genes *ChLMI1* and *AtLMI1* share comparable expression domains, indicating that the distinct expression pattern of *RCO* reflects regulatory diversification from *ChLMI1* (fig. S7). Consistent with this view, *RCO* 5′ regulatory sequences drive reporter gene expression in more proximal and internal regions of the *A. thaliana* leaf than those of *ChLMI1*, *AtLMI1*, and *Aethionema arabicum LMI1* (Fig. 2, H to L; fig. S8, A to D; and fig. S9). Because *LMI1* transcripts do not accumulate at the leaf base of *A. arabicum*, which is an early divergent member of the Brassicaceae, these comparisons indicate that the *RCO* expression pattern represents an evolutionary novelty that arose after gene duplication through neofunctionalization (*11*).

To evaluate whether protein sequence specificity also contributes to *RCO* function, we expressed *ChLMI1* under the regulatory regions of *RCO* and found that this transgene complemented the *rco* mutant phenotype (Fig. 2, M

to O). Thus, RCO and ChLMI1 proteins are functionally equivalent in this developmental context, and these results suggest that the species-specific action of *RCO* in leaflet formation reflects diversification of gene expression from its paralog *ChLMI1*. The absence of *RCO*, a leaf complexity gene, from the *A. thaliana* genome suggests that *RCO* plays a key role in shaping leaf diversity. If so, and given that the simple leaf shape of *A. thaliana* is evolutionarily derived (*12*), introducing *RCO* into *A. thaliana* should reverse some of the effects of evolution and increase leaf complexity. As predicted, *A. thaliana* transgenic lines carrying a *C. hirsuta RCO* genomic clone (*RCOg*) produced deep lobes never seen in the wild type, suggesting that *RCO* is sufficient to increase *A. thaliana* leaf complexity (Fig. 2, P and Q, and fig. S10). These transgenic lines lacked pleiotropic effects, consistent with a specific role of *RCO* in leaf margin morphogenesis (fig. S10). Furthermore, introducing a genomic clone of the *RCO* ortholog of *A. lyrata* (*AlRCOg*), a lobed leaf species, into *A. thaliana* also produced lobed leaves (Fig. 2R). This phenotype is likely to result from AlRCO protein activity in the leaf base because the *AlRCO::GUS* reporter is expressed in a similar domain to the *C. hirsuta RCO* gene (Fig. 2S and fig. S8D). The finding that *RCO::GUS* and *AlRCO::GUS* are expressed in the *A. thaliana* leaf base, despite the absence of an *RCO*-type gene, suggests that at least part of the ancestral regulatory landscape that promotes leaf complexity through *RCO* activation has been retained in this species and highlights the importance of the leaf base as an organizing region for leaf growth (*13*). Collectively, these results indicate that localized *RCO* action is a key factor in determining leaf shape complexity in the Brassicaceae, and loss of this gene contributed to leaf simplification in *A. thaliana*.

We next considered how *RCO* regulates leaflet development. Auxin is required for leaflet development, and failure to organize discrete auxin activity maxima along the leaf margin reduces leaf complexity (*3*, *14*). However, the auxin activity marker *DR5* and the auxin efflux carrier PINFORMED1 were similarly expressed in leaf primordia of *rco* and the wild type, with sequential discrete auxin activity maxima forming at the leaf margin in both lobes and leaflets (Fig. 3, A and B, and fig. S11). Thus, RCO does not contribute to the establishment or maintenance of local auxin activity maxima that control leaflet initiation. *RCO* is expressed immediately adjacent to leaflet primordia (Fig. 2, A to D, and fig. S5, A to F), so drawing on classic ideas of leaf shape patterning (*15*), we hypothesized that *RCO* may influence growth locally to enable separation of individual leaflets. To test this hypothesis, we used MorphoGraphX software to analyze time-lapse images of leaflet growth (*16*) (Fig. 3, C to H). In wild-type (WT) plants, we observed that cell expansion and proliferation is inhibited in the marginal region between initiating leaflets (Fig. 3, D and E, and fig. S12, A to C). By contrast, in the *rco* mutant these cells proliferate and grow faster than in the wild type, filling up the space between leaflets (Fig. 3, G and H, and fig. S12, A to C). Conversely, cells within leaflets grow and proliferate fast in both genotypes (Fig. 3, C to H, and fig. S12, A to C). Thus, *rco* mutant leaflets initiate and grow in a comparable way to the wild type but fail to separate properly from each other due to incomplete growth repression at their boundaries, resulting in a simplified leaf. We also analyzed *A. thaliana* leaves expressing *RCOg* and found significant repression of growth and cell proliferation adjacent to emerging serrations that transformed these small protrusions into deep lobes (fig. S12, D to F, and fig. S13). Thus, *RCO* contributes to growth repression between adjacent leaflets.

We found that *ChLMI1* rescues the *rco* phenotype when expressed from the *RCO* promoter and the smooth leaf margin of *A. thaliana lmi1* when expressed in its genomic context (Fig. 2O and fig. S14). Growth derepression probably contributes to this *lmi1* phenotype and that of a classical pea mutant, in which mutation in an unusual *LMI1*-like gene converts filamentous leaf tendrils into laminate leaflets (*7*, *17*). Additionally, both *LMI1* and *RCO* repress growth when overexpressed in *A. thaliana* (fig. S15). To understand the degree of conservation of the growth-regulating function of *LMI*- and *RCO*-type genes during crucifer evolution, we evaluated the ability of selected genes from the phylogeny (fig. S2) to modify *A. thaliana* leaf shape when expressed under the control of the *RCO* promoter. With the exception of the *RCO-B* gene of *C. rubella*, all sequences assayed produced deep lobes in *A. thaliana* leaves (fig. S16). Because these sequences included *LMI1* from the early divergent crucifer *Aethionema* and the basal eudicot *Aquilegia*, it follows that the potential for LMI1 protein to repress growth evolved before the appearance of *RCO* in the Brassicaceae, and probably before the split of eudicots from other seed plants. Within the Brassicaceae, evolution of *RCO* through gene duplication created a new version of these growth repressors that is active in the morphogenetically important leaf base, thus contributing to diversification of leaf shape (Fig. 4).

Leaflet formation in *C. hirsuta* requires the *RCO* homeobox gene that arose through gene duplication and is only expressed at initiating leaflets. Thus, evolutionary changes in leaf complexity can arise through factors distinct to those acting at the shoot apical meristem. RCO does not appear to act through the well-characterized auxin-based patterning that underpins leaflet and serration formation (Fig. 3, A and B, and fig. S11), or to regulate transcription of *CUP-SHAPED COTYLEDON (CUC)* or *KNOTTED1-like homeobox (KNOX)* genes that influence this patterning (fig. S17) (*3*, *18*, *19*). One possibility is that *RCO* acts parallel to or downstream of these genes to regulate leaf complexity, thus providing a means to uncouple growth and patterning inputs during evolution. *RCOg* expression in *A. thaliana* leaves transforms serrations into deep lobes by locally repressing growth adjacent to each serration (fig. S13); however, complete transformation into leaflets may require other genes that are active in *C. hirsuta* but not in *A. thaliana* leaves, such as *CUC1* or *KNOX* genes (*20*, *21*). *RCO*-mediated shape diversification follows a broad principle of regulatory evolution: that morphological diversity is driven by changes in gene expression that minimize fitness costs by circumventing pleiotropy (*22*, *23*). It will be interesting to explore whether the high tendency toward gene duplication in plants (*24*) was a major driver for

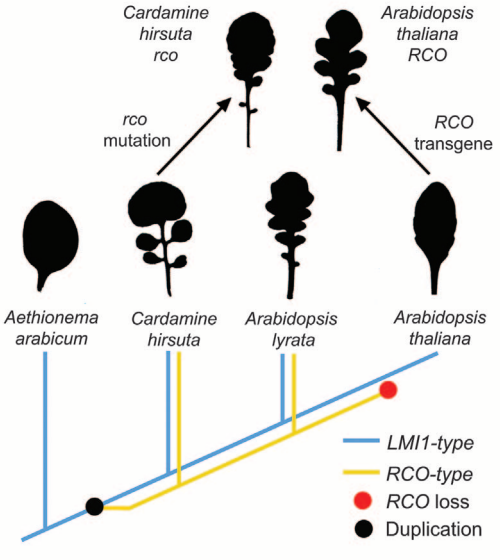


Fig. 4. Evolution of *RCO* and its consequences for diversification of leaf shape in crucifers (based on phylogeny presented in figs. S2 and S3). The genome of *Aethionema arabicum*, a simple-leaf early divergent crucifer, contains a single *LMI1*-type gene. *RCO*-type genes arose from duplication of an *LMI1*-type gene after the divergence of *Aethionema* from core Brassicaceae. The genomes of dissected-leaf *C. hirsuta* and lobed-leaf *A. lyrata* contain both *LMI1*-type and *RCO*-type genes. In the lineage that gave rise to *A. thaliana*, the *RCO*-type gene was secondarily lost, contributing to evolution of a simple leaf. Consistent with a role for *RCO* in promoting leaf complexity, removing its activity from *C. hirsuta* in the *rco* mutant leads to leaf simplification, whereas introducing *RCO*-type function into *A. thaliana* results in a more complex leaf. Silhouettes are from adult leaves and are not to scale. *A. lyrata* and *C. hirsuta* also contain *LMI1-like3*, a third copy in this gene cluster that arose through duplication of *RCO*-type but is not shown in this diagram because it is not expressed in *C. hirsuta* (fig. S1) and has not been characterized functionally.

evolution of regulatory variants that underlie trait diversification between species, as we have shown here for *RCO*.

References and Notes

1. I. Efroni, Y. Eshed, E. Lifschitz, *Plant Cell* **22**, 1019–1032 (2010).
2. N. Ori *et al.*, *Nat. Genet.* **39**, 787–791 (2007).
3. M. Barkoulas, A. Hay, E. Kougioumoutzi, M. Tsiantis, *Nat. Genet.* **40**, 1136–1141 (2008).
4. S. Hake, N. Ori, *Nat. Genet.* **31**, 121–122 (2002).
5. S. K. Floyd, J. L. Bowman, *J. Plant Res.* **123**, 43–55 (2010).
6. Materials and methods are available as supplementary materials on *Science* Online.
7. L. A. Saddic *et al.*, *Development* **133**, 1673–1682 (2006).
8. M. A. Beilstein, N. S. Nagalingum, M. D. Clements, S. R. Manchester, S. Mathews, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18724–18728 (2010).
9. T. L. P. Couvreur *et al.*, *Mol. Biol. Evol.* **27**, 55–71 (2010).
10. C. D. Bailey *et al.*, *Mol. Biol. Evol.* **23**, 2142–2160 (2006).
11. S. Ohno, *Evolution by Gene Duplication* (Springer, Berlin, 1970).
12. P. Piazza *et al.*, *Curr. Biol.* **20**, 2223–2228 (2010).
13. E. E. Kuchen *et al.*, *Science* **335**, 1092–1096 (2012).
14. D. Koenig, E. Bayer, J. Kang, C. Kuhlemeier, N. Sinha, *Development* **136**, 2997–3006 (2009).
15. W. Whaley, C. Whaley, *Am. J. Bot.* **29**, 195 (1942).
16. D. Kierzkowski *et al.*, *Science* **335**, 1096–1099 (2012).
17. J. Hofer *et al.*, *Plant Cell* **21**, 420–428 (2009).
18. Y. Berger *et al.*, *Development* **136**, 823–832 (2009).
19. G. D. Bilsborough *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3424–3429 (2011).
20. T. Blein *et al.*, *Science* **322**, 1835–1839 (2008).
21. A. Hay, M. Tsiantis, *Nat. Genet.* **38**, 942–947 (2006).
22. B. Prud'homme, N. Gompel, S. B. Carroll, *Proc. Natl. Acad. Sci. U.S.A.* **104** (suppl. 1), 8605–8612 (2007).
23. D. Stern, V. Orgogozo, *Science* **323**, 746–751 (2009).
24. M. Freeling, *Annu. Rev. Plant Biol.* **60**, 433–453 (2009).
25. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

Acknowledgments: We thank A. Hudson, G. Coupland, Y. Eshed, and N. Gompel for critical comments on the manuscript; P. Sarchet and S. Broholm for mutant screening and providing the mapping cross; I. A. Al-Shehbaz for helpful discussions on leaf shape in *Aethionema*; S. Hake, M. Aida, G. Coupland, S. Sampson, and D. Wagner for materials; and R. Berndtgen, S. Höhmann, E. Rabbinowitsch, J. Baker, and Z. Lewis for technical assistance. This work was supported by Biotechnology and Biological Sciences Research Council grants BB/H011455/1 (M.T.) and BB/H006974/1 (M.T. and A.H.), Deutsche Forschungsgemeinschaft 'Adaptomics' grant TS 229/1-1 (M.T. and A.H.) and grant SFB 680 (M.T.), the Gatsby Charitable Foundation (M.T.), Human Frontier Science Program grant RGP0047/2010 (M.T.), a core grant from the Max Planck Society (M.T.), and NSF Plant Genome Research Mid-Career Investigator Award 1238731 (C.D.B.). M.T. also acknowledges support of the Cluster of Excellence on Plant Sciences. The data reported in this paper are tabulated in the supplementary materials. Sequences have been submitted to GenBank for archiving under the following accession numbers: *ChLMI1*, KF939590; *RCO*, KF939591; and *LMI1-like3*, KF939592.

Supplementary Materials

www.sciencemag.org/content/343/6172/780/suppl/DC1
Materials and Methods
Figs. S1 to S17
Table S1 to S4
References (*26*–*55*)

11 November 2013; accepted 22 January 2014
10.1126/science.1248384

A Viral RNA Structural Element Alters Host Recognition of Nonself RNA

Jennifer L. Hyde,[1] Christina L. Gardner,[2] Taishi Kimura,[3] James P. White,[1] Gai Liu,[5] Derek W. Trobaugh,[2] Cheng Huang,[4] Marco Tonelli,[6] Slobodan Paessler,[4] Kiyoshi Takeda,[3] William B. Klimstra,[2] Gaya K. Amarasinghe,[5] Michael S. Diamond[1,5,7]*

Although interferon (IFN) signaling induces genes that limit viral infection, many pathogenic viruses overcome this host response. As an example, 2'-O methylation of the 5′ cap of viral RNA subverts mammalian antiviral responses by evading restriction of *Ifit1*, an IFN-stimulated gene that regulates protein synthesis. However, alphaviruses replicate efficiently in cells expressing *Ifit1* even though their genomic RNA has a 5′ cap lacking 2'-O methylation. We show that pathogenic alphaviruses use secondary structural motifs within the 5′ untranslated region (UTR) of their RNA to alter *Ifit1* binding and function. Mutations within the 5′-UTR affecting RNA structural elements enabled restriction by or antagonism of Ifit1 in vitro and in vivo. These results identify an evasion mechanism by which viruses use RNA structural motifs to avoid immune restriction.

Eukaryotic mRNA contains a 5′ cap structure with a methyl group at the N-7 position (cap 0). In higher eukaryotes, methylation also occurs at the 2′-O position of the penultimate and antepenultimate nucleotides to generate cap 1 and 2 structures, respectively. Many viral mRNAs also display cap 1 structures. Because cytoplasmic viruses cannot use host nuclear capping machinery, some have evolved viral methyltransferases for N-7 and 2′-O capping or mechanisms to "steal" the cap from host mRNA (*1*). Whereas N-7 methylation of mRNA is critical for efficient translation (*2*), cytoplasmic viruses encoding mutations in their viral 2′-O-methyltransferases are inhibited by IFIT proteins (*3*–*7*), which belong to a family of interferon (IFN)–stimulated genes (ISGs) induced after viral infection [reviewed in (*8*)]. Thus, 2′-O methylation of host mRNA probably evolved, in part, to distinguish self from nonself RNA (*9*, *10*).

Alphaviruses are positive-strand RNA viruses that replicate in the cytoplasm and lack 2′-O methylation on the 5′ end of their genomic RNA (*11*, *12*) and thus should be restricted by IFIT proteins. To assess the role of IFIT1 in limiting alphavirus replication, we silenced its expression in human HeLa cells and then infected them with Venezuelan equine encephalitis virus (VEEV) strain TC83, an attenuated New World alphavirus. In cells with reduced IFIT1 expression, TC83 replicated to higher levels (Fig. 1A). To determine whether this phenotype occurred in vivo, wild-type (WT) and *Ifit1*$^{-/-}$ C57BL/6 mice were infected with TC83. In contrast to WT mice, *Ifit1*$^{-/-}$ mice succumbed to TC83 infection (Fig. 1B) and sustained a higher viral burden (Fig. 1, C and D, and fig. S1), especially in the brain and spinal cord.

We next analyzed the growth of TC83 in mouse embryonic fibroblasts (MEFs). Although untreated WT and *Ifit1*$^{-/-}$ MEFs supported TC83 infection equivalently (Fig. 1E), IFN-β pretreatment preferentially inhibited replication in WT cells. However, an absence of *Ifit1* was sufficient to restore infection. A similar trend was observed with *Ifit1*$^{-/-}$ dendritic cells and cortical neurons (fig. S2, A and B). TC83 infection in *Ifit1*$^{-/-}$ MEFs remained partially inhibited by IFN-β treatment, indicating that additional ISGs restrict viral replication (*13*–*15*). The similarity of infection by TC83 in untreated WT and *Ifit1*$^{-/-}$ MEFs probably reflects the ability of alphaviruses to antagonize the induction of type I IFN and ISGs (*16*, *17*).

TC83 was generated after passage of the virulent Trinidad donkey (TRD) VEEV strain and contains two changes that attenuate virulence (*18*). One mutation occurs at nucleotide 3 (nt 3, G3A) in the 5′-UTR and increases the sensitivity of TC83 to type I IFN (*17*). We hypothesized that the 5′-UTR mutation might explain the differential sensitivity to *Ifit1* and the pathogenicity of TC83 and TRD. To begin to test this hypothesis, WT and *Ifit1*$^{-/-}$ mice were infected with TRD (Fig. 1F). WT and *Ifit1*$^{-/-}$ mice succumbed to TRD infection without differences in survival time or mortality. Thus, in contrast to TC83, TRD was relatively resistant to the antiviral effects of *Ifit1*.

To determine whether the effect of the G3A mutation was independent of the TC83 structural genes, which contain a second attenuating mutation (*19*), we assessed replication in WT and *Ifit1*$^{-/-}$ MEFs of two isogenic chimeric VEEV/Sindbis (SINV) viruses (*20*); these encode the 5′-UTR and nonstructural proteins of TRD and structural

[1]Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA. [2]Department of Microbiology and Molecular Genetics, Center for Vaccine Research, University of Pittsburgh, Pittsburgh, PA 15261, USA. [3]Laboratory of Immune Regulation, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan. [4]Department of Pathology, University of Texas Medical Branch at Galveston, Galveston, TX 77555, USA. [5]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA. [6]National Magnetic Resonance Facility at Madison, University of Wisconsin, Madison, 433 Babcock Drive, Madison, WI 53706, USA. [7]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110, USA.

*Corresponding author. E-mail: diamond@borcim.wustl.edu

proteins of SINV, and differ only at nt 3 [(G3)VEE/SINV and (A3)VEE/SINV] (fig. S3, A and B). In IFN-β–pretreated WT MEFs, (A3)VEE/SINV was not recovered from culture supernatants (Fig. 2A). However, in IFN-β–treated *Ifit1*$^{-/-}$ MEFs, (A3)VEE/SINV infection was partially restored. In contrast, (G3)VEE/SINV replicated equivalently in IFN-β–treated WT and *Ifit1*$^{-/-}$ MEFs (Fig. 2B), indicating that a G at nt 3 renders VEEV resistant to inhibition by *Ifit1*.

RNA secondary structure algorithms predicted differences in base pairing at the 5′ end of the UTR of G3 and A3 RNA [fig. S3A and (*20*, *21*)]. The imino region of a two-dimensional nuclear Overhauser effect spectroscopy nuclear magnetic resonance spectrum revealed that A3 RNA displayed less secondary structure and base pairing than G3 RNA (fig. S4, A and B) and fewer cross peaks in the corresponding $^{1}H/^{15}N$ heteronuclear single-quantum coherence (HSQC) spectrum (fig. S4, C and D). On the basis of these data, we hypothesized that the stable stem-loop structure in the 5′-UTR of TRD compensated for the absence of 2′-O methylation of alphavirus RNA. To determine whether the secondary structure or primary sequence modulated *Ifit1* susceptibility, we analyzed the growth of VEE/SINV containing the A3 nt mutation that also had compensatory mutations that were predicted to restore the 5′-UTR stem-loop (Fig. 2, C and D, and fig. S3C). Although two of the mutants tested (A3U24 and A3U24;A20U) showed increased [relative to (A3)VEE/SINV] but limited growth in IFN-β–treated WT MEFs, a third mutant (A3U24;20_21insC) replicated to levels comparable to (G3)VEE/SINV in IFN-β–treated WT and *Ifit1*$^{-/-}$ MEFs. Mutants that replicated less well in IFN-β–treated WT MEFs (A3U24 and A3U24;A20U) were predicted to have less stable minimum free energy structures relative to (A3U24;20_21insC)VEE/SINV and (G3)VEE/SINV. To further define the requirements in the 5′-UTR for evasion of *Ifit1* restriction, we evaluated additional viral mutants: one that changed the sequence of the A3U24 loop but retained the less stable stem structure of the parent A3U24 5′-UTR [(LOOP)VEE/SINV] (*21*), and two G3 variants with more stable hairpins (G3;C19C20)VEE/SINV that contained additional nucleotide repeats (AUG and AUG_2) appended to the 5′ end (fig. S5A). The latter (AUG_n)VEE/SINV mutants were relevant because RNA recognition by IFIT proteins reportedly requires a 5′ overhang of 3 to 5 nts (*22*). Alteration of the loop sequence [(LOOP)VEE/SINV] did not relieve *Ifit1*-mediated restriction (fig. S5B). However, G3 mutants with an overhang of 3 or more nts at the 5′ end became sensitive to *Ifit1*-dependent antiviral effects (fig. S5C).

To assess whether nucleotide changes altered the stability of the VEEV 5′-UTR, we monitored RNA unfolding by circular dichroism spectrometry (fig. S6). Changes in ellipticity as a function of temperature were analyzed (Fig. 2, E to I, and table S1); we observed several maxima, presumably corresponding to major cooperative unfolding events (Fig. 2, E to I). We detected more-pronounced maxima near 75°C in all but the A3 RNA, confirming that A3 and G3 RNA have different stabilities. The A3U24;20_21insC mutant RNA displayed the most stable secondary structure. Computational analyses suggested that even closely related RNA sequences (such as A3 and A3U24) have

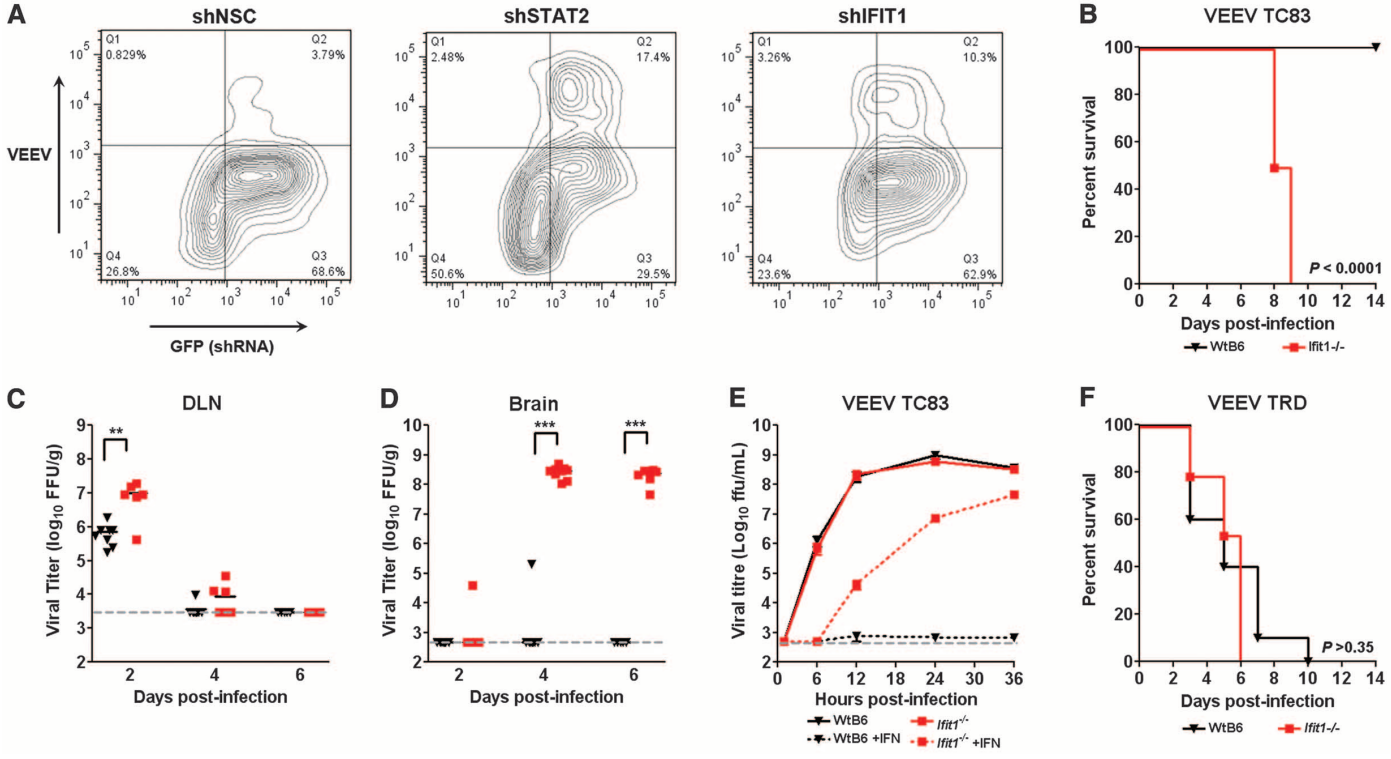


Fig. 1. VEEV TC83 but not TRD is restricted by *Ifit1*. (**A**) Flow cytometry contour plots showing infection of TC83 in IFN-β–treated HeLa cells transduced with short hairpin RNA (shRNA) against a scrambled nonsilencing control (NSC), human STAT2, or human IFIT1 (shNSC versus shIFIT1, $P < 0.003$). One representative experiment of four is shown. This phenotype was confirmed with a second shRNA against IFIT1. (**B**) Survival of 4-week-old WT mice ($n = 10$) and *Ifit1*$^{-/-}$ mice ($n = 10$) after subcutaneous (s.c.) infection with 10^6 focus-forming units (FFU) of TC83. Results are pooled from three independent experiments. *P* values for survival were calculated using the log-rank test. (**C** and **D**) Viral burden in 4-week-old WT or *Ifit1*$^{-/-}$ mice infected s.c. with 10^6 FFU of TC83, as measured in (C) a draining popliteal lymph node (DLN) and (D) the brain. Results are from five to nine mice per tissue. Asterisks indicate statistically significant differences, as judged by an unpaired *t* test (**$P < 0.005$, ***$P < 0.0001$). Dashed lines indicate the limit of detection of the assay. (**E**) WT and *Ifit1*$^{-/-}$ MEFs were pretreated with 10 IU/ml of IFN-β for 12 hours or left untreated, and then infected with TC83 [with a multiplicity of infection (MOI) of 0.1]. Supernatants were harvested for virus titration (WT versus *Ifit1*$^{-/-}$, $P > 0.2$; WT + IFN-β versus *Ifit1*$^{-/-}$ + IFN, 12, 24, and 36 hours after infection, $P < 0.03$). Each point represents the average of three experiments performed in triplicate, and error bars represent the standard error of the mean (SEM). *P* values were determined by an unpaired *t* test. (**F**) Survival curves of 8-week-old WT mice ($n = 10$) and *Ifit1*$^{-/-}$ mice ($n = 24$) after s.c. infection with 50 plaque-forming units (PFU) of TRD. Results are pooled from two independent experiments. *P* values for survival were calculated using the log-rank test.

different ensemble free energy and diversity (table S2). Differences in the base pairing probability were noted, which further support structural differences between A3 and G3 RNA (fig. S7). We also measured melting temperature (T_m) values (table S1), which showed an inverse correlation between *Ifit1* susceptibility and base-pairing stability. These analyses suggest that G3 and A3U24;20_21insC 5′-UTR RNAs adopt more stable conformations, which correlates with antagonism of *Ifit1*.

To validate that changes at nt 3 determined sensitivity to *Ifit1* independently of other VEEV-encoded factors, we repeated experiments with isogenic variants of TC83 and an enzootic VEEV strain, ZPC-738 (Fig. 3, A to D, and fig. S3D). Whereas TC83 replicated poorly in IFN-β–treated WT MEFs, the isogenic nt 3 mutant TC83 A3G showed increased replication (Fig. 3A), confirming that the A3G mutation confers resistance to type I IFN. However, unlike that seen with (G3)VEE/SINV (Fig. 2B), the phenotype of TC83 A3G in IFN-β–treated WT MEFs did not fully recapitulate the restoration seen in IFN-β–treated *Ifit1*$^{-/-}$ MEFs (compare Fig. 3A to Fig. 3B), suggesting that additional viral elements may be inhibited by *Ifit1*. Infection of the mutant ZPC-738 G3A in IFN-β–treated WT MEFs was decreased as compared to WT ZPC-738, whereas the infection of WT and G3A ZPC-738 was equivalent in IFN-β–treated *Ifit1*$^{-/-}$ MEFs (Fig. 3, C and D).

To assess whether nt 3 mutation reciprocally affects virulence, we infected WT and *Ifit1*$^{-/-}$ mice with TC83, ZPC-738, and paired isogenic variants (Fig. 3, E and F). In WT mice, ZPC-738 G3A was attenuated as compared to the WT virus. However, no difference in mortality and only a small difference in survival kinetics were observed in *Ifit1*$^{-/-}$ mice infected with ZPC-738 WT or G3A. In comparison, we observed increased lethality in WT mice infected with TC83 A3G relative to TC83. We also noted a slight decrease in the survival kinetics of *Ifit1*$^{-/-}$ mice infected with A3G as compared to TC83 WT, suggesting that the A3G change may have additional effects aside from antagonizing *Ifit1* function.

To determine whether structures in the 5′-UTR of other alphaviruses functioned analogously, we introduced mutations at either nt 5 or 8 into SINV (Fig. 3, G and H, and fig. S3E). These mutations were selected because they altered the virulence of SINV in rats (*23*, *24*) and were predicted to change the 5′-UTR secondary structure (fig. S3E). An A–to-G substitution at nt 5 resulted in increased viral replication relative to that of the parental virus in IFN-β–pretreated WT MEFs but not in IFN-β–treated *Ifit1*$^{-/-}$ MEFs, suggesting that the A5G phenotype was specific to *Ifit1*. Conversely, a substitution at nt 8 (G8U) resulted in a decrease in replication in IFN-β–treated WT MEFs relative to WT SINV, which was restored to comparable levels in IFN-β–treated *Ifit1*$^{-/-}$ MEFs. This experiment establishes that mutations within the 5′-UTR of an Old World alphavirus also affect *Ifit1* antagonism, suggesting that secondary structure at the 5′-UTR might be a more universal mechanism to circumvent *Ifit1*-mediated restriction.

IFIT1 binds flavivirus RNA lacking 2′-O methylation and blocks translation and binding of eukaryotic translation initiation factors (*6*, *7*, *25*). To determine whether *Ifit1* differentially affected translation of alphavirus RNA with different 5′-UTR RNA structures, we transfected type 0 capped WT and G3A mutant translation reporter RNA encoding a luciferase gene fused to nsP1 (fig. S3F) (*26*) into IFN-β–treated or untreated MEFs (Fig. 4, A to D). In WT MEFs treated with IFN-β (Fig. 4A), G3 RNA exhibited greater translation reporter activity relative to A3 RNA. We also detected greater translation of G3 reporter RNA in untreated WT MEFs (Fig. 4B), suggesting that basal *Ifit1* expression in these cells may limit A3 RNA translation. However, we observed a greater increase in A3 reporter RNA translation relative to G3 in *Ifit1*$^{-/-}$ MEFs that were treated with IFN-β or left untreated (Fig. 4, C and D). The higher level of A3 versus G3 RNA translation in *Ifit1*$^{-/-}$ MEFs was not unexpected, because (A3)VEE/SINV replicates more efficiently than (G3)VEE/SINV in cells lacking type I IFN induction (*20*). Although A3 RNA has a translation advantage in cells defective in innate immune responses, the G3 nucleotide confers resistance to *Ifit1*.

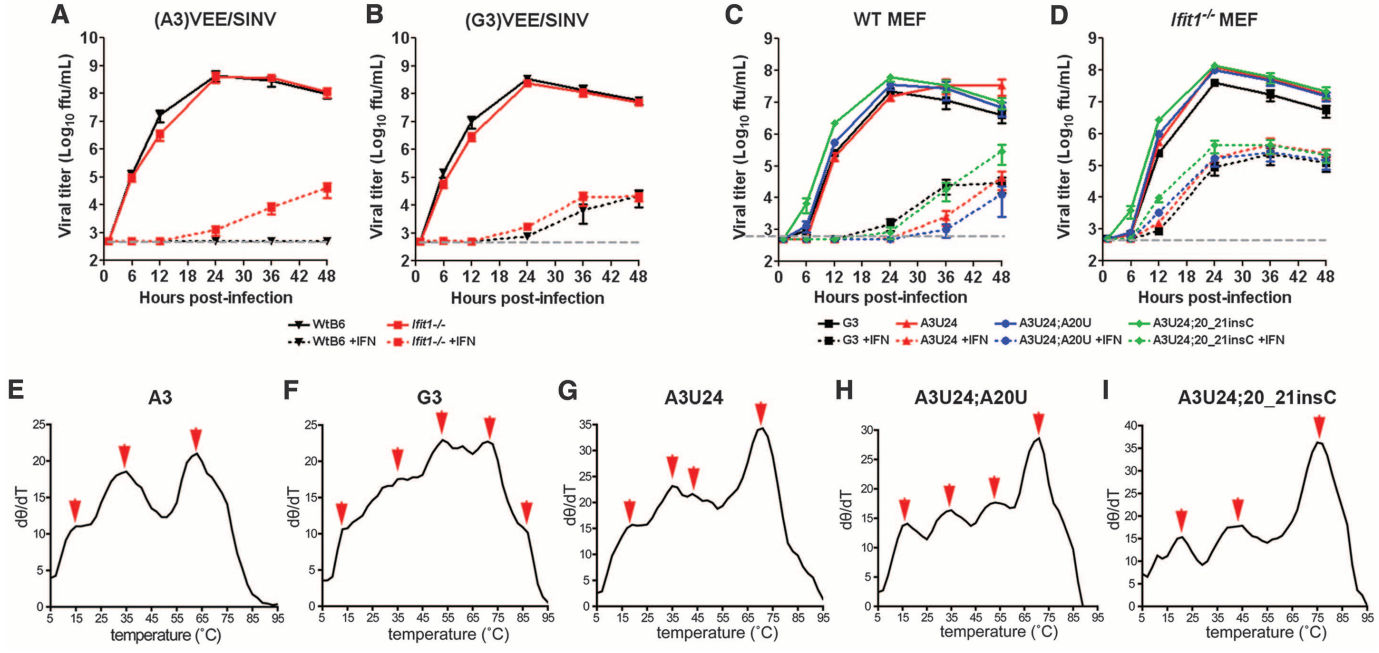


Fig. 2. Mutations in the 5′-UTR determine *Ifit1* sensitivity in vitro. (**A** and **B**) Growth kinetics of (A3)VEE/SINV and (G3)VEE/SINV viruses in WT and *Ifit1*$^{-/-}$ MEFs. Cells were pretreated with 1 IU/ml of IFN-β for 12 hours or left untreated, and then infected with (A3)VEE/SINV or (G3)VEE/SINV (MOI of 0.1). Supernatants were harvested at indicated times for virus titration [(A3)VEE/SINV: WT + IFN-β versus *Ifit1*$^{-/-}$ + IFN-β, 36 and 48 hours after infection, $P < 0.006$]. Each point represents the average of three independent experiments performed in triplicate, and error bars represent the SEM. *P* values were determined using an unpaired *t* test. Dashed lines indicate the limit of detection of the assay. (**C** and **D**) Growth kinetics of (G3)VEE/SINV, (A3U24)VEE/SINV, (A3U24;A20U)VEE/SINV, and (A3U24;20_21insC)VEE/SINV viruses in WT (C) and *Ifit1*$^{-/-}$ (D) MEFs. Experiments and analysis were performed as in (A). (**E** to **I**) Thermal denaturation of A3, G3, A3U24, A3U24;A20U, and A3U24;20_21insC RNA as measured by circular dichroism spectroscopy at 210 nm. RNA was heated from 5° to 95°C at a rate of 1°C/min, and readings were collected every 1°C to monitor unfolding. Data are represented as the change in molar ellipticity as a function of temperature ($d\theta/dT$), and red arrows indicate major maxima. One representative experiment of two is shown.

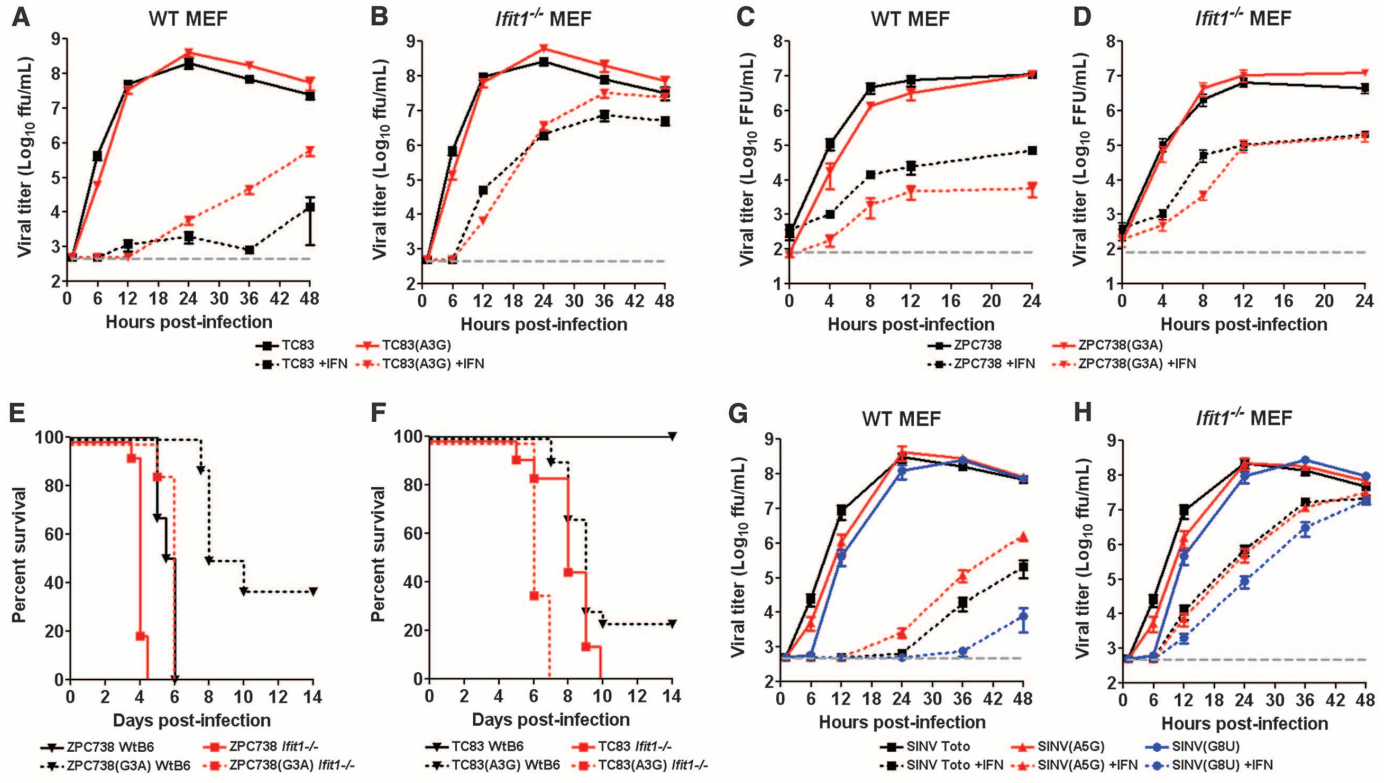


Fig. 3. Mutations that alter the secondary structure of the 5′-UTR affect pathogenicity in vivo. (**A** to **D**) Growth kinetics of isogenic TC83 WT and A3G [(A) and (B)] or ZPC-738 WT and G3A [(C) and (D)] in WT and *Ifit1*$^{-/-}$ MEFs. Cells were pretreated with 10 IU/ml of IFN-β for 12 hours (TC83) or 100 IU/ml of IFN-β for 8 hours (ZPC738), or left untreated, and then infected with the respective viruses (MOI of 0.1) [TC83 versus TC83(A3G): WT + IFN-β, 36 and 48 hours after infection, $P < 0.006$; ZPC738 versus ZPC738(G3A): WT + IFN-β, 24 hours after infection, $P < 0.0001$]. Each point represents the average of two (ZPC-738) or three (TC83) independent experiments performed in triplicate, and error bars represent the SEM. *P* values were determined using the unpaired *t* test. Dashed lines indicate the limit of detection of the assay. (**E** and **F**) Survival studies of isogenic ZPC-738 WT and G3A (E) and TC83 WT and A3G (F) viruses in WT and *Ifit1*$^{-/-}$ mice. Mice were infected s.c. with 10^1 PFU of ZPC-738 (WT, $n = 6$; *Ifit1*$^{-/-}$, $n = 15$) or ZPC-738(G3A) (WT, $n = 8$; *Ifit1*$^{-/-}$, $n = 15$) and 10^6 PFU of TC83 (WT, $n = 18$; *Ifit1*$^{-/-}$, $n = 13$) or TC83(A3G) (WT, $n = 21$; *Ifit1*$^{-/-}$, $n = 8$). ZPC738 versus ZPC738(G3A): WT mice, survival $P = 0.0002$; mean time to death (MTD) of 5.5 versus 8.3 days, $P = 0.0002$. ZPC738 versus ZPC738(G3A): *Ifit1*$^{-/-}$ mice, MTD of 4.0 versus 5.8 days, $P < 0.0001$. TC83 versus TC83(A3G): WT mice, survival $P < 0.0001$; TC83 versus TC83(A3G): *Ifit1*$^{-/-}$ mice, MTD of 8.2 versus 6.3 days, $P < 0.003$. Experiments were performed twice for ZPC-738 viruses and four times for TC83 viruses. *P* values for survival were determined as in Fig. 1. *P* values for MTD were determined using an unpaired *t* test. (**G** and **H**) Growth kinetics of SINV Toto, A5G, and G8U SINV in WT (G) and *Ifit1*$^{-/-}$ (H) MEFs. Cells were pretreated with 1 IU/ml of IFN-β for 12 hours or left untreated, and then infected with the respective viruses at an MOI of 0.1. SINV Toto versus A5G: WT MEFs + IFN-β, $P < 0.05$; SINV Toto versus G8U, WT MEFs + IFN-β, $P < 0.05$. Experiments and analysis were performed as in (A).

We hypothesized that alphavirus mutants with different 5′-UTR structural stabilities might interact with *Ifit1* in a manner that is less compatible with translation. We used electrophoretic mobility shift assays (EMSAs) (Fig. 4, E to G) to determine whether TRD 5′-UTR RNA containing an A3 or G3 and a type 0 cap differentially interacted with *Ifit1* (Fig. 4E). We observed significant binding of *Ifit1* to A3 RNA but less binding to G3 RNA, suggesting that the secondary structure of the G3 RNA probably inhibited interaction with *Ifit1*. This conclusion was supported by dot-blot binding studies, which showed a 2- to 10-fold greater affinity [dissociation constant (K_D) ~30 nM] of cap 0 A3 RNA as compared to G3 RNA for *Ifit1*, depending on the incubation conditions (Fig. 4H and fig. S8). The binding of *Ifit1* to cap 0 RNA was specific, as it was competed by excess unlabeled 5′-ppp A3 RNA (fig. S7). Exogenous 2′-O methylation of A3 and G3 RNA, which generates a type 1 cap, resulted in less *Ifit1* binding (Fig. 4F), which agrees with flavivirus studies (*6*, *7*). When EMSA experiments were repeated in the absence of capping, TRD 5′-UTR RNA containing an A3 or G3 and a free 5′-ppp differentially and weakly recognized *Ifit1* (Fig. 4G), which is consistent with experiments demonstrating that single-stranded RNA, but not double-stranded RNA containing a free 5′-ppp, is bound by IFIT1 (*22*). Excess A3 5′-ppp RNA compared to G3 5′-ppp RNA preferentially competed for *Ifit1* binding to type 0 cap A3 RNA [inhibition constant (K_i) of 3 and 48 μM for A3 and G3 5′-ppp RNA, respectively; fig. S9]. These results suggest that secondary structure in the context of an uncapped RNA can alter *Ifit1* binding and may contribute to why negative-stranded viruses with 5′-ppp genomic RNA and highly structured 5′-UTRs (such as filoviruses) are resistant to type I IFN and *Ifit1*-mediated control. Our results also establish that *Ifit1* has a higher affinity for RNA with a type 0 cap than with a free 5′-ppp moiety.

Alphaviruses use a stable 5′-UTR stem-loop structure to antagonize *Ifit1* antiviral activity. Although some IFIT proteins bind 5′-ppp RNA (*22*, *27*), it remains to be determined how *Ifit1* differentially recognizes capped RNA that displays or lacks 2′-O methylation and how alphavirus 5′-UTR stem-loop structures affect this. Our experiments suggest that genomic RNA elements can function to evade host cell–intrinsic immunity. Thus, structural elements in viral or virus-associated RNA can bind antiviral proteins irreversibly to block function (*28*, *29*) or attenuate binding of host antiviral proteins. It is intriguing to consider that viral RNA structural elements that antagonize *Ifit1* recognition may have become targets for other RNA sensors (such as RIG-I and MDA5). Finally, these results may be relevant to pharmaceutical approaches that use mRNA as therapeutics or vaccine design strategies for attenuating alphaviruses and other viruses.

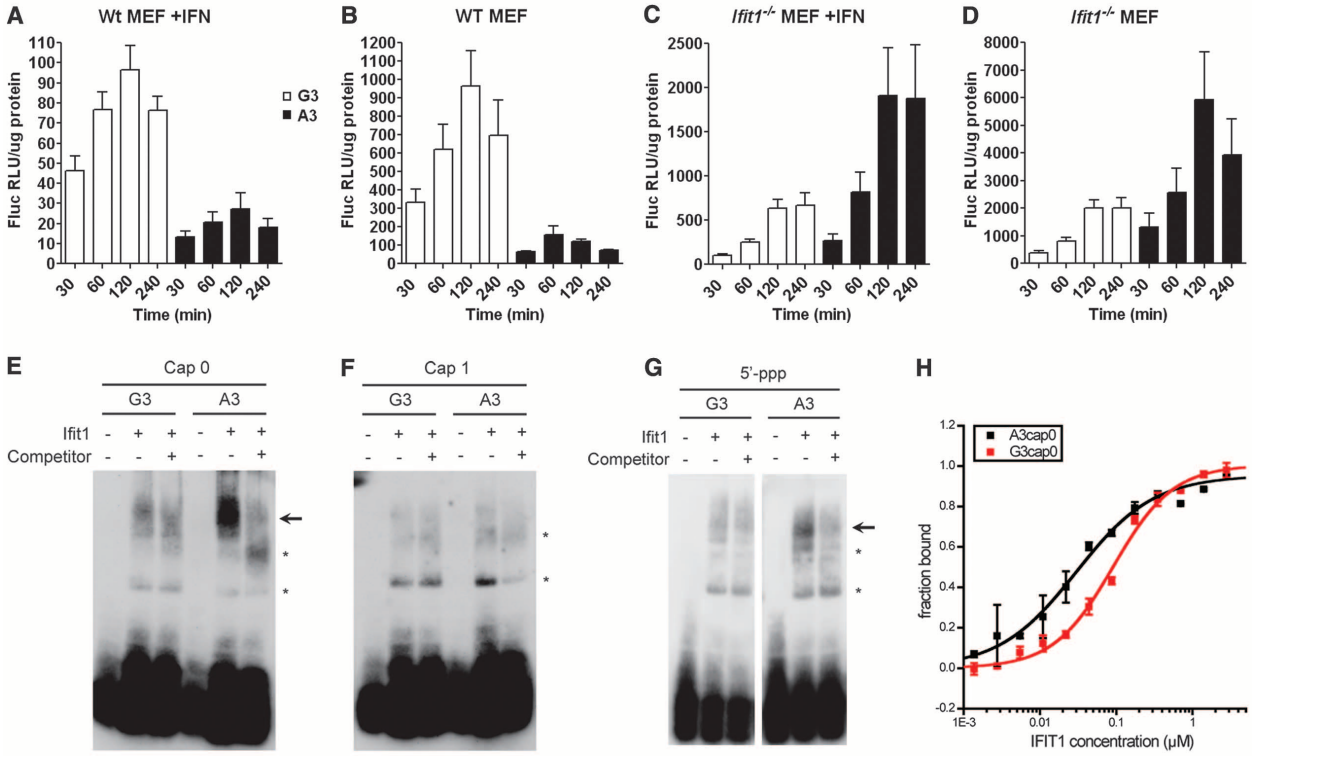


Fig. 4. The nt G3 in the 5′-UTR evades translational inhibition by altering *Ifit1*-RNA binding. (**A** to **D**) Luciferase assays of A3 and G3 TRD translation reporters. WT and *Ifit1*$^{-/-}$ MEFs were untreated or treated with 100 IU/ml IFN-β for 8 hours, and then electroporated with in vitro synthesized and type 0–capped reporter RNA. Cell lysates were harvested at the indicated time points and assayed for luciferase activity. Each bar represents the average of four independent experiments performed in triplicate. WT MEFs + IFN-β: G3 versus A3, $P < 0.0004$; WT MEFs, no treatment: G3 versus A3, $P < 0.005$; *Ifit1*$^{-/-}$ MEFs + IFN-β, G3 versus A3, $P < 0.05$ (30, 60, and 120 min). Error bars represent the SEM. *P* values were determined using an unpaired *t* test. (**E** to **G**) EMSA of A3 and G3 VEEV 5′-UTR RNA bound to recombinant *Ifit1*. G3 and A3 VEEV 5′-UTR RNA were synthesized in vitro using T7 polymerase (5′-ppp) and then treated with (E) an N-7 methylguanosine capping reagent (cap 0), (**F**) an N-7 methylguanosine capping reagent and an exogenous 2′-O methyltranferase (cap 1), or (**G**) no enzymes (5′-ppp). All RNA was labeled with biotin and competed with 3 μg of homologous unlabeled RNA. Cap 0 and cap 1 RNAs were heated at 95°C; 5′-ppp RNA was heated at 70°C, as no specific binding was observed after heating at 95°C. Binding assays were performed with 1 μg of *Ifit1*. EMSA data are representative of at least three independent experiments. Arrows indicate specific binding of RNA to *Ifit1*, whereas asterisks indicate nonspecific binding (not competed with unlabeled RNA). G3 and A3 5′-ppp paired samples were run simultaneously on the same gel and cropped as individual panels for presentation purposes. (**H**) Quantification of *Ifit1*-A3/G3 RNA binding by filter-binding assay at 4°C. The fraction bound of A3 cap 0 (black squares) and G3 cap 0 (red squares) was normalized to maximum binding and plotted against *Ifit1* concentration. Data from A3 (black) and G3 (red) were fitted using the Hill equation. A3 cap 0 $K_D = 0.030 \pm 0.004$ μM; G3 cap 0 $K_D = 0.091 \pm 0.007$ μM. One representative experiment of three performed in triplicate is shown.

References and Notes

1. S. J. Plotch, M. Bouloy, I. Ulmanen, R. M. Krug, *Cell* **23**, 847–858 (1981).
2. S. Muthukrishnan, G. W. Both, Y. Furuichi, A. J. Shatkin, *Nature* **255**, 33–37 (1975).
3. S. Daffis *et al.*, *Nature* **468**, 452–456 (2010).
4. R. Züst *et al.*, *Nat. Immunol.* **12**, 137–143 (2011).
5. K. J. Szretter *et al.*, *PLOS Pathog.* **8**, e1002698 (2012).
6. T. Kimura *et al.*, *J. Virol.* **87**, 9997–10003 (2013).
7. M. Habjan *et al.*, *PLOS Pathog.* **9**, e1003663 (2013).
8. M. S. Diamond, M. Farzan, *Nat. Rev. Immunol.* **13**, 46–57 (2013).
9. C. M. Wei, A. Gershowitz, B. Moss, *Cell* **4**, 379–386 (1975).
10. C. M. Wei, B. Moss, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 318–322 (1975).
11. E. Hefti, D. H. Bishop, D. T. Dubin, V. Stollar, *J. Virol.* **17**, 149–159 (1975).
12. R. F. Pettersson, H. Söderlund, L. Kääriäinen, *Eur. J. Biochem.* **105**, 435–443 (1980).
13. J. W. Schoggins *et al.*, *Nature* **472**, 481–485 (2011).
14. S. Karki *et al.*, *PLOS ONE* **7**, e37398 (2012).
15. Y. Zhang, C. W. Burke, K. D. Ryman, W. B. Klimstra, *J. Virol.* **81**, 11246–11255 (2007).
16. J. Yin, C. L. Gardner, C. W. Burke, K. D. Ryman, W. B. Klimstra, *J. Virol.* **83**, 10036–10047 (2009).
17. L. J. White, J. G. Wang, N. L. Davis, R. E. Johnston, *J. Virol.* **75**, 3706–3718 (2001).
18. R. M. Kinney *et al.*, *J. Virol.* **67**, 1269–1277 (1993).
19. B. J. Johnson, R. M. Kinney, C. L. Kost, D. W. Trent, *J. Gen. Virol.* **67**, 1951–1960 (1986).
20. R. Kulasegaran-Shylini, V. Thiviyanathan, D. G. Gorenstein, I. Frolov, *Virology* **387**, 211–221 (2009).
21. R. Kulasegaran-Shylini, S. Atasheva, D. G. Gorenstein, I. Frolov, *J. Virol.* **83**, 8327–8339 (2009).
22. Y. M. Abbas, A. Pichlmair, M. W. Górna, G. Superti-Furga, B. Nagar, *Nature* **494**, 60–64 (2013).
23. D. Kobiler *et al.*, *J. Virol.* **73**, 10440–10446 (1999).
24. K. L. McKnight *et al.*, *J. Virol.* **70**, 1981–1989 (1996).
25. P. Kumar *et al.*, *Nucleic Acids Res.* (2013).
26. M. Z. Tesfay *et al.*, *J. Virol.* **82**, 2620–2630 (2008).
27. A. Pichlmair *et al.*, *Nat. Immunol.* **12**, 624–630 (2011).
28. R. P. O'Malley, T. M. Mariano, J. Siekierka, M. B. Mathews, *Cell* **44**, 391–400 (1986).
29. J. Kitajewski *et al.*, *Cell* **45**, 195–200 (1986).

Acknowledgments: NIH grants U19 AI083019 (M.S.D.), R01 AI104972 (M.S.D.), R01 AI083383 (W.B.K.), and Training Grant AI049820 (D.W.T.) and the Institute for Human Infections and Immunity, University of Texas Medical Branch (UL1TR000071) supported this work. The authors thank I. Frolov, M. MacDonald, and C. Rice for their generosity with alphavirus reagents and experimental advice. This study made use of the National Magnetic Resonance Facility at Madison, Wisconsin, which is supported by NIH grants P41RR02301 (Biomedical Research Technology Program/National Center for Research Resources) and P41GM66326 (National Institute of General Medical Sciences). The luciferase reporter gene constructs and virulent VEEV strains are subject to material transfer agreements with the University of Pittsburgh and the University of Texas Medical Branch, respectively. Use of the virulent VEEV strain (TRD) requires Biosafety Level 3 facilities and U.S. Department of Agriculture approval. The authors have no financial conflicts to disclose. The data reported in this manuscript are tabulated in the main paper and in the supplementary materials.

Supplementary Materials

www.sciencemag.org/content/343/6172/783/suppl/DC1
Materials and Methods
Author Contributions
Figs. S1 to S9
Tables S1 to S4
References (*30*–*46*)

13 November 2013; accepted 17 January 2014
10.1126/science.1248465

A Common Cellular Basis for Muscle Regeneration in Arthropods and Vertebrates

Nikolaos Konstantinides[1,2,3]* and Michalis Averof[1,3,4]*

Many animals are able to regenerate amputated or damaged body parts, but it is unclear whether different taxa rely on similar strategies. Planarians and vertebrates use different strategies, based on pluripotent versus committed progenitor cells, respectively, to replace missing tissues. In most animals, however, we lack the experimental tools needed to determine the origin of regenerated tissues. Here, we present a genetically tractable model for limb regeneration, the crustacean *Parhyale hawaiensis*. We demonstrate that regeneration in *Parhyale* involves lineage-committed progenitors, as in vertebrates. We discover *Pax3/7*-expressing muscle satellite cells, previously identified only in chordates, and show that these cells are a source of regenerating muscle in *Parhyale*. These similarities point to a common cellular basis of regeneration, dating back to the common ancestors of bilaterians.

Regeneration relies on specific populations of progenitor cells, which serve as the source of new cells in the regenerated tissues. Progenitors may be undifferentiated stem cells or differentiated cells that have the capacity to dedifferentiate, proliferate, and redifferentiate to produce new functionally specialized cells (*1*). Their identity and degree of commitment are relevant for addressing fundamental questions in regenerative biology, such as the role of cellular memory and plasticity during regeneration.

Although the capacity to regenerate is widespread in animals, the evolutionary origins of regeneration remain unexplored. Among animals with extensive regenerative abilities, progenitor cells have been identified only in planarians, vertebrates, and cnidarians. These animals use different strategies to replace missing tissues. In planarians, all tissues regenerate from a common pool of pluripotent progenitor cells (*2*, *3*), whereas in vertebrates, different cell types arise from distinct progenitors (*4*–*7*). Cnidarians use progenitor cells that are specialized to different degrees in different species (*8*–*10*). In most animal phyla, we lack the tools to rigorously identify these progenitor cells and to map their lineage commitments. Thus, we do not know if similar types of progenitors exist across diverse phyla, whether there are shared regenerative strategies, and which of these strategies is most ancient.

Here, we establish the crustacean *Parhyale hawaiensis* as a genetically tractable model for limb regeneration. *Parhyale* can fully regenerate all amputated appendages throughout their lifetime (Fig. 1A). Regeneration restores all of the cell types that can be observed in adult limbs, including epidermis, neurons, and muscles. Several of these cell types can be visualized with the use of gene trap lines and reporter constructs (fig. S1). The speed of regeneration varies among individuals and correlates with the frequency of moulting (fig. S2). Young adults typically need 5 to 8 days to regenerate a patterned thoracic limb, such as the one shown in Fig. 1B.

We used cellular, morphological, and genetic markers to define a timeline for limb regeneration in *Parhyale* young adults (Fig. 1C). Wound closure takes place within a day of amputation, as seen by the development of a melanized scab at the wound surface (Fig. 1, D to D″). This is followed by the formation of a blastema, consisting of proliferating cells underlying the wound, which can be visualized by EdU incorporation 2 to 3 days after amputation (Fig. 1, E to E″, and fig. S3). Approximately 4 to 6 days after amputation, the distal tip of the newly regenerated appendage becomes apparent, visualized with the *Distal*^{DsRed} gene trap (*11*) (Fig. 1, F to F″, and movie S1). During the following days, the regenerating limb grows in size, acquiring its characteristic pattern of limb segments. The axis of the limb is folded (often S-shaped) to accommodate the growing appendage in the limited space available within the exoskeleton of the amputated limb (Fig. 1B). The fully regenerated limb is revealed during the next moult. Muscles, visualized with the *PhMS-DsRed* reporter (*12*), regenerate after the epidermis, within a week from moulting (fig. S1B′).

To address whether pluripotent or lineally restricted progenitor cells give rise to the new tissues of regenerated limbs, we marked individual cell lineages in early embryos and followed their regenerative contributions in adults. *Parhyale* embryos have a stereotypic early cell lineage: At the eight-cell stage, three blastomeres (El, Er, and Ep) are fated to produce the ectoderm, three (ml, mr, and Mav) give rise to mesoderm, one (en) produces cells that localize in the gut, and one (g) gives rise to the germ line (*13*) (Fig. 2A). We stably marked each of these blastomere lineages by injecting early embryos with a *Minos* transposon carrying a fluorescence marker [enhanced green fluorescent protein (EGFP) or DsRed] driven by the *PhHS* promoter, which is activated in all cell types after heat shock (*14*). Integration of the transposon in the genome of individual blastomeres produced mosaic animals in which the marked lineages could be identified (Fig. 2, A to C). By injecting ~4000 early embryos and screening the survivors as late embryos and juveniles, we identified 79 individuals in which specific cell lineages were marked. These individuals were raised to adulthood, subjected to limb amputations, and allowed to regenerate. The contribution of each marked lineage was then assessed on the regenerated limbs (Fig. 2, B and C, and supplementary materials and methods). Results are summarized in Fig. 2D.

We consistently observed that the descendants of blastomeres El, Er, and Ep gave rise to regenerated ectodermal cell types, namely epidermis and neurons, but never to mesodermal cells, such as muscle (Fig. 2B). Conversely, the descendants of blastomeres ml, mr, and Mav gave rise to regenerated muscle, but not to epidermis or neurons (Fig. 2C). We did not observe any contribution from the en and g lineages to regenerated appendages. We found no blastomere lineages contributing to both ectodermal and mesodermal cell types, implying an absence of pluripotent progenitors and of transdifferentiation across ectoderm and mesoderm. These results demonstrate that, in *Parhyale*, regenerative progenitor cells have a developmental potential that is restricted with respect to germ layers. These progenitors may be stem cells or differentiated cells that have retained the capacity to proliferate.

We also found that marked cell lineages only contribute to regenerating limbs locally, within the body region that was originally populated by each lineage: El and ml contribute to regeneration of appendages on the left side of the body, Er and mr to appendages in the right, Ep to posterior thoracic appendages, and Mav to regenerated antennae (Fig. 2D). The fact that no lineage contributed to all limbs suggests that there is no central pool of progenitor cells for the whole body—the progenitor cells reside locally.

Our experiments show that *Parhyale* limb regeneration involves at least two distinct types of progenitor cells: ectodermal and mesodermal. These progenitors derive from distinct cell lineages, have predetermined (lineally restricted) regenerative capacities, and are present near the regenerating tissue. This is highly reminiscent of vertebrates—in which distinct progenitors give rise to different ectodermal and

[1]Institute of Molecular Biology and Biotechnology (IMBB), Foundation for Research and Technology Hellas (FORTH), 70013 Heraklio, Crete, Greece. [2]Graduate Programme in Molecular Biology and Biomedicine and Medical School, University of Crete, Greece. [3]Institut de Génomique Fonctionnelle de Lyon (IGFL), École Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon, France. [4]Centre National de la Recherche Scientifique (CNRS), France.

*Corresponding author. E-mail: michalis.averof@ens-lyon.fr (M.A.); nikolaos.konstantinides@ens-lyon.fr (N.K.)

mesodermal cells during limb, fin, and tail regeneration (*4*–*6*)—and differs from the strategy used by planarians involving pluripotent stem cells (*2*, *3*).

While studying the *PhMS-DsRed* reporter line, where DsRed is expressed in muscles (*12*), we noticed another DsRed-expressing cell type that is tightly associated with muscles within each limb (Fig. 3A). These compact cells reminded us of the satellite cells, which can serve as muscle progenitors during muscle maintenance, growth, and regeneration in vertebrates (*4*, *15*–*19*). Satellite cells are characterized by expression of Pax3/7 family transcription factors (*19*, *20*). Using two antibodies that recognize Pax3/7 proteins in a wide range of animals (*21*) and in situ hybridization, we saw that the muscle-associated cells of *Parhyale* also express Pax3/7 (Fig. 3, A to C, and fig. S6). The expression level is variable but is consistently above background in these cells (fig. S7). By genetically marking mesodermal cell lineages (as described earlier), we could demonstrate that these Pax3/7-positive cells have a mesodermal origin (Fig. 3D). Based on these characteristics, we refer to these mesodermal Pax3/7-positive cells as satellite-like cells (SLCs).

Consistent with a possible involvement of SLCs in regeneration, Pax3/7-positive cells proliferate in the amputated limb stump and contribute to the regeneration blastema 2 to 3 days after amputation (Fig. 3E). To directly test whether SLCs can act as muscle precursors during regeneration, we transplanted isolated SLCs from the limbs of *PhMS-EGFP* transgenic animals into the amputated limbs of wild-type animals and examined the contribution of those EGFP-marked cells after regeneration (Fig. 4A and supplementary materials). The EGFP-marked cells expressed Pax3/7 (fig. S7) and were free of muscle cells. Out of 72 limbs that received marked SLCs, 12 contained one or a few EGFP-expressing muscle fibers after regeneration (Fig. 4, B to D). To test whether muscle fibers could derive from unlabeled cells that were inadvertently transplanted together with the EGFP-marked SLCs, we also transplanted unlabeled (non–EGFP-expressing) cells from *PhMS-EGFP* donors to wild-type recipients; no EGFP-expressing muscles were observed after regeneration in 34 limbs. These results demonstrate that SLCs are capable of functioning as progenitor cells for regenerated muscle. Our experiments do not exclude the involvement of other types of muscle progenitors, in addition to SLCs.

Our study has revealed a number of key similarities between arthropod and vertebrate

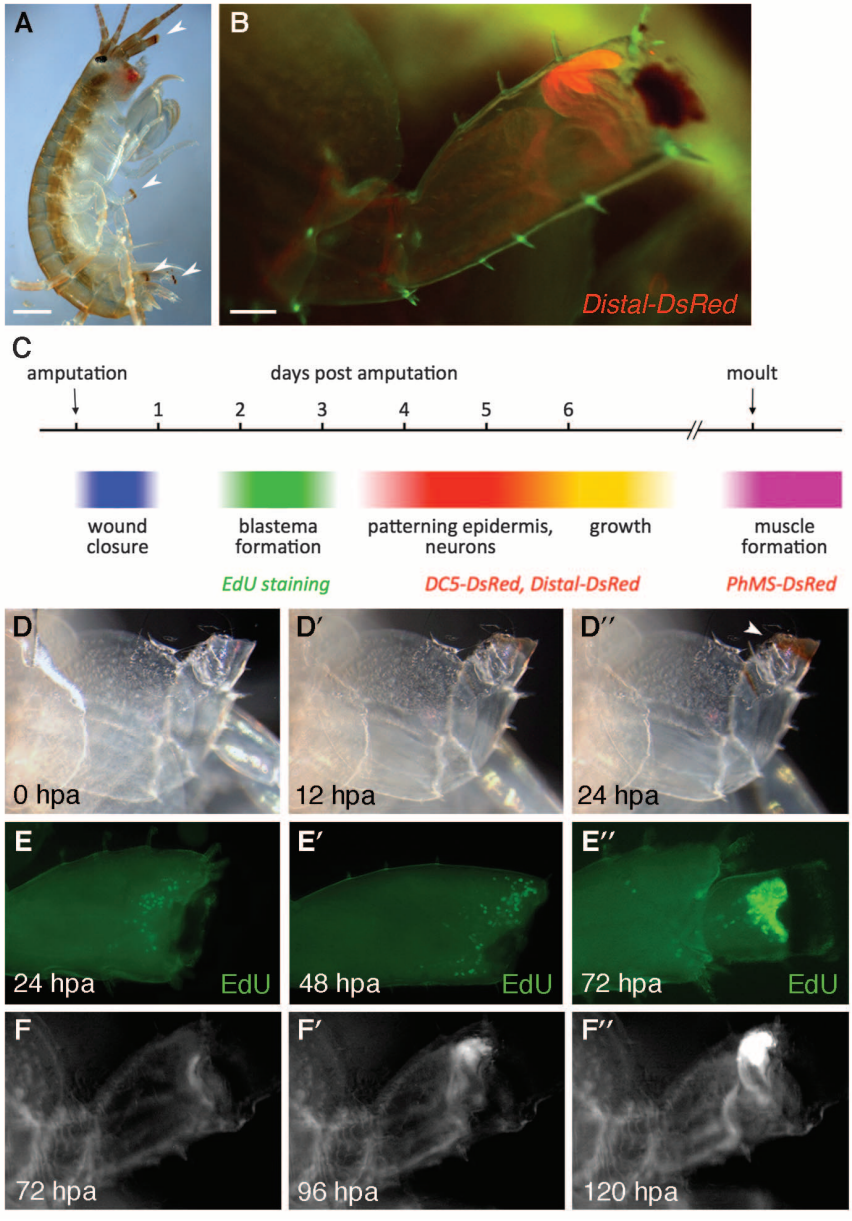
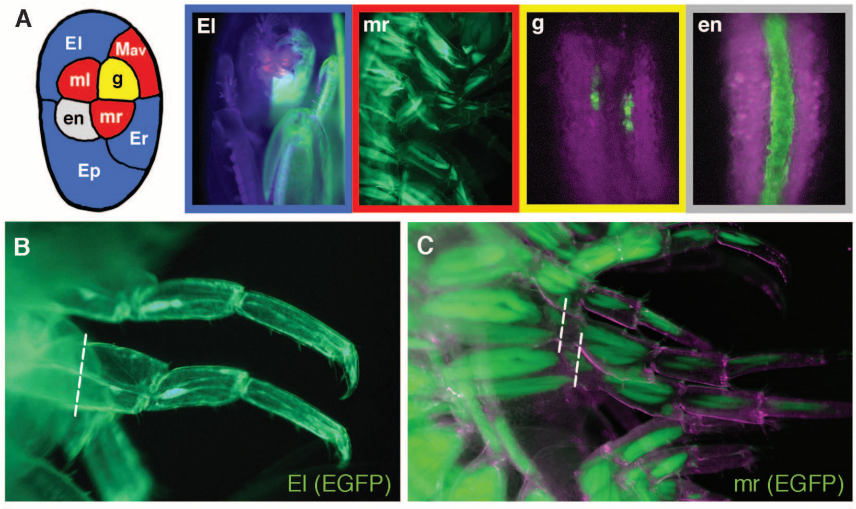


Fig. 1. Limb regeneration in *Parhyale*. (**A**) Adult *Parhyale* with amputated antenna, thoracic leg, and uropods (arrowheads). Scale bar, 1 mm. (**B**) Regenerating thoracic leg visualized by the *Distal*^{DsRed} gene trap (*11*) (red), within the cuticle of the amputated limb (green autofluorescence). Scale bar, 100 μm. (**C**) Timeline of *Parhyale* leg regeneration, representing the average speed of regeneration in ~6-month-old adults. (**D** to **F″**) Different phases of leg regeneration are characterized by melanized scab formation at the wound surface (arrowhead) (D, D′, D″); proliferating cells in the blastema, labeled by EdU (green) (E, E′, E″); and morphogenesis of the leg visualized by the *Distal*^{DsRed} gene trap (F, F′, F″) (time-lapse shown in movie S1). hpa, hours postamputation.



D

Marked lineages	Regenerated tissues	Ectoderm Left	Ectoderm Right	Ectoderm Posterior	Mesoderm Left	Mesoderm Right	Mesoderm Anterior
	El (n=19)	26/27	0/27	0/7	0/20	0/19	0/23
	Er (n=18)	0/25	21/21	0/10	0/19	0/18	0/21
	Ep (n=11)	0/18	0/21	10/10	0/14	0/20	0/21
	ml (n=12)	0/6	0/9	0/8	10/11	0/11	0/10
	mr (n=10)	0/13	0/7	0/9	0/14	11/11	0/8
	Mav (n=6)	0/7	0/6	0/8	0/11	0/10	12/12
	en (n=13)	0/21	0/38	0/22	0/25	0/25	0/34
	g (n=8)	0/17	0/21	0/16	0/18	0/14	0/21

Fig. 2. Contribution of marked cell lineages to regenerated tissues. (**A**) At the eight-cell stage, the *Parhyale* embryo has blastomeres contributing specifically to the ectoderm (blue), mesoderm (red), gut (gray), and germ line (yellow) (*13*). Images show EGFP-marked lineages in the ectoderm (El, ventral view), mesoderm (mr, lateral view), germ line (g, dorsal view), and gut (en, dorsal view). (**B** and **C**) EGFP-marked mosaics showing the contributions of the El and mr lineages during regeneration. After amputation, the El lineage contributed to epidermis and neurons (B), whereas the mr lineage contributed to muscles (C). Amputation planes are marked by dashed lines; limbs are visualized by reflected light in (C) (magenta). (**D**) Summary of lineage contributions to regenerated limb tissues (number of marked limbs/number of limbs tested). n, number of animals in which the particular blastomere lineage was labeled.

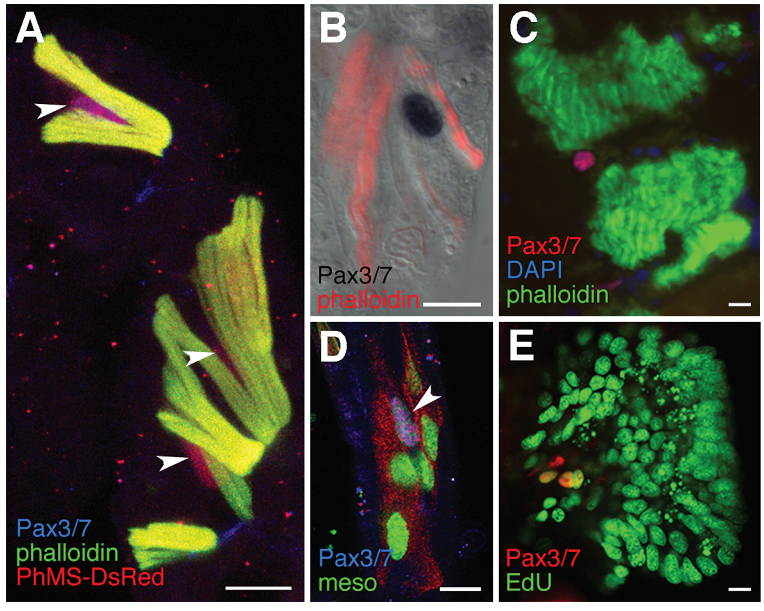


Fig. 3. Satellite-like cells (SLCs) in *Parhyale*. (**A**) SLCs (arrowheads) in the ischium and merus of a thoracic limb, from a late embryo carrying the *PhMS-DsRed* reporter, stained with antibodies for DsRed (red) and Pax3/7 (blue) and with phalloidin to mark muscles (green). The *PhMS* regulatory sequence carries putative MyoD binding sites (*12*). (**B**) SLC in *Parhyale* limb stained with an antibody for Pax3/7 (black) and phalloidin (red). SLC nuclei have a diameter of 5 to 10 μm and occupy most of the cell's volume. (**C**) SLC in an adult *Parhyale* limb sectioned transversely and stained with an antibody for Pax3/7 (red), 4′,6-diamidino-2-phenylindole (DAPI) (blue), and phalloidin (green). SLC nuclei are associated with muscle fibers in late embryos, juveniles, and adults. (**D**) SLCs have a mesodermal origin, seen by the colocalization of Pax3/7-positive nuclei (blue) with a marker for mesoderm (arrowhead) (seven out of seven SLCs scored in late embryos). Mesodermal cells were clonally marked with nuclear-localized EGFP (green) and membrane-localized tdTomato (red). (**E**) Pax3/7-expressing cells (red) contribute to the regeneration blastema (EdU, green). The amputation plane was at the edge of the blastema, on the right. Four Pax3/7-positive cells are seen in the proximal part of the blastema, three of which are positive for EdU (on average, 2.5 Pax3/7-positive cells per blastema, scored on 12 blastemas). Individual fluorescence channels for these images are shown in fig. S4. Scale bars, 10 μm.

limb regeneration. First, regenerated ectodermal and mesodermal cells derive from distinct progenitor cells, rather than a common pool of pluripotent progenitors. Second, the progenitors are present locally in the amputated limb stump. Third, muscle regeneration involves a similar progenitor cell type, the satellite-like cells, which are tightly associated with muscles before regeneration and contribute to newly formed muscle fibers. We suggest that these similarities reflect cell types and repair strategies that evolved in the common ancestors of protostomes and deuterostomes in Precambrian times.

References and Notes

1. E. M. Tanaka, P. W. Reddien, *Dev. Cell* **21**, 172–185 (2011).
2. P. W. Reddien, A. Sánchez Alvarado, *Annu. Rev. Cell Dev. Biol.* **20**, 725–757 (2004).
3. D. E. Wagner, I. E. Wang, P. W. Reddien, *Science* **332**, 811–816 (2011).
4. C. Gargioli, J. M. Slack, *Development* **131**, 2669–2679 (2004).
5. M. Kragl *et al.*, *Nature* **460**, 60–65 (2009).
6. S. Tu, S. L. Johnson, *Dev. Cell* **20**, 725–732 (2011).
7. Y. Rinkevich, P. Lindau, H. Ueno, M. T. Longaker, I. L. Weissman, *Nature* **476**, 409–413 (2011).
8. H. R. Bode, *J. Cell Sci.* **109**, 1155–1164 (1996).
9. W. A. Müller, R. Teo, U. Frank, *Dev. Biol.* **275**, 215–224 (2004).
10. T. C. G. Bosch, *Bioessays* **31**, 478–486 (2009).
11. Z. Kontarakis *et al.*, *Development* **138**, 2625–2630 (2011).

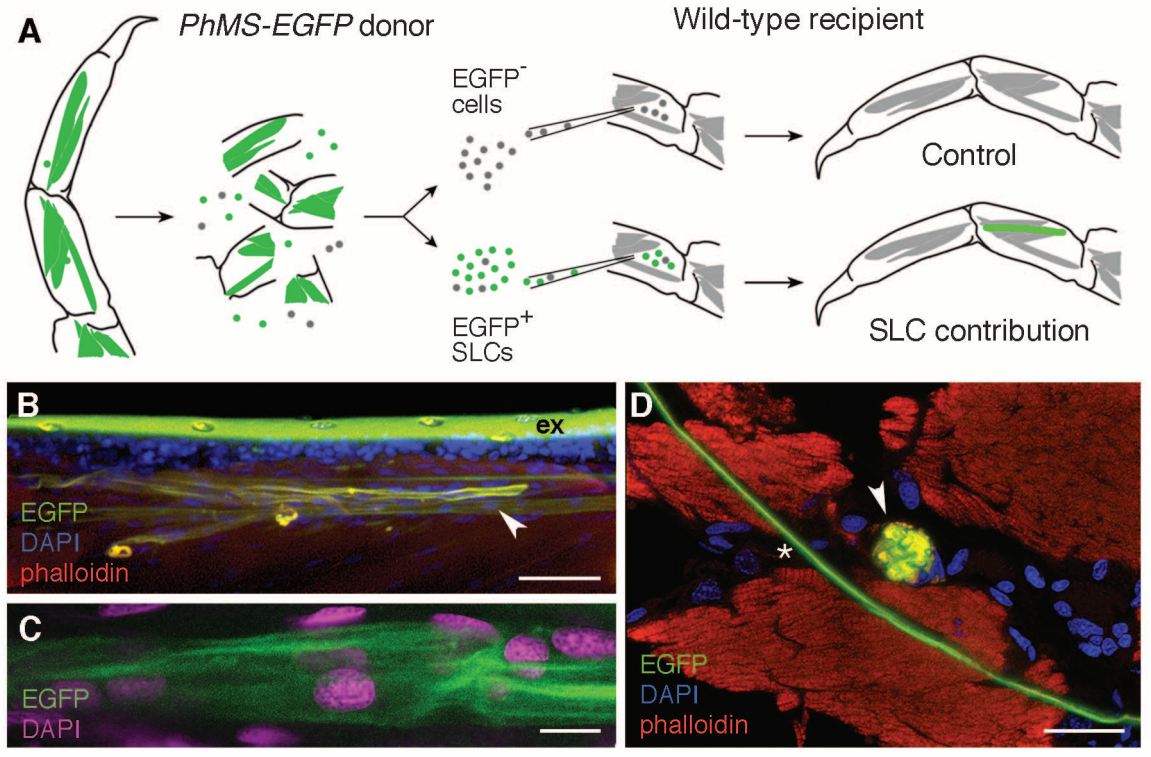


Fig. 4. Transplanted SLCs contribute to muscle regeneration. (**A**) Schematic representation of the transplantation experiment designed to test the contribution of SLCs to regenerated muscle. EGFP-marked SLCs were taken from the limbs of transgenic donors carrying the *PhMS-EGFP* construct (*12*) and transplanted into freshly amputated limbs of nontransgenic recipients. In control experiments, non–EGFP-expressing cells were transplanted using the same donor and recipient lines. (**B**) SLC-derived muscle fibers were detected by EGFP fluorescence (arrowhead, green) in recipient limbs also stained with phalloidin (red) and DAPI (blue). ex, autofluorescence of the chitinous exoskeleton. Scale bar, 50 μm. (**C**) Higher-magnification view of EGFP-expressing muscle. Scale bar, 10 μm. (**D**) Transverse section of a regenerated recipient limb, showing an EGFP-expressing muscle fiber (arrowhead), stained with an antibody for EGFP (green) and phalloidin to mark muscles (red). The asterisk marks autofluorescence from a slice of cuticle that was displaced during sectioning. Scale bar, 20 μm.

12. A. Pavlopoulos, M. Averof, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7888–7893 (2005).
13. M. Gerberding, W. E. Browne, N. H. Patel, *Development* **129**, 5789–5801 (2002).
14. A. Pavlopoulos *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13897–13902 (2009).
15. J. I. Morrison, P. Borg, A. Simon, *FASEB J.* **24**, 750–756 (2010).
16. A. Sacco, R. Doyonnas, P. Kraft, S. Vitorovic, H. M. Blau, *Nature* **456**, 502–506 (2008).
17. C. Lepper, T. A. Partridge, C. M. Fan, *Development* **138**, 3639–3646 (2011).
18. R. Sambasivan *et al.*, *Development* **138**, 3647–3656 (2011).
19. Y. X. Wang, M. A. Rudnicki, *Nat. Rev. Mol. Cell Biol.* **13**, 127–133 (2012).
20. I. M. L. Somorjai, R. L. Somorjai, J. Garcia-Fernàndez, H. Escrivà, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 517–522 (2012).
21. G. K. Davis, J. A. D'Alessio, N. H. Patel, *Dev. Biol.* **285**, 169–184 (2005).

Acknowledgments: We thank A. Pavlopoulos, Z. Kontarakis, E. Kabrani, N. Patel, and M. Grillo for sharing reagents and unpublished information and Z. Kontarakis, K. Echeverri, E. Houliston, M. Telford, V. Laudet, F. Alwes, and two anonymous referees for suggesting improvements. Our research was supported by the Heraklitos II program of the Ministry of Education (Greece) and the European Social Fund, and by grant ANR-12-CHEX-0001-01 of the Agence Nationale de la Recherche (France).

Supplementary Materials

www.sciencemag.org/content/343/6172/788/suppl/DC1
Materials and Methods
Figs. S1 to S7
References (*22*–*24*)
Movie S1

22 July 2013; accepted 20 December 2013
Published online 2 January 2014;
10.1126/science.1243529

Somites Without a Clock

Ana S. Dias,[1]* Irene de Almeida,[1]* Julio M. Belmonte,[2] James A. Glazier,[2] Claudio D. Stern[1]†

The formation of body segments (somites) in vertebrate embryos is accompanied by molecular oscillations (segmentation clock). Interaction of this oscillator with a wave traveling along the body axis (the clock-and-wavefront model) is generally believed to control somite number, size, and axial identity. Here we show that a clock-and-wavefront mechanism is unnecessary for somite formation. Non-somite mesoderm treated with Noggin generates many somites that form simultaneously, without cyclic expression of Notch-pathway genes, yet have normal size, shape, and fate. These somites have axial identity: The Hox code is fixed independently of somite fate. However, these somites are not subdivided into rostral and caudal halves, which is necessary for neural segmentation. We propose that somites are self-organizing structures whose size and shape is controlled by local cell-cell interactions.

The mesoderm of the embryo, from which the cardiovascular and musculoskeletal systems arise, derives from the primitive streak (PS) during gastrulation. A high level of bone morphogenetic protein (BMP) at the posterior PS generates ventral mesoderm (blood vessels, lateral and extraembryonic mesoderm), whereas lower levels near the anterior tip generate paraxial mesoderm, from which somites (future striated muscle and axial skeleton) develop (*1*). Somites are epithelial spheres that form sequentially from head to tail on either side of the spinal cord. The combination of a molecular clock (cell-autonomous Notch and Wnt oscillations) and a wave traveling the length of the paraxial mesoderm (*2*, *3*) is thought to regulate the number, size, timing of formation, and

[1]Department of Cell and Developmental Biology, University College London, Gower Street, London WC1E 6BT, UK. [2]Department of Physics, Biocomplexity Institute, University of Indiana at Bloomington, Simon Hall MSB1, 047G, 212 South Hawthorne Drive, Bloomington, IN 47405–7003, USA.

*These authors contributed equally to this work.
†Corresponding author. E-mail: c.stern@ucl.ac.uk

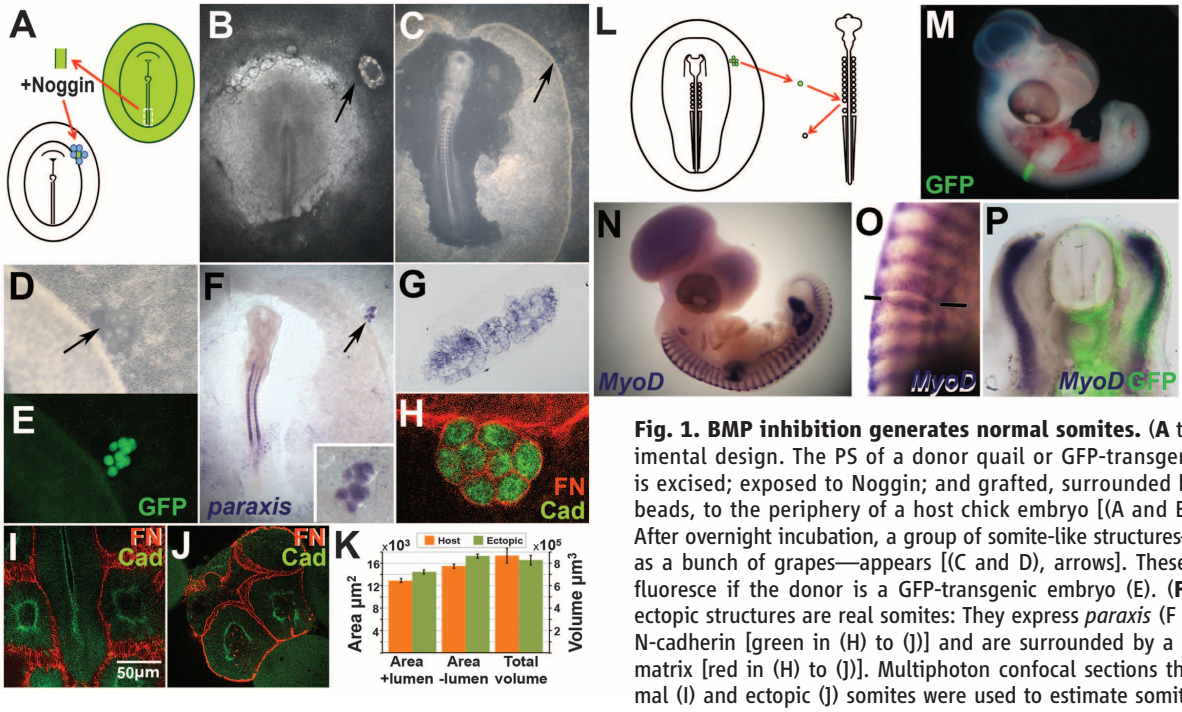


Fig. 1. BMP inhibition generates normal somites. (**A** to **E**) Experimental design. The PS of a donor quail or GFP-transgenic embryo is excised; exposed to Noggin; and grafted, surrounded by Noggin-beads, to the periphery of a host chick embryo [(A and B), arrows]. After overnight incubation, a group of somite-like structures—arranged as a bunch of grapes—appears [(C and D), arrows]. These structures fluoresce if the donor is a GFP-transgenic embryo (E). (**F** to **P**) The ectopic structures are real somites: They express *paraxis* (F and G) and N-cadherin [green in (H) to (J)] and are surrounded by a Fibronectin matrix [red in (H) to (J)]. Multiphoton confocal sections through normal (I) and ectopic (J) somites were used to estimate somite sizes (K). When an ectopic somite is grafted instead of a somite in an older embryo (L), the graft incorporates well (M). After 2 to 3 days, the grafted somite appropriately expresses *MyoD* (N to P).

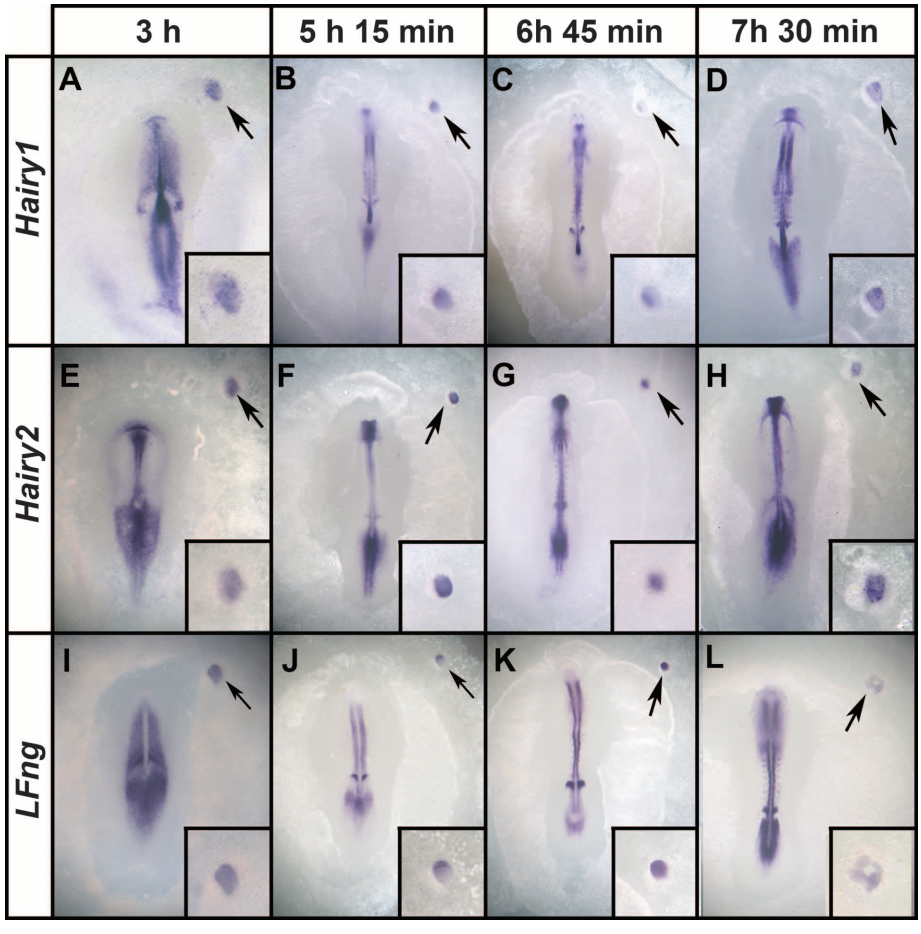


Fig. 2. Ectopic somites form without cyclic expression of segmentation clock genes. Embryos were fixed at 45-min intervals (examples shown at 3, 5.15, 6.45, and 7.5 hours after grafting to a host embryo) and stained for expression of *Hairy1* (**A** to **D**), *Hairy2* (**E** to **H**), and *LFng* (**I** to **L**). The in situ embryos were developed to reveal the segmentation clock in the presomitic cells of the host. Although patterns of expression in the presomitic mesoderm of the host are dynamic, no major differences in expression are seen in the graft (insets). Arrows mark the graft region, which is shown magnified in the insets.

axial identity (*4*–*6*) of somites. Because the BMP antagonist Noggin is sufficient to transform ventral cells to a dorsal (somite) fate (*7*, *8*), we applied Noggin as evenly as possible (*9*) to dorsalize posterior PS explants from quail or green fluorescent protein (GFP)–transgenic chick embryos and thus to test whether somites could be generated independently of a segmentation clock (*10*, *11*). Explants from stage-5 (*12*) embryos were incubated in Noggin for 3 hours, then grafted into a remote (extraembryonic) region of a host chick embryo surrounded by Noggin-soaked beads (Fig. 1, A and B). A few hours later (total 9 to 12 hours), 6 to 14 somite-like structures had formed, arranged as a "bunch of grapes" (Fig. 1, C to E) rather than in linear sequence. Like normal somites, these structures express *paraxis* (*8*) (Fig. 1, F and G) and consist of epithelial cells around a lumen (Fig. 1, G to J), with apical N-cadherin and a Fibronectin-positive basal lamina (Fig. 1, H to J). The size of each somite-like structure is normal: Fig. 1K compares ectopic and normal somite volumes calculated from living embryos and multiphoton cross-sectional areas with and without the lumen (*t* tests $P = 0.496$, 0.401, and 0.493, respectively).

To test whether the ectopic somites can give rise to normal somite derivatives, we replaced individual recently formed somites in 10 to 14 somite secondary hosts with ectopic GFP-transgenic somites (Fig. 1L). After 2 to 3 days (stages 19 to 25), the grafted somite was well integrated (Fig. 1M) and expressed the sclerotome/vertebral marker *Pax1* (fig. S1) ($n = 6$ experiments) and the dermomyotome/muscle marker *MyoD* (Fig. 1, N to P) ($n = 4$) in the correct positions. Some blood vessels were also generated (fig. S1), which may be normal somite derivatives (*13*, *14*) or cells retaining their original lateral fate. Thus, the structures in the "bunch of grapes" are indeed somites.

To test whether somites form sequentially or simultaneously, we used time-lapse microscopy to film ectopic GFP-transgenic somite formation. About 6 to 14 somites form in just 2 hours (9 to 11 hours after grafting) (fig. S2 and movies S1 and S2). The finding that so many somites can form almost synchronously suggests that the ectopic somites form independently of a clock. To assess the molecular clock, we examined embryos at different time points before ectopic somite formation for expression of clock genes *Hairy1* (Fig. 2, A to D), *Hairy2* (Fig. 2, E to H), and *LFng* (Fig. 2, I to L) at 45-min intervals between 3 and 7.5 hours after exposure of PS explants to Noggin. Although host embryos displayed typical (*10*) strong variations in the pattern of expression, the explants showed only subtle variations, not like a prepattern of the somites that would later form. Moreover, when examining many embryos for each marker at a particular time point (fig. S3), oscillatory expression was evident in the host embryo, but the explants (insets) show comparatively uniform expression. Examination of *Dapper1* and *-2* expression suggests that Noggin-treated mesoderm can generate somites without passing through a presomitic-like state (fig. S4). These results strongly suggest that the ectopic somites form simultaneously and without cyclic expression of clock genes.

Each somite is normally subdivided into two halves, rostral and caudal, a property subsequently required for segmentation of the peripheral nervous system (*15*). To test whether the ectopic somites are subdivided, we examined expression of caudal (*Hairy1*, *Hairy2*, *LFng*, *Uncx4.1*, and *Meso2*) and rostral (*EphA4*) markers. None of them revealed subdivision of the ectopic somites. *Hairy1* [0 of 22 embryos (0/22)], *Meso2* (0/22), and *EphA4* (0/19) were not expressed (Fig. 3, A to C); *LFng* (22/24) (Fig. 3D) and *Hairy2* (8/8) (Fig. 3E) were expressed weakly and uniformly throughout the somites; and *Uncx4.1* (13/19) (Fig. 3F) was patchy (Fig. 3F). Therefore, ectopic somites seem to lack coherent rostrocaudal identity, because the patterns of different genes are inconsistent with each other. As neural crest cells and motor axons normally only migrate through the rostral half of the sclerotome (*15*), we used this as an additional test of somite patterning. An ectopic GFP-somite was grafted instead of a normal somite in a secondary host (Fig. 1L). At stage 22 to 25, the patterns of motor axon growth (Fig. 3, G to I) and neural crest migration (Fig. 3, J to O) were disrupted. Abnormalities included an enlarged gap between motor roots (Fig. 3, G to I), fusion of adjacent ventral roots and dorsal root ganglia (Fig. 3, J to L), or several small ganglia formed within a grafted somite (Fig. 3, M to O), as if the somite contained random islands of permissive (non-caudal) cells exploited by axons and crest cells. These results suggest that the ectopic somites are not subdivided into rostral and caudal halves, consistent with the proposal (*16*) that the clock is required for this feature of segmentation.

During normal development, the occipital somites (the most cranial four or five somites) form almost simultaneously rather than in sequence (movie S3) and lack expression of some rostral and/or caudal markers (*17*–*19*). Could the ectopic somites be occipital? We examined expression of Hox genes (*20*, *21*) (Fig. 4, A to P): *Hoxb3* (Fig. 4, A and C) and *Hoxb4* (Fig. 4, E and G) were both expressed (Fig. 4, B, D, F, and H), suggesting that they are not occipital. *Hoxb6* and *Hoxb9* were not expressed (Fig. 4, F and J), suggesting that they are cervical (somite eight or nine). The posterior PS of stage-5 donor embryos expresses similar genes: *Hoxb3* and *b4*, but not *b6* or *b9* (Fig. 4, A, E, I, and M); the latter start to be expressed later (stage 7 or 8) (Fig. 4, C, G, K, and O). We therefore tested

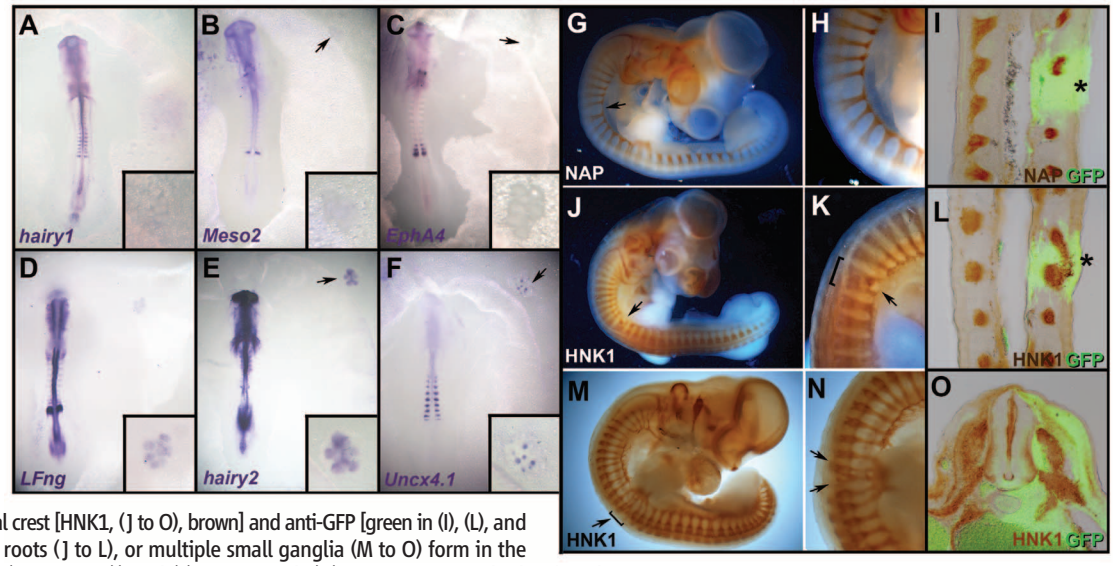


Fig. 3. Ectopic somites are not subdivided into rostral and caudal halves. (**A** to **F**) Ectopic somites were analyzed for expression of caudal (*Hairy1*, *Meso2*, *LFng*, *Hairy2*, and *Uncx4.1*) and rostral (*EphA4*) markers. *Hairy1* (A), *Meso2* (B), and *EphA4* (C) are not expressed; *LFng* and *Hairy2* (D and E) are weak and uniform; and *Uncx4.1* is expressed as random patches (F). Insets show a magnified view of the graft. (**G** to **O**) As a further test of rostrocaudal patterning, embryos grafted as in Fig. 1L were stained for motor axons [neurofilament-associated protein NAP, (G to I), brown] or neural crest [HNK1, (J to O), brown] and anti-GFP [green in (I), (L), and (O)]. A large gap (G to I), fused roots (J to L), or multiple small ganglia (M to O) form in the ectopic somite (arrows, asterisks). Sections (I) and (L) are coronal, (O) is transverse at the level of the graft.

whether somites made from PSs from older embryos (stage 8) express these markers. Indeed, they do (Fig. 4, L and P). This confirms that the Hox code imparting axial identity to cells is already present in the PS (*22*), independently of the segmentation clock (*6*), and suggests that the axial identity of the ectopic somites is specified according to which Hox genes are expressed in the posterior PS at the time of explantation, even though this region does not normally contribute to somites. Therefore, either exit of cells from the PS or, more likely, inhibition of BMP by Noggin arrests the molecular clock controlling expression of Hox genes that impart axial identity (*23*). In vivo, this may happen as presomitic cells leave the BMP-expressing PS and lie next to the notochord, the endogenous source of Noggin.

The clock-and-wavefront model requires both an oscillator and a wave. In zebrafish, changing the period of molecular oscillations affects somite number and size (*5*, *6*). We show that somites can form without oscillations of segmentation clock genes; all of their properties are normal, except for their rostrocaudal subdivision. Moreover, waves and gradients are also unnecessary, because the spatial organization and simultaneous formation of the ectopic somites does not seem compatible with this. We therefore propose that the main function of the clock is to subdivide somites into rostral and caudal halves and to couple this to somite formation.

If clock-and-wavefront mechanisms are not required to control somite formation, what does? Our observations implicate local cell-cell interactions. Embryological experiments (*24*) suggest that somites are self-organizing structures, regulated by intrinsic properties of the cells and packing constraints for cells undergoing mesenchymal-to-epithelial conversion as they form spheres. We tested this in computer simulations using CompuCell-3D (*25*, *26*), with the following assumptions: (i) A cell mass is exposed to Noggin evenly and simultaneously; (ii) in response, cells polarize and elongate; (iii) polarized cells secrete extracellular matrix; (iv) polarized cells have to be exposed to extracellular space at both their apical and basal surfaces; (v) tight junctions form at the apical ends; and (vi) misplaced cells rearrange their polarity and attach to their appropriate ends (*27*). This causes cells to become arranged in spherical masses around a lumen (movie S4) (*9*). After a transition period of intense cell rearrangement, the somites stabilize. The number of cells they contain is relatively invariant, and their structure is similar to that seen in vivo. There is no tendency to merge into a giant structure, nor do very small stable somites form. We propose that somite size and shape can be controlled entirely by local cell interactions, such as adhesion and packing constraints of cells transitioning between the mesenchyme and a polarized epithelium (*28*). Inhibition of BMP by Noggin may be a trigger for this conversion, consistent with the abnormal somite formation in Noggin-null mice (*29*), and may also "freeze" molecular determinants of axial identity (Hox code). In normal embryos, the segmentation clock and associated wave are likely to play a role in regulating the timing of somite formation and coupling this to the subdivision of each somite into rostral and caudal subcompartments.

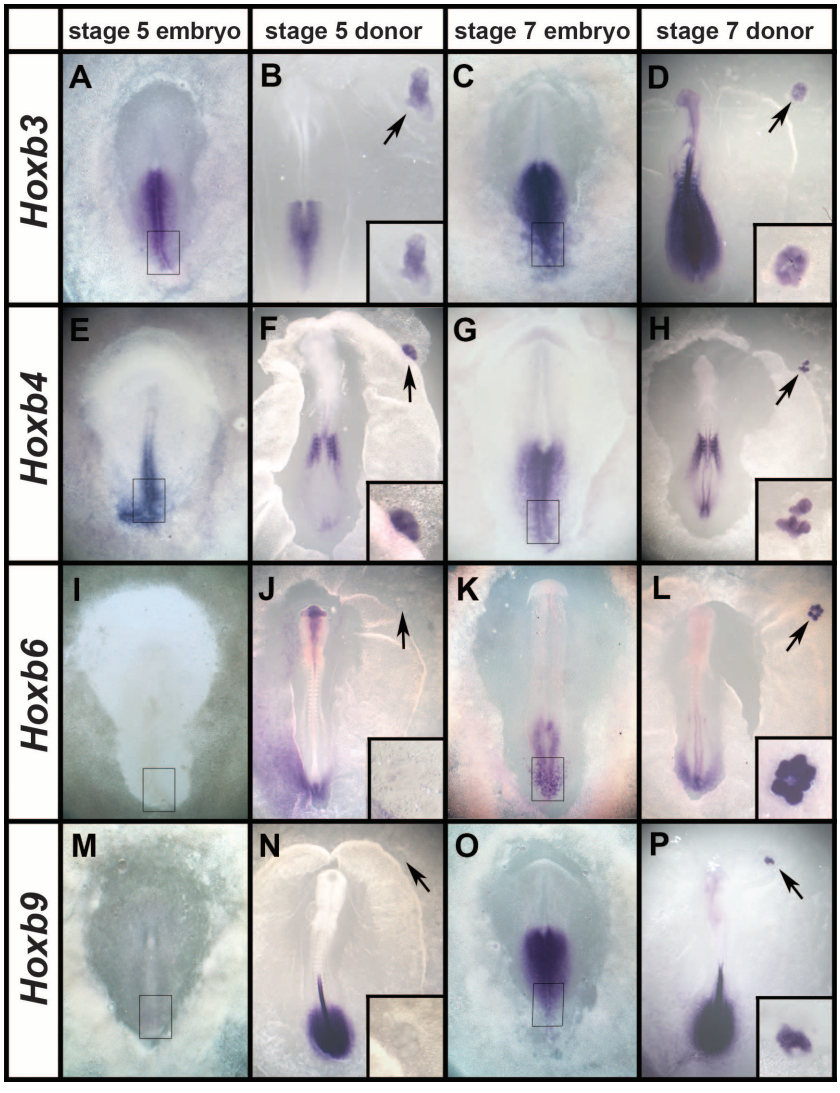


Fig. 4. Ectopic somites have trunk identity, fixed according to the Hox genes expressed in the donor PS. (**A** to **P**) At stage 5, the posterior PS expresses Hoxb3 (A) and b4 (E), but not b6 (I) or b9 (M). Ectopic somites made from posterior streak explants from these stages show a similar pattern of expression (B, F, J, and N). At stages 7 and 8, the posterior streak expresses all four genes (C, G, K, and O), as do the ectopic somites formed from it (D, H, L, and P). Arrows point to the graft region, shown magnified in the insets.

References and Notes

1. I. Muñoz-Sanjuán, A. H. Brivanlou, in *Gastrulation: From Cells to Embryo*, C. D. Stern, Ed. (Cold Spring Harbor Press, New York, 2004), pp. 475–489.
2. J. Dubrulle, M. J. McGrew, O. Pourquié, *Cell* **106**, 219–232 (2001).
3. J. Dubrulle, O. Pourquié, *Nature* **427**, 419–422 (2004).
4. J. Cooke, E. C. Zeeman, *J. Theor. Biol.* **58**, 455–476 (1976).
5. Y. Harima, Y. Takashima, Y. Ueda, T. Ohtsuka, R. Kageyama, *Cell Rep.* **3**, 1–7 (2013).
6. C. Schröter, A. C. Oates, *Curr. Biol.* **20**, 1254–1258 (2010).
7. A. Tonegawa, N. Funayama, N. Ueno, Y. Takahashi, *Development* **124**, 1975–1984 (1997).
8. A. Streit, C. D. Stern, *Mech. Dev.* **85**, 85–96 (1999).
9. See materials and methods and other supplementary materials on *Science* Online.
10. I. Palmeirim, D. Henrique, D. Ish-Horowicz, O. Pourquié, *Cell* **91**, 639–648 (1997).
11. A. C. Oates, L. G. Morelli, S. Ares, *Development* **139**, 625–639 (2012).
12. V. Hamburger, H. L. Hamilton, *J. Morphol.* **88**, 49–92 (1951).
13. R. S. Beddington, P. Martin, *Mol. Biol. Med.* **6**, 263–274 (1989).
14. C. D. Stern, S. E. Fraser, R. J. Keynes, D. R. Primmett, *Development* **104** (suppl.), 231–244 (1988).
15. R. J. Keynes, C. D. Stern, *Nature* **310**, 786–789 (1984).
16. Y. Takahashi, T. Inoue, A. Gossler, Y. Saga, *Development* **130**, 4259–4268 (2003).

17. C. Jouve, T. Iimura, O. Pourquie, *Development* **129**, 1107–1117 (2002).
18. T. M. Lim, E. R. Lunn, R. J. Keynes, C. D. Stern, *Development* **100**, 525–533 (1987).
19. S. Rodrigues, J. Santos, I. Palmeirim, *Gene Expr. Patterns* **6**, 673–677 (2006).
20. A. C. Burke, C. E. Nelson, B. A. Morgan, C. Tabin, *Development* **121**, 333–346 (1995).
21. S. J. Gaunt, *Int. J. Dev. Biol.* **38**, 549–552 (1994).
22. T. Iimura, O. Pourquié, *Nature* **442**, 568–571 (2006).
23. S. A. Wacker, C. L. McNulty, A. J. Durston, *Dev. Biol.* **266**, 123–137 (2004).
24. C. D. Stern, R. Bellairs, *J. Embryol. Exp. Morphol.* **81**, 75–92 (1984).
25. S. D. Hester, J. M. Belmonte, J. S. Gens, S. G. Clendenon, J. A. Glazier, *PLOS Comput. Biol.* **7**, e1002155 (2011).
26. M. H. Swat *et al.*, *Methods Cell Biol.* **110**, 325–366 (2012).
27. G. G. Martins *et al.*, *PLOS ONE* **4**, e7429 (2009).
28. Y. Nakaya, S. Kuroda, Y. T. Katagiri, K. Kaibuchi, Y. Takahashi, *Dev. Cell* **7**, 425–438 (2004).
29. J. A. McMahon *et al.*, *Genes Dev.* **12**, 1438–1452 (1998).

Acknowledgments: This work was funded by Medical Research Council (grant G0700095) and the NIH (National Institute of General Medical Sciences grants R01GM076692 and R01GM077138). N. De-Oliveira helped with cloning, C. Thrasivoulou with confocal microscopy, and A. Norris with vibratome sectioning. S. Evans, J. Sharpe, A. Streit, and S. Thorsteinsdóttir provided information and constructive comments.

Supplementary Materials

www.sciencemag.org/content/343/6172/791/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Table S1
References
Movies S1 to S4

23 October 2013; accepted 13 December 2013
Published online 9 January 2014;
10.1126/science.1247575

An Antifreeze Protein Folds with an Interior Network of More Than 400 Semi-Clathrate Waters

Tianjun Sun, Feng-Hsu Lin, Robert L. Campbell, John S. Allingham, Peter L. Davies*

When polypeptide chains fold into a protein, hydrophobic groups are compacted in the center with exclusion of water. We report the crystal structure of an alanine-rich antifreeze protein that retains ~400 waters in its core. The putative ice-binding residues of this dimeric, four-helix bundle protein point inwards and coordinate the interior waters into two intersecting polypentagonal networks. The bundle makes minimal protein contacts between helices, but is stabilized by anchoring to the semi-clathrate water monolayers through backbone carbonyl groups in the protein interior. The ordered waters extend outwards to the protein surface and likely are involved in ice binding. This protein fold supports both the anchored-clathrate water mechanism of antifreeze protein adsorption to ice and the water-expulsion mechanism of protein folding.

A driving force for protein folding is formation of a hydrophobic core (*1*, *2*). As aliphatic and aromatic side chains pack together in the protein core, they release constrained waters into the surrounding solvent with an overall gain in entropy that helps power the folding process. Therefore, almost all globular proteins reported to date have a dry protein core. There are two main schools of thought for how a protein's hydrophobic core is formed. In the dewetting mechanism, waters collectively evaporate from the partially formed core. This is followed by the spontaneous collapse of the core, which stabilizes the protein by reducing the solvent-accessible surface area of core residues (*3*). In the expulsion mechanism, an initial structural collapse forms a near-native intermediate with a partially solvated hydrophobic core. This is followed by water expulsion from the hydrophobic core to form the native state (*4*). In this model, waters are thought to function as a lubricant that helps the hydrophobic core find its optimally packed state (*4*). Here we report the crystal structure of an antifreeze protein (AFP) with a water-rich core that offers an alternative view on the role of water in protein folding.

Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON K7L 3N6, Canada.

*Corresponding author. E-mail: peter.davies@queensu.ca

Antifreeze proteins are a class of proteins that adsorb to the surface of ice crystals to prevent their growth. We determined the crystal structure of the AFP Maxi, a large isoform of the 3-kD, alanine-rich type I AFP from winter flounder (*Pseudopleuronectes americanus*). Type I AFP is a monomeric α helix with an 11-residue periodicity (*5*) that binds to a pyramidal ice plane (*6*). Its ice-binding residues (Thr, i+4 Ala, i+8 Ala) are arrayed along one face of the helix (*7*). Ice-binding residues are thought to organize and stabilize ordered waters to merge with, and freeze to, the quasi-liquid layer of water next to the ice lattice (*8*–*10*). Maxi is five times as long as type I and forms a homodimer with no stable monomer form, but is otherwise similar in alanine richness (65%), helicity (>95% α helix), and 11-residue periodicity (*11*–*13*).

The crystal structure of Maxi was determined at 1.8 Å resolution from needle-shaped crystals (table S1) grown in arginine buffer at pH 9.6. Maxi is a rodlike four-helix bundle with a length of 145 Å and an average diameter of 22 Å (Fig. 1A). Both 290 Å–long helix monomers fold exactly in the middle through 180° so that their N and C termini are side by side. In the dimer, the two hairpins are aligned in an antiparallel manner without overlap such that the four-helix bundle has a twofold rotational axis of symmetry (indicated by the central curved arrow). The two N-terminal helices lie adjacent to each other in an antiparallel orientation, as do the two C-terminal helices. At the secondary-structure level, Maxi is composed of tandem 11-residue repeats (T/IxxxAxxxAxx, where x is any residue) that each form three helical turns with an average of 3.7 residues per turn (R1-3 and R1′-3′), as opposed to 3.6 residues per turn in the classic α helix. The only departures from this pattern are the central sections of two seven-residue segments (two helical turns each) and the terminal capping regions. The capping regions comprise the N and C termini of one monomer and the hairpin loop region of the other (Fig 1B). Notably, the two chains of Maxi associate with minimal protein-protein interactions, as do the two arms of the hairpins (Fig. 1C). This is in sharp contrast to a standard four-helix bundle protein like "Repressor of primer" (Rop), where the longer aliphatic and aromatic side chains form a compact hydrophobic core with only two waters inside (Fig. 1D). Rop was selected for comparison because, like Maxi, it is also a dimer of hairpin α helices with a nonoverlapping antiparallel alignment (*14*).

Direct contact between the Maxi monomers is largely limited to the capping structures and the center region. In the capping structures, Leu^{94} and Tyr^{103} near the hairpin loop of one monomer, together with Ile^{2} at the N terminus and Phe^{191} at the C terminus of the other monomer, pack together to form a local hydrophobic core with stacking of the two aromatic side chains (Fig. 1B). On the hairpin loop, Lys^{100} donates hydrogen bonds to cap the backbone carbonyl groups of Ala^{191}, Ala^{192}, and Ala^{194} on the C terminus, whereas the carbonyl groups of Leu^{94} and Ile^{97} and the side chain of Asn^{98} accept hydrogen bonds from Arg^{7}. In the center region, Ile^{51} and Ile^{54} from both N-terminal helical arms of Maxi form a hydrophobic cluster, as does Val^{144} and Val^{148} from both C-terminal helical arms. However, the intervening helical repeats pack loosely through sparse van der Waals interactions, as shown by a representative cross section through one repeat (Fig. 1C).

The apparent loose packing of the four helices in the 11-residue repeat regions generates internal space that is just wide enough to accommodate a single layer of water (Fig. 1C and Fig. 2, A and B). The water molecules that occupy the gap form an extensive polypentagonal network around the inner-projecting residues (mostly Ala

and Thr, fig. S1), with an occasional tetragonal and hexagonal water ring. In the 11-residue repeat regions, the network comprises two monolayers of polypentagonal waters (Fig. 2, A and B) that cross each other at ~90° (Fig. 3 and fig. S3). The intrachain sheet, which lies between the N-terminal arms and the C-terminal arms of Maxi, contains 19 pentagonal rings per 11-residue repeat (labeled A to R in Fig. 2A). The pattern for each repeat is almost identical, and three repeats spanning nine helical turns are shown. Occasional large holes in the sheet are generated where side chains like Thr^{132} and Ala^{68} from the same monomer form close protein-protein contacts. The interchain water web that forms the interface between the two monomers contains seven pentagonal rings (labeled S to Z in Fig. 2B). Again, a few large holes appear in this sheet due to rare interchain side-chain contacts like those between Thr^{122} and Ala^{173} from different monomers. A molecular dynamics simulation of Maxi in a box of water

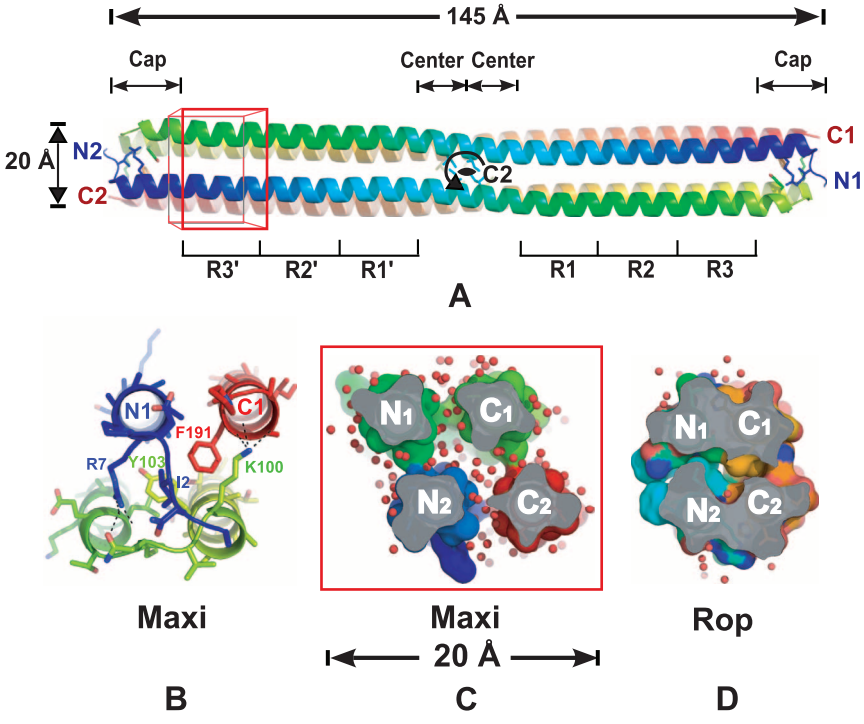


Fig. 1. Four-helix bundle structure of Maxi has an open core. (**A**) Four-helix bundle structure of Maxi in ribbon format, showing how the antiparallel hairpin monomers 1 and 2 align. N and C termini are indicated. The dyad axis of symmetry is shown by the arrow C_2. Capping regions, 11-residue repeats, and center sections are labeled as Cap, R, and Center, respectively. Residues in van der Waals contact in the cap and central regions are indicated by their side chains. (**B**) End-on view of the capping structure of Maxi; hydrogen bonds are illustrated by black dotted lines. (**C**) End-on view of a cross section of Maxi in surface representation marked in (A) by the red box. Red spheres are crystallographic waters present in the 20 Å–deep section. (**D**) Similar-sized section through Rop, a representative four-helix bundle structure.

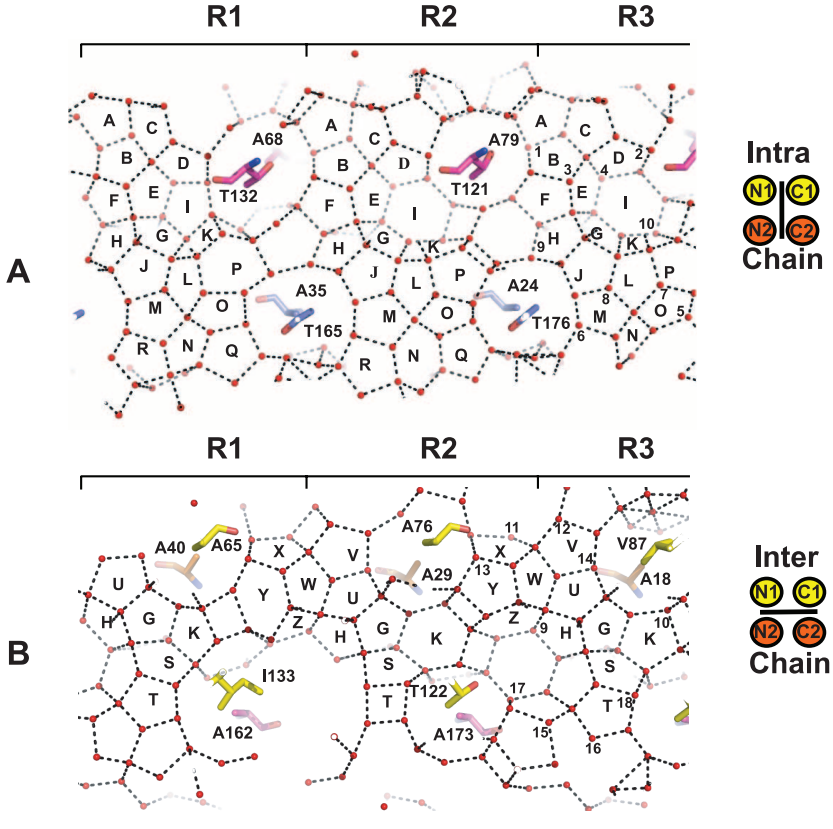


Fig. 2. The interior water structure associated with the repeat regions of Maxi. (**A**) Sheet of pentagonal intrachain water rings labeled A to R. Waters are showed as red spheres and hydrogen bonds by black dashes. Residues forming intrachain contacts are identified and shown in stick representation. Waters that hydrogen bond to the carbonyl groups of Maxi in R3 are numbered. (**B**) A sheet of interchain pentagonal waters at right angles to the sheet in (A), as indicated in the schematic diagrams on the right. Notations are the same as in (A). The pentagonal rings in this sheet are labeled from S to Z. In addition, rings G, H, and K at the intersection of two sheets are also labeled.

showed that the predicted interior water densities match extremely well with the ones in the protein core of the crystal structure (fig. S2).

Fifty years ago, Scheraga *et al.* theorized that a partial cage of pentagonal water rings will form when hydrophobic side chains in α helices or β sheets are separated by a single layer of water molecules (*15*). The formation of pentagonal rings has negative free energy when two hydrated aliphatic side chains approach each other during protein folding to a distance where they are separated by a single water layer. This proposal has been borne out in the present work, where the spaces inside Maxi's four-helix bundle are just wide enough to allow a single layer of waters to fit in and form a semi-clathrate structure. Before this study, five pentagonal "ice-like" water clusters had been reported in crystals of crambin [Protein Data Bank (PDB) code: 1CRN], a 46-residue seed storage protein, but only as a result of intermolecular packing (*16*). It was predicted that these localized pentagonal clusters might occur not only at intermolecular hydrophobic contacts, but also around adjacent, hydrophobic side chains in α helices or β sheets. This prediction has been supported by a molecular dynamics simulation study of streptavidin (*17*), where a five-membered water ring was formed in between two hydrophobic groups in the binding cavity. The crystal structure of Maxi has revealed the natural association of pentagonal water clusters within a protein on a large scale.

Two-dimensional phases of water are produced by nanoscale confinement between nonpolar materials (nanopores) (*18*). These thin water layers exhibit unusual properties compared to bulk water and are typically studied by molecular simulation. Maxi contains a monolayer of amorphous ice/water between mainly hydrophobic surfaces. The pattern of the water sheet in Maxi is different from those observed by molecular simulations and experimental approaches (*19*, *20*). This is mainly due to the inner-surface architecture of the protein. Because most nanopores have flat surfaces, water molecules inside tend to arrange themselves in layers parallel to the surface (*21*). By contrast, the inward-pointing side chains that form the inner surface of Maxi are molecularly rough, so the five-membered water rings form cages on individual residues, illustrating their semi-clathrate hydrate structure (Fig. 4D). In addition, unlike most nanopores used for simulations, the inner surface of Maxi interacts with about one in four of the waters through hydrogen bonds, in addition to van der Waals interactions (as described below).

The 3.7-residue repeat of the helices in Maxi results in altered Φ and Ψ angles compared to those of classic α helices. This deviation likely causes the backbone carbonyl groups to be slightly tilted outwards (*13*), allowing them to form the key intrahelical hydrogen bonds while also hydrogen bonding with solvent waters (Fig. 4B), a duality not seen in Rop and other typical α helices (Fig. 4C). In addition, Maxi is an alanine-rich protein, and these small side chains make the backbone carbonyl groups more exposed to solvent. Therefore, most of the carbonyl groups in Maxi are involved in hydrogen-bonding interactions with waters, which helps to keep this rather hydrophobic protein highly solvated and freely soluble in flounder blood. For typical protein structures determined at 2.0 Å resolution, the number of water molecules located by crystallography is roughly equal to the number of amino acid residues (*22*). By contrast, in the asymmetric unit of Maxi, the ratio of water molecules to amino acid residues is almost three times this value (2.9), even though Maxi does not contain many polar residues to hydrogen bond with waters. Because most of the residues in the internal space are also alanine, hydration occurs around all four helices. Maxi uses the interior bifurcated carbonyl groups to help anchor the more than 400 internal ordered waters, ~25% of which are directly involved in backbone hydrogen-bonding interactions. The hydrogen-bonding interactions in R3 are listed in table S2.

The hydrated protein core of Maxi suggests that this protein uses the water-expulsion folding mechanism but does not complete the water-discharge step. The alanine richness of the protein core may have provided an ideal opportunity to see retention of waters during the folding. These waters serve to glue the four helical arms of Maxi together through a stabilizing network of hydrogen bonds that are anchored to backbone carbonyl groups in the interior (*23*). Because water-mediated protein association is less

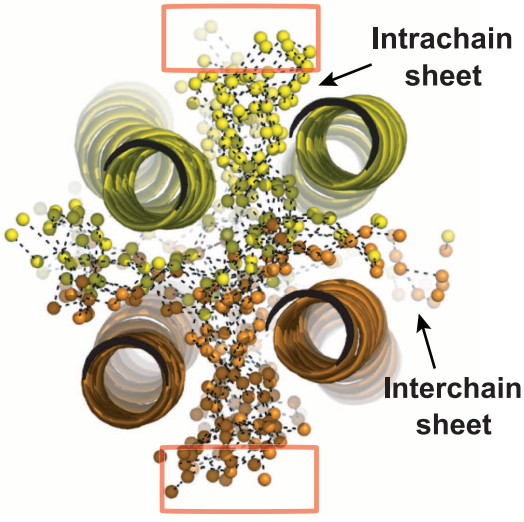


Fig. 3. Interior water network in Maxi. Section through Maxi showing network of interior waters as spheres colored by association with the yellow or orange monomers. Hydrogen bonds are indicated by black dotted lines. Waters that make multiple matches to the ice lattice and are not sterically hindered by the protein are boxed in red.

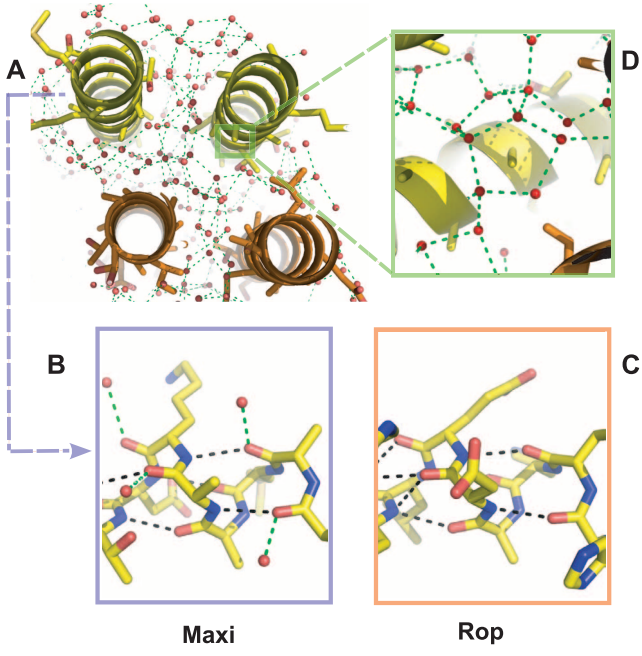


Fig. 4. Water interactions with Maxi. (**A**) Cross section through R2, illustrating the interior water network and the intersection of the intra- and intermolecular water sheets. Waters are shown as red spheres, and hydrogen bonds are shown as green dotted lines. (**B**) A section of helix from Maxi in stick format showing bifurcation of intramolecular hydrogen bonds. Hydrogen bonds between water (red spheres) and carbonyl O atoms are shown as green dotted lines; intrahelical hydrogen bonds are shown as black dashed lines. N atoms are blue and O atoms, red. (**C**) A section of helix from Rop in stick format showing normal intramolecular hydrogen bonds. (**D**) An example of semi-clathrate hydrate structure in the internal space is shown as a close-up view of the green boxed region in (A).

stable than direct protein association (*4*), this could explain why this antifreeze protein is thermolabile and irreversibly denatures at temperatures above 16°C.

In addition to the internal waters, a second unexpected feature of Maxi's structure was that the putative ice-binding residues (Thr/Ile, i+4 Ala, i+8 Ala) occur on the inward-pointing surfaces of all four helices (colored pink in fig. S1). These residues were identified by homology to the small, monomeric type I AFP isoform that binds ice through this highly conserved hydrophobic face using regularly spaced methyl groups from the Ala and Thr/Ile (*7*). The same ice-binding residues are well conserved in each 11-residue repeat of Maxi and occupy equivalent quadrants of the helices throughout the molecule. Their side chains have a function similar to that of ice-binding residues because they cooperate to form and anchor the interior ordered waters that stabilize the protein structure (Fig. 4A and table S2). The average B-factor for all waters in the crystallographic asymmetric unit is 26 $Å^2$. However, most of the waters in the inner network have B-factors that are lower than 20 $Å^2$ (colored blue and green in fig. S4). Outer waters (colored red) are more disordered.

Three lines of evidence suggest that the crystal structure of Maxi is representative of the solution form and that the protein does not undergo a conformational change to bind to ice crystals through the Thr/Ile, i+4 Ala, i+8 Ala residues. First, Olijve *et al*. have determined the solution structure of Maxi at low temperature using small-angle x-ray scattering (*24*). In solution, Maxi has a cylindrical shape and dimensions that are consistent with the four-helix bundle crystal structure and eliminate a previously proposed model where the two helices form a fully extended helix dimer (*13*). Second, the measurement of intrinsic fluorescence transfer between Tyr and Phe suggests that the exquisite capping structure seen in the crystal (Fig. 1B) is also present in solution (fig. S3). Lastly, cross-linking experiments show that the juxtaposition of residues on neighboring chains of the helix bundle is the same in solution as it is in the crystal (figs. S4 and S5). Moreover, where these cross-links prevent the opening of the helix bundle to expose the "ice-binding residues," there is no appreciable loss of antifreeze activity (table S3).

How then does Maxi bind to ice? Close examination of Maxi's crystal structure shows positioned waters extending outwards between all four helices from the core to the surface. At the periphery, they form a network of ordered waters that are unobstructed by the helices and available to merge and freeze with the quasi-liquid layer on the surface of ice (*8*–*10*). It is possible to fit clusters of crystallographic surface waters at the face formed by the N- and C-terminal helices (boxed regions in Fig. 3) into the ice lattice on numerous planes, three of which are shown in fig. S6, A to C. Consistent with this result, when we used fluorescence-based ice plane affinity analysis (*25*) to determine which ice planes adsorb Maxi, all surfaces of the single ice-crystal hemisphere were bound by the fluorescently tagged AFP (fig. S7D). Unlike other AFPs characterized to date, which have ice-binding residues located on their surface, Maxi is the only one in which these residues are buried inside. This further supports the anchored-clathrate water mechanism by which AFPs adsorb to ice (*10*) because it suggests that this AFP cannot directly bind to its ligand but must do so through the ordered surface waters.

References and Notes

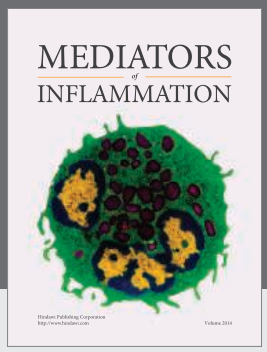
1. W. Kauzmann, *Adv. Protein Chem.* **14**, 1–63 (1959).
2. C. Tanford, J. Reynolds, *Nature's Robots: A History of Proteins* (Oxford Univ. Press, New York, 2001).
3. P. R. ten Wolde, D. Chandler, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6539–6543 (2002).
4. Y. Levy, J. N. Onuchic, *Annu. Rev. Biophys. Biomol. Struct.* **35**, 389–415 (2006).
5. F. Sicheri, D. S. Yang, *Nature* **375**, 427–431 (1995).
6. C. A. Knight, C. C. Cheng, A. L. DeVries, *Biophys. J.* **59**, 409–418 (1991).
7. J. Baardsnes *et al*., *FEBS Lett.* **463**, 87–91 (1999).
8. C. Yang, K. A. Sharp, *Proteins* **59**, 266–274 (2005).
9. D. R. Nutt, J. C. Smith, *J. Am. Chem. Soc.* **130**, 13066–13073 (2008).
10. C. P. Garnham, R. L. Campbell, P. L. Davies, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7363–7367 (2011).
11. C. B. Marshall, G. L. Fletcher, P. L. Davies, *Nature* **429**, 153 (2004).
12. L. A. Graham, C. B. Marshall, F. H. Lin, R. L. Campbell, P. L. Davies, *Biochemistry* **47**, 2051–2063 (2008).
13. C. B. Marshall, A. Chakrabartty, P. L. Davies, *J. Biol. Chem.* **280**, 17920–17929 (2005).
14. D. W. Banner, M. Kokkinidis, D. Tsernoglou, *J. Mol. Biol.* **196**, 657–675 (1987).
15. H. A. Scheraga, G. Nemethy, I. Z. Steinberg, *J. Biol. Chem.* **237**, 2506–2508 (1962).
16. M. M. Teeter, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6014–6018 (1984).
17. T. Young, R. Abel, B. Kim, B. J. Berne, R. A. Friesner, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 808–813 (2007).
18. J. C. Rasaiah, S. Garde, G. Hummer, *Annu. Rev. Phys. Chem.* **59**, 713–740 (2008).
19. K. Koga, H. Tanaka, X. C. Zeng, *Nature* **408**, 564–567 (2000).
20. J. Bai, C. A. Angell, X. C. Zeng, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5718–5722 (2010).
21. J. Bai, X. C. Zeng, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21240–21245 (2012).
22. O. Carugo, D. Bordo, *Acta Crystallogr. D Biol. Crystallogr.* **55**, 479–483 (1999).
23. M. Ahmad, W. Gu, T. Geyer, V. Helms, *Nat. Commun.* **2**, 261 (2011).
24. L. C. Olijve *et al*., *RSC Advances* **3**, 5903 (2013).
25. C. P. Garnham *et al*., *Biochemistry* **49**, 9063–9071 (2010).

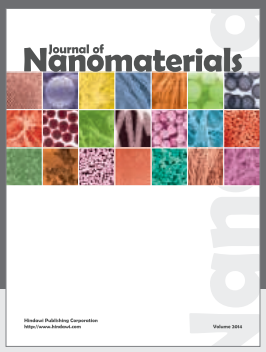
Acknowledgments: We are grateful to D. McLeod and K. Munro of the Protein Function Discovery Facility at Queen's University for mass spectrometry and intrinsic fluorescence analyses, respectively. We thank E. Lazo and V. Stojanoff of the X6A beamline at Brookhaven National Laboratories for help with x-ray data acquisition and processing. This research was funded by the Canadian Institutes of Health Research. P.L.D. holds the Canada Research Chair in Protein Engineering. J.S.A holds a Canada Research Chair (Tier 2) in Structural Biology and is an Ontario Early Researcher Award recipient. The coordinates and x-ray data for Maxi have been deposited with the PDB with accession code 4EK2.

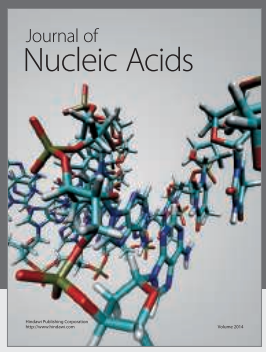
Supplementary Materials

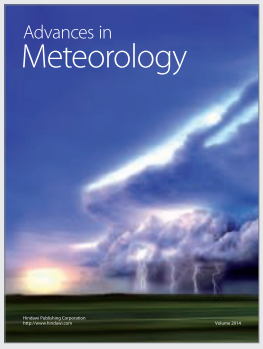
www.sciencemag.org/content/343/6172/795/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S3
References (*26*–*35*)

21 October 2013; accepted 8 January 2014
10.1126/science.1247407



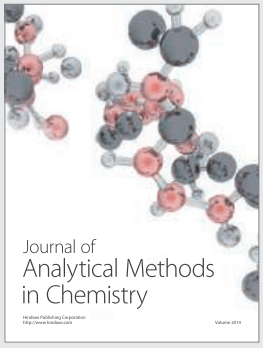


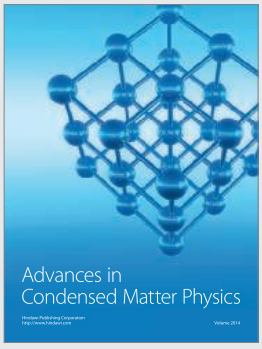


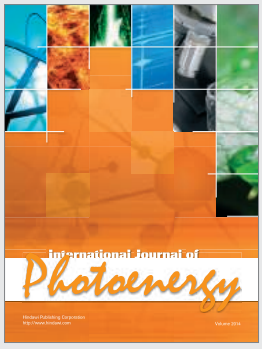


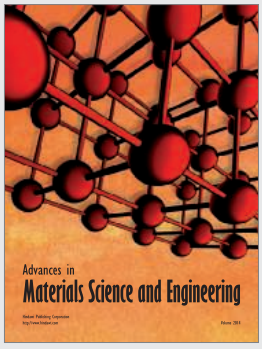
Hindawi

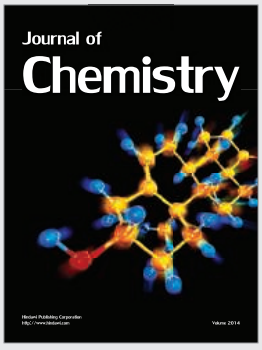
Submit your manuscripts at
http://www.hindawi.com

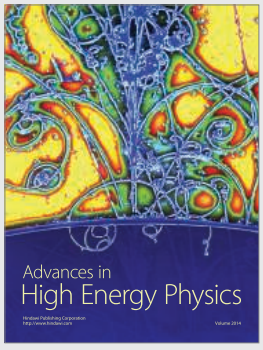












2014 (30th) Japan Prize Laureates

"Electronics, Information and Communication" Field

Contribution to building the cornerstone for today's information society

"By developing semiconductor lasers for high capacity, long distance optical fiber transmission and enabling people to communicate via submarine optical cables around the world"

Dr. Yasuharu Suematsu
Japan

How they have contributed to Society through science and technology?

Dr. C. David Allis
USA

"Life Science" Field

Contribution to progressing studies of life science

"By discovering and demonstrating histone modification as a fundamental gene regulatory mechanism and advancing new areas of epigenetic research"

Japan prize selection criteria

The Japan Prize is awarded to scientists and engineers from around the world who are recognized as having achieved original and dramatic accomplishments that greatly enhance the progress of science and technology, thereby contributing to the peace and prosperity of mankind.

JAPAN PRIZE

▶ Visit for details on the Japan Prize and its laureates
www.japanprize.jp

JAPAN PRIZE

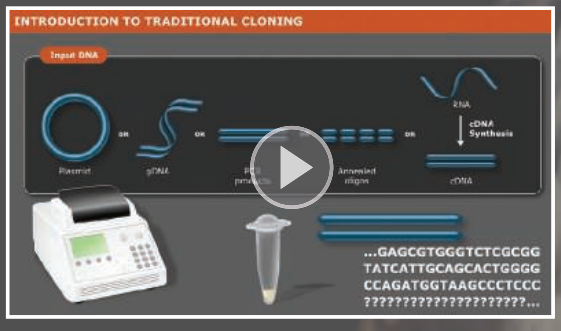
NEW ENGLAND BioLabs® Inc.

Clone with Confidence.

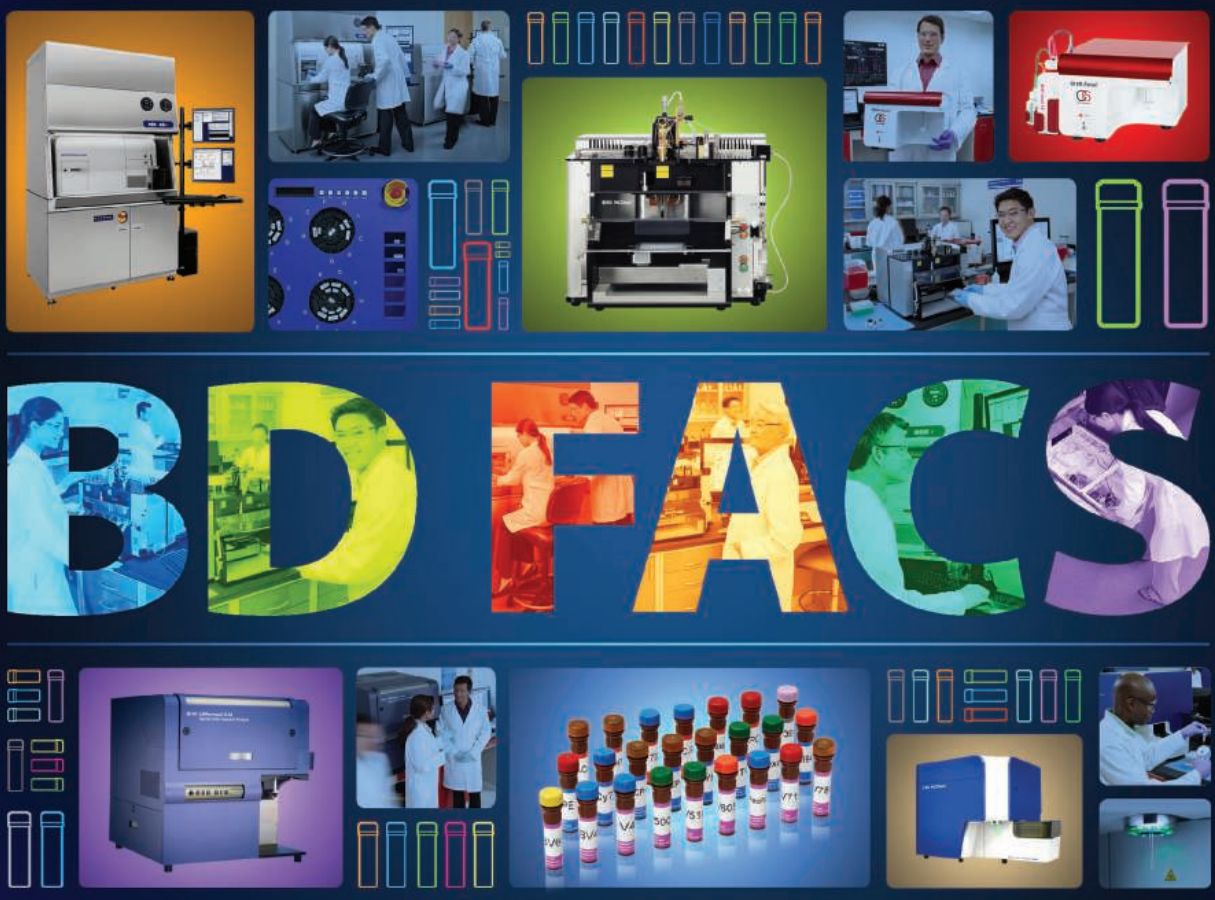
Whether you are performing your first cloning experiment, or constructing multi-fragment gene assemblies, NEB® has the solution for you. Our high quality reagents are available for every workflow, and include specialized enzymes, competent cells, and novel solutions – such as Gibson Assembly®. When you are looking to clone with confidence, think of NEB.

Explore the wise choice at **CloneWithNEB.com.**

Visit CloneWithNEB.com to view online tutorials describing various cloning workflows.



BD FACS™ Research Systems to Empower the Next Generation of Research



BD Helping all people live healthy lives

40 years of trust.

Across the past 40 years we've been honored to be the brand you've trusted to facilitate and speed your work to discover more about life around us and to improve the human condition.

BD FACS™ brand research systems stand for quality, relentless innovation, and tireless dedication to service.

Our platforms continue to keep pace with the accelerating speed of discovery, out-innovating the field so you have reliable and consistent results from experiment to experiment, operator and sample protection from fully integrated biosafety and cell sorting, more space on your benchtop, and more flexibility for your experiments.

The BD FACS brand assurance is realized further through our knowledgeable associates around the world, who provide training, application support, and field services so you can achieve optimal performance and productivity.

We thank you for trusting your discoveries to our technologies and wish you success in 2014.

For more information please visit **bdbiosciences.com/go/facs.**

BD Biosciences
2350 Qume Drive
San Jose, CA 95131
bdbiosciences.com

WEBINAR

Wednesday, March 12, 2014

12 noon Eastern, 9 a.m. Pacific, 4 p.m. UK, 5 p.m. CET

Speakers

Miguel E. Quiñones-Mateu, Ph.D.
UHCMC/CWRU
Cleveland, OH

Second speaker to be announced.

The Sequencing Continuum for Clinical Research:

From Sanger to Next Gen

Researchers today have a many choices when deciding which sequencing technology to use for their clinical research. Sanger sequencing, long the "gold standard" in clinical research sequencing technology, has met significant competition with the advent of next generation sequencing (NGS). NGS technologies are now commonplace in clinical research laboratories where they have enabled rapid advances in the gathering and analysis of genetic information. However, with these advances have come additional challenges involving validation as these technologies become more widespread and move closer to future clinical application. In this webinar, our expert speakers will discuss the relative benefits of Sanger and NGS technologies and their application in different fields of clinical research.

During the webinar, the speakers will:

- Compare and contrast Sanger sequencing and NGS approaches
- Describe the most relevant applications for both Sanger sequencing and NGS
- Discuss future clinical research applications of both Sanger sequencing and NGS
- Answer your questions live!

Webinar sponsored by

ion torrent
by life technologies™

Brought to you by the *Science*/AAAS Custom Publishing Office

Science
AAAS

DEEP-IMAGING MULTIPHOTON SYSTEM

The new FluoView FVMPE-RS is a dedicated multiphoton microscope system, which enables high-precision, ultrafast scanning and stimulation. This system allows researchers to see deep within specimens, take measurements at the highest speeds, and capture images, even when working under the most demanding conditions. With its high-speed and high-precision performance, the FVMPE-RS is designed for electrophysiology and optogenetics studies. It is also a good match for applications such as high-speed calcium and in vivo imaging, peristalsis and blood flow studies, mosaic imaging, connectomics and functional brain imaging, stem cell research, and any field that requires precise co-localization, uncaging, simultaneous imaging/stimulation, extensive real-time signal processing, or multipoint mapping. Its design offers ready adaptability for researchers who design their own custom-built systems as well. The precision timing on the system allows for microsecond repeatability and control of multiple imaging and stimulation protocols as well as millisecond repeatability over days of time-lapse imaging.

Olympus

For info: +49-40-23773-5913 | www.microscopy.olympus.eu

SEQUENCING PANEL

The TruSight One Sequencing Panel is the industry's broadest sequencing panel, targeting 4,813 genes with known associated clinical phenotypes. Clinical research laboratories can use this panel to expand existing menus, streamline workflows, or create an entire portfolio of sequencing options, with benefits including increased productivity, reduced handling errors, and decreased costs. The power of the TruSight One Sequencing Panel is enhanced by Illumina's user-friendly VariantStudio analysis and reporting software. To coincide with the introduction of the TruSight One Sequencing Panel, VariantStudio will offer new features that expand annotation and filtering capabilities. These features include support for enabling family-based filtering (mother, father, child, and siblings), providing variant classifications, and generating ready-to-use reports. In addition, the TruSight One Sequencing Panel can be used to create dozens of "virtual subpanels" to fit the needs of any clinical researcher seeking to understand the genetic basis of disease.

Illumina

For info: 800-809-4566 | www.illumina.com/powerofone

DNA METHYLATION ANALYSIS KITS

New, high-efficiency BioArray conversion kits for detection of 5 mC and 5 hmC modifications (Express DNA Methylation Kit, Blood & Tissue DNA Methylation Kit, and BioArray 5-hmC Methylation Kit) allow for resolution of single-nucleotide changes without the introduction of DNA damage. Unique features such as capacity for multiple sample inputs, rapid processing time, and high conversion efficiency rate, provide enhanced flexibility without sacrificing performance. For researchers requiring downstream analysis products, new enzyme-linked immunosorbent assay (ELISA)-based detection kits for 5 mC and 5 hmC exhibit high specificity comparable to mass spectrometry analysis and are amenable to high throughput processing. Complementing the ELISA kits is a new real-time polymerase chain reaction-based detection platform, the BioPanel DNA Methylation Detection Kit for Human Pluripotent Stem Cells. This product is designed to quantify the percentage of DNA methylation in six gene promoters, RAB25, NANOG, PTPN6, MGMT, GBP3, and LYST.

Enzo Life Sciences

For info: 800-942-0430 | www.enzolifesciences.com

ION CHANNEL DRUG SCREENING

The new SyncroPatch 384 Patch Engine (PE) propels ion channel drug discovery to a new level. Designed for seamless integration into process-automated drug screening environments, the Patch Engine is equipped with 384 patch clamp amplifiers and an advanced 384 channel liquid handling robot. The SyncroPatch 384 PE is the high-quality, automated patch clamp system able to finally thrust gold standard electrophysiology from secondary to primary ion channel drug screening. Allowing for up to 20,000 data points per day, it is the most efficient platform on the market for high quality, ion channel recordings. This efficiency is primarily due to fully parallel measurements from 384 cells, the 384-channel pipettor, and exceptionally efficient control and analysis software. Both hardware and software have been fully tested and validated with leading players in industrial ion channel drug development to provide optimum performance in true high throughput ion channel screening.

Nanion Technologies

For info: +49-89-2189-97972 | www.nanion.de

ddPCR LIBRARY QUANTIFICATION KIT

The new droplet digital polymerase chain reaction (ddPCR) library quantification kit is designed for Ion Torrent library preparation. Used with Bio-Rad's QX200 Droplet Digital PCR system, the new kit provides researchers with the ability to precisely and directly measure amplifiable library concentrations. The Ion AmpliSeq library kit is used to prepare libraries for Ion Torrent next generation sequencing systems. Using the ddPCR library quantification kit to quantify Ion AmpliSeq gDNA and RNA libraries maximizes the number of useable reads, enables consistent loading, and optimizes the utilization of every sequencing run. The resulting data provide additional measures of library quality not provided by other methods, including the percentage of nonamplifiable species such as adapter dimers and the size range of library inserts. Additional key benefits of the ddPCR library quantification kit for Ion Torrent systems include superior performance, visualization of library quality, and efficient utilization of sequencing runs.

Bio-Rad Laboratories

For info: 800-424-6723 | www.bio-rad.com/ion-torrent

AAAS 2015 ANNUAL MEETING

12–16 FEBRUARY • SAN JOSE, CA

TEAM SAN JOSE

Call for Symposium Proposals

Symposium proposals for the 2015 AAAS Annual Meeting are now being solicited. To submit a proposal, visit **www.aaas.org/meetings.** The deadline for submission is **25 April 2014.**

Innovations in Information and Imaging

Science and technology are being transformed by new ways to collect and use information. Progress in all fields is increasingly driven by the ability to organize, visualize, and analyze data. Advances in information and imaging technologies are generating novel applications in fields such as biochemistry, computer science, particle physics, genomics, and oceanography, and creating ways to interpret data across disciplines. This transformation makes scientific information more open, available, and accessible globally. The escalating amount of data, and advances in data analysis, are changing the ways we discover answers to scientific and societal problems. Thoughtful consideration of how information is used for societal benefit, evaluated for potential risks, and communicated beyond the scientific community will allow this revolution to reach its full potential.

AAAS

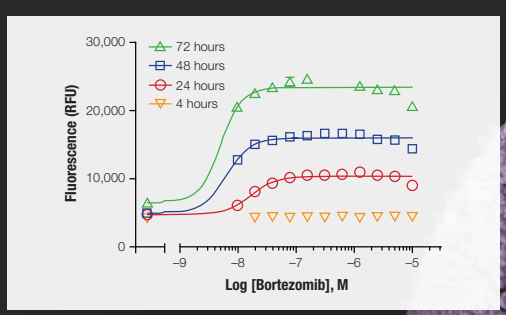
ADVANCING SCIENCE, SERVING SOCIETY

Promega

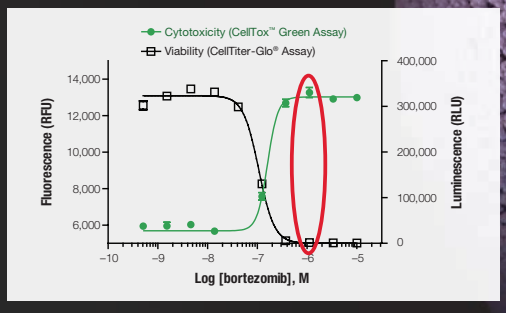
CellTox™ Green

More Biology, Less Work

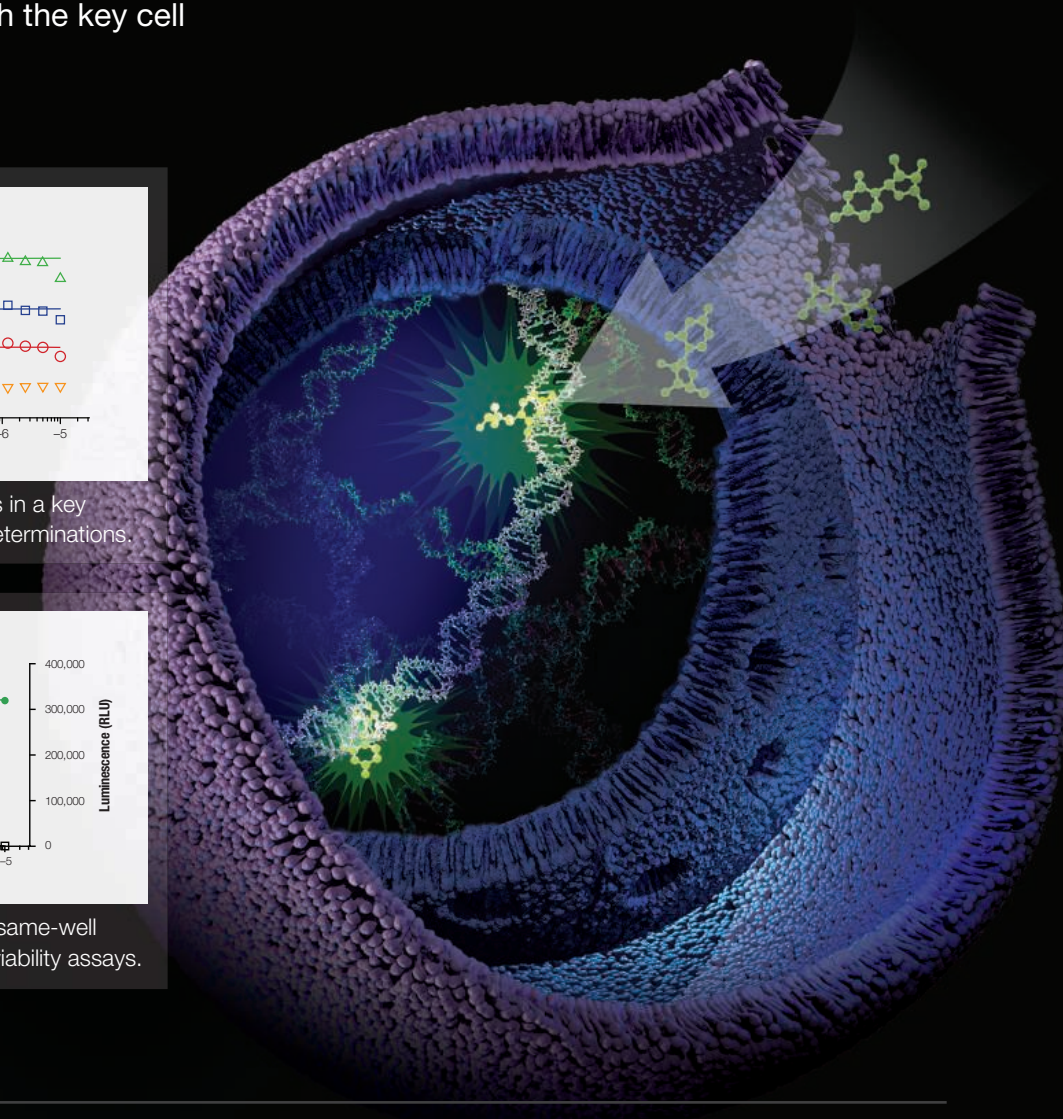
CellTox™ Green enables real-time mechanistic toxicity monitoring with a simple Add & Read protocol. Multiplexing with CellTiter-Glo® allows investigators to monitor temporal changes of membrane-modulated cytotoxicity in parallel with the key cell viability biomarker, ATP.



Easily monitor temporal changes in a key cytotoxicity biomarker for IC_{50} determinations.



Get more informative data from same-well multiplexing of cytotoxicity and viability assays.



To see how easy better biology can be, request a free sample at:
www.promega.com/realtimecytotoxicity

Science Careers Advertising

For full advertising details, go to ScienceCareers.org and click For Employers, or call one of our representatives.

Tracy Holmes
Worldwide Associate Director
Science Careers
Phone: +44 (0) 1223 326525

THE AMERICAS
E-mail: advertise@sciencecareers.org
Fax: 202-289-6742

Tina Burks
Phone: 202-326-6577

Marci Gallun
Phone: 202-326-6582

Online Job Posting Questions
Phone: 202-312-6375

EUROPE / INDIA / AUSTRALIA / NEW ZEALAND / REST OF WORLD
E-mail: ads@science-int.co.uk
Fax: +44 (0) 1223 326532

Axel Gesatzki
Phone: +44 (0)1223 326529

Sarah Lelarge
Phone: +44 (0) 1223 326527

Kelly Grace
Phone: +44 (0) 1223 326528

JAPAN
Yuri Kobayashi
Phone: +81-(0)90-9110-1719
E-mail: ykobayas@aaas.org

CHINA / KOREA / SINGAPORE / TAIWAN / THAILAND
Ruolei Wu
Phone: +86-1367-1015-294
E-mail: rwu@aaas.org

Science Careers
From the journal *Science* AAAS

ScienceCareers.org

2014 *Science*/UCSF Career Event

March 21, 2014 • San Francisco, CA
UCSF Mission Bay Community Center • 10–4:45

A Day of Recruiting Opportunities and Career Workshops

***Science* Careers** and **UCSF** have joined forces to deliver an exciting event that includes five career workshops and a chance to meet face to face with recruiters. Two of the five workshops will focus on career opportunities in Asia.

► **Job seekers:** Visit the Mission Bay campus for a chance to get valuable advice from career experts and to meet with recruiters from some of the top scientific organizations. The combination of valuable career development content and exciting career opportunities makes this a "must-attend" event for scientists in the bay area. For more details and to register, visit **UCSF2014.ScienceCareers.org**

► **Employers:** Save time and money by meeting hundreds of scientists in person. If your organization would like to recruit at this event, please call 202 326-6577 for more information or e-mail **advertise@sciencecareers.org**

UCSF
University of California San Francisco

Science Careers
From the journal *Science* AAAS
ScienceCareers.org

For careers in science, there's only one Science

Research Positions

Careers
Applied Science, Technology, and Engineering Research

Idaho National Laboratory INL

INL's Energy and Environment Science and Technology Directorate is seeking outstanding , highly creative and motivated early to mid-level career professionals to join our multi-disciplinary research teams. The Directorate is INL's principal multi-mission organization focused on research to advance clean energy systems, advanced transportation, advanced process technology and related sciences.

Multiple positions are available in the following areas:
• Bioenergy research including biomass characterization, conversion, preprocessing and molecular biology. • Analytical chemistry specializing in laser spectroscopy and mass spectrometry. • Materials science and physics with focus on performance in harsh environments, nondestructive evaluation, and radiation based imaging. • Membrane science with a focus on dense film separation and filtration. • Geology and Hydrology with a focus on hydraulic fracturing and geothermal evaluation. • Scientific visualization research in a CAVE environment with a focus on immerse visualization, virtual reality, graphics programming, and large format display technologies.

The Idaho National Laboratory is a science-based, applied engineering national laboratory dedicated to supporting the U.S. Department of Energy's mission in nuclear energy research, science, and national defense. With 3,800 scientists, researchers and support staff, the laboratory works with national and international governments, universities and industry partners to discover new science, develop technologies that underpin the nation's nuclear and renewable energy, national security and environmental missions. The Laboratory is a multi-program national laboratory. It currently performs a range of research and development activities associated with energy and national security. The laboratory currently has more than 150,000 sq. ft. of modern research facilities with a significant amount of wet laboratory space that are supported by a vast array of state-of-the-art research instrumentation. Idaho Falls is conveniently situated near many national treasures such as Yellowstone National Park, Teton National Park, Jackson, WY, etc. For more information about the area, please visit **www.visitidahofalls.com** and **www.visitidaho.org**. Interested parties should visit our website at **www.inl.gov**.

INL is an Equal Opportunity Employer M/F/D/V

VCU
VIRGINIA COMMONWEALTH UNIVERSITY

Seeking Applications for Multiple Faculty Positions

School of Engineering

The School of Engineering at Virginia Commonwealth University in Richmond, Virginia, is seeking qualified candidates for tenure-track faculty positions at the Assistant or Associate Professor level for Fall 2014. Founded in 1996, the School stands as a remarkable example of public-private partnership. Due to our expansion plans, we are seeking faculty in all five departments – biomedical engineering, chemical and life science engineering, computer science, electrical and computer engineering, and mechanical and nuclear engineering.

Our School is seeking candidates with research experience and publications that make a positive difference in our community and human kind. The School's strategic plan places particular emphasis on the growth of basic and applied research related to six areas of multi-disciplinary research including: sustainability and energy engineering, micro and nano electronic systems, pharmaceutical engineering, mechanobiology and regenerative medicine, security and mining of big data, and device design and development. Collaborative research and scholarly activity with the medical campus faculty and its associated Schools will be required, as well as engagement with the VCU Hospitals (as appropriate).

The School is projecting an undergraduate enrollment of 2000 and a graduate student enrollment of 500 with a faculty of 115 by 2020. Candidates must have demonstrated experience working in and fostering a diverse faculty, staff, and student environment or commitment to do so as a faculty member at VCU.

For further information concerning these jobs, or to apply, please go to:
http://www.pubinfo.vcu.edu/facjobs/searchunitNew.asp?Item=Engineering

Virginia Commonwealth University is an Equal Opportunity/Affirmative Action Employer. Women, minorities and persons with disabilities are encouraged to apply.

Technical University of Denmark DTU

THE H.C. ØRSTED POSTDOC PROGRAMME

Cofunded by Marie Curie Actions – COFUNDPostdocDTU

Technical University of Denmark - DTU invites highly talented young researchers who have obtained outstanding results during their PhD studies and who have demonstrated excellence and potential in their fields of study, to apply for one of the stipends under the COFUNDPostdocDTU programme. The programme is named after Hans Christian Ørsted, discoverer of electromagnetism and founder of the university, and cofunded by Marie Curie Actions.

The programme will contribute to the researchers' career development, broadening and deepening their individual competencies by exposing them to an international and multidisciplinary research environment.

DTU welcomes applications from all relevant candidates irrespective of age, gender, disability, race, religion or ethnic background.

Applications must be based on the details of the full text announcement which can be downloaded at www.dtu.dk/COFUNDpostdoc

Application deadline: **Sunday 6 April 2014**

DTU is a technical university providing internationally leading research, education, innovation and public service. Our staff of 5.000 advance science and technology to create innovative solutions that meet the demands of society; and our 9.000 students are educated to address the technological challenges of the future. DTU is an independent academic university collaborating globally with business, industry, government, and public agencies.

Further details: career.dtu.dk

SCIENCE AND TECHNOLOGY AT A GLOBAL SCALE - SET THE STANDARDS FOR THE FUTURE
See our PhD-programmes at dtu.dk/phd

University of Colorado
Anschutz Medical Campus

Assistant/Associate Professor
Department of Biochemistry and Molecular Genetics

The Department of Biochemistry and Molecular Genetics at the University of Colorado School of Medicine invites applications for faculty in the area of **structural biology**, with a focus on the use of **single particle cryo-electron microscopy** or **mass spectrometry-based proteomics**. Ideal candidates will integrate these methods with a variety of structural biology and biochemical techniques to address fundamental questions in molecular biology. We are seeking an entry level (Assistant Professor) or mid-career (Associate Professor) colleague.

Successful candidates will be expected to establish an innovative, independent and collaborative research program and participate in teaching. They will join a highly interactive, interdisciplinary group of faculty, students, and fellows, and enjoy access to state-of-the-art equipment and facilities on our new campus. Candidates must hold a PhD (or equivalent) degree and have a strong record of research accomplishment.

Applications are accepted electronically at **www.jobsatcu.com**, refer to job posting **F01090**. Applicants should submit a current CV, a cover letter, a statement of Research Accomplishments and plans (2 page maximum), and the contact information of at least 3 professional references.

We will begin reviewing applications **March 15, 2014**.

The University of Colorado strongly supports the principle of diversity. We encourage applications from women, ethnic minorities, persons with disabilities and all veterans. The University of Colorado is committed to diversity and equality in education and employment.

UNIVERSITY OF CAMBRIDGE

www.jobs.cam.ac.uk

Research Associate

Department of Pharmacology • £28,132-£36,661 (depending on experience)

Multidrug transporters mediate the active extrusion of cytotoxic agents away from their intracellular targets. They are pharmacologically important proteins in microbial pathogens that are associated with some of the most devastating diseases in the world; in this capacity they can impair antimicrobial chemotherapy.

A post-doctoral position is available for up to 2.5 years in the Department of Pharmacology, University of Cambridge, under supervision of Dr Hendrik van Veen. Van Veen's research group aims at the molecular mechanisms of recognition and transport of chemotherapeutic drugs by efflux pumps from pathogenic bacteria and cancer cells.

The project will focus on the molecular mechanisms of tripartite drug efflux pumps, such as AcrAB-TolC and MacAB-TolC in Escherichia coli, and will involve biochemical and electrophysiological techniques in cells and artificial membrane vesicle systems to study substrate transport and its energetics. The project is funded by a programme grant of the Human Frontier Science Program Organization (HFSPO).

Applicants should be highly motivated, enthusiastic individuals, capable of thinking and working independently, and must have experience with membrane proteins. Candidates should have or shortly expect to obtain a PhD in a related subject.

Limited funding: The funds for this post are available for 2.5 years in the first instance.

Further information can be found on our website at: http://www.phar.cam.ac.uk/research/vanveen

To apply online for this vacancy and to view further information about the role, please visit: http://www.jobs.cam.ac.uk/job/3081

This will take you to the role on the University's Job Opportunities pages. There you will need to click on the 'Apply online' button and register an account with the University's Web Recruitment System (if you have not already) and log in before completing the online application form.

Applications, to include a CV, cover sheet (CHRIS/6, Parts I and III only) including the names and addresses of two referees, a brief statement (two sides maximum) of future research plans and a list of publications, should be submitted to the Departmental Administrator, Ms Jessica Dunne, Department of Pharmacology, Tennis Court Road, Cambridge, CB2 1PD, Tel: (01223) 334002, e-mail: recruitment@phar.cam.ac.uk

Please quote reference PL02632 on your application and in any correspondence about this vacancy.

Closing date: Friday 7 March 2014

The University values diversity and is committed to equality of opportunity.

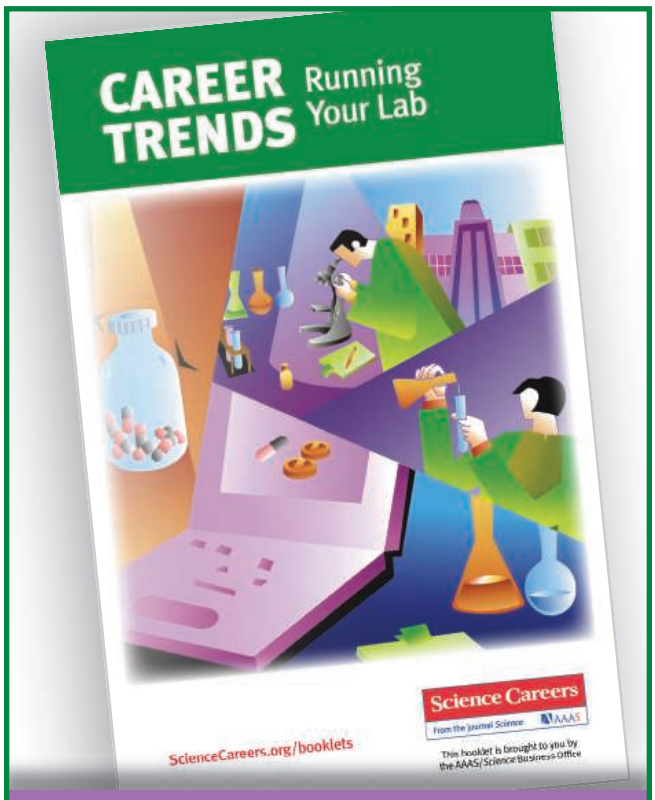
The University has a responsibility to ensure that all employees are eligible to live and work in the UK.

generationnext

MedImmune is committed to investing in the talented, scientific minds of the next generation. We're looking for motivated and innovative post-doctoral scientists who bring a passion for great ideas, fresh thinking to scientific challenges, and a desire to make a difference.

Visit MedImmune.com/careers

MedImmune



Download your free copy today at ScienceCareers.org/booklets

Prescott Medical Communications Group

Medical Writer

The Prescott Medical Communications Group, a health care communications company serving the pharmaceutical and biotechnology industries, has immediate openings for scientific writers/editors. Candidates must possess an advanced biomedical science degree (MS, PhD, PharmD, MD) and excellent oral/written communications skills. Previous experience in CME or agency setting is desirable, as is a background in neuroscience and/or cardiopulmonary medicine. The available positions are full-time in-house, and will require residing in the Chicago area with occasional domestic/international travel. PMCG offers an unparalleled opportunity for professional development in a fast-paced and intellectually challenging environment. Our healthcare package is the best: BCBS Healthcare Savings Accounts, or PPO, in addition to a host of top-of-the-line benefits.

Please send employment history, salary requirements, and three writing samples (e.g., manuscript, slide deck, and scientific poster) to:

Jim Bachleda
Prescott Medical Communications Group
205 N. Michigan Avenue, Suite 3400
Chicago IL 60601
Fax: 312.528.3901
Email: jbachleda@prescottmed.com

Max Planck Institute of Immunobiology and Epigenetics

Max-Planck-Institut für Immunbiologie und Epigenetik

We are offering

Postdoctoral Positions in Immunology

The Max Planck Institute of Immunobiology and Epigenetics in Freiburg, Germany, is offering several Postdoctoral Positions in the Department of Developmental Immunology (Head: Dr. Thomas Boehm). The positions are available for initial two-year appointments with the possibility of extension.

The MPI in Freiburg is an international research institute at the cross-road of Southern Germany, Switzerland and France. The working language is English. State-of-the-art infrastructure and service units including mouse and zebrafish facilities, flow cytometry, imaging, mass spectrometry and proteomics units are available.

Research in the Department of Developmental Immunology covers a broad range of topics. We are interested in the functional characteristics of the alternative adaptive immune system in lampreys (see Nature 470, 90, 2011; Nature 501, 435, 2013), the structure of the immune systems of cartilaginous and bony fishes (see Nature 505, 174, 2014; PNAS 110, 6043, 2013; J. Immunol. 186, 7060, 2011) and the function of the thymus (see Nature 441, 992, 2006; Cell 138, 186, 2009; Cell 149, 159, 2012).

Your qualifications:
We seek enthusiastic, highly motivated and science-driven postdoctoral fellows to join our research activities in these areas. Applications from scientists with proven experience in bioinformatics, cellular immunology and mouse genetics are particularly welcome.

Please ask three referees to send recommendation letters directly to kirk@ie-freiburg.mpg.de

We offer:
Salaries are in accordance with the postdoctoral fellowship guidelines of the Max Planck Society or TVöD and commensurate with experience.

Application deadline: 15.03.2014

Handicapped applicants with equal qualifications will be given preferential treatment. The Max Planck Society seeks to increase the number of women in areas where they are underrepresented, and therefore explicitly encourages women to apply. A childcare facility is directly attached to the Institute.

Max Planck Institute of Immunobiology and Epigenetics
Ms. Stallone

Please apply online via the Jobmarket at our website. We are looking forward to receiving your complete application documents.

http://www1.ie-freiburg.mpg.de/jobs

USDA

United States Department of Agriculture
National Institute of Food and Agriculture

The NATIONAL INSTITUTE OF FOOD AND AGRICULTURE (NIFA) in the DEPARTMENT OF AGRICULTURE is seeking to fill the Senior Executive Service (SES) position of **Deputy Director of the Institute of Food Production and Sustainability (IFPS)**.

The Deputy Director of IFPS is responsible for scientific and managerial leadership and direction in formulating and implementing policies and programs that support research, education, and extension programs leading to new science to sustainably boost U.S. agricultural production, improve global capacity to meet the growing food demand, and foster innovation in fighting hunger by addressing food security. The incumbent serves as a principal scientific and management advisor in administering, evaluating, planning, directing, and coordinating activities related to the mission and function of NIFA, in execution of policies and practices of grants management, and in the allocation of resources to carry out these policies and practices. More information about NIFA can be found at **http://www.nifa.usda.gov/**.

A copy of the job announcement (**AG-22-014-0012**) is available at **https://www.usajobs.gov/**. All applications must be received by **March 12, 2014**.

U.S. CITIZENSHIP REQUIRED. USDA IS AN EQUAL OPPORTUNITY PROVIDER AND EMPLOYER.

CARNEGIE INSTITUTION FOR SCIENCE

PRESIDENT

The Carnegie Institution for Science invites nominations and applications for the position of President.

The Carnegie Institution for Science was founded by Andrew Carnegie in 1902 "to encourage, in the broadest and most liberal manner, investigation, research, and discovery and the application of knowledge to the improvement of mankind..." One of the few organizations that allows scientists the freedom to independently explore new directions, the Carnegie Institution has remained at the forefront of scientific discovery since its founding.

The Carnegie Institution is supported by an endowment of approximately $1 billion, and is currently comprised of six research departments: the Observatories, in Pasadena, California, and Las Campanas, Chile; the Department of Terrestrial Magnetism and the Geophysical Laboratory, both in Washington, D.C.; the Department of Embryology in Baltimore, Maryland; and the Departments of Plant Biology and of Global Ecology in Stanford, California.

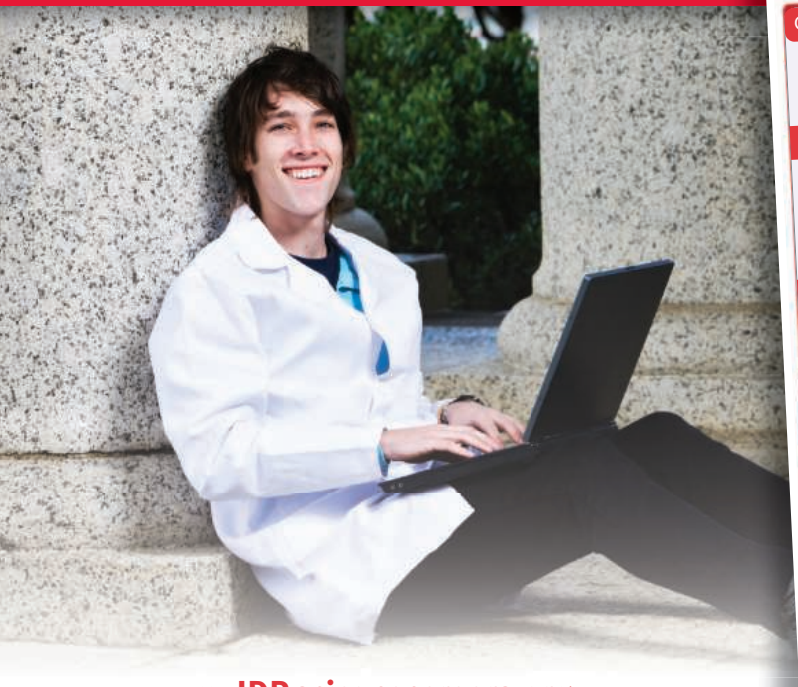
The Carnegie Institution seeks a bold, inspirational leader to serve as its next President. The successful candidate will be an eminent scholar with an outstanding record of peer-reviewed research. She/he will have successfully led a large and complex organization, will have interest and experience in attracting resources for scientific endeavors, and will possess unquestionable integrity. The President will be a catalyst for great science.

All nominations and applications will be treated with the strictest confidence. Interested candidates should submit a C.V. and a summary of accomplishments to:
Carnegie Institution for Science Presidential Search Committee
carnegie@russellreynolds.com

The Carnegie Institution for Science is an equal opportunity / affirmative action employer.

For your career in science, there's only one **Science**

A career plan customized *for* you, *by* you.



myIDP.sciencecareers.org

Recommended by leading professional societies and endorsed by the National Institutes of Health, an individual development plan will help you prepare for a successful and satisfying scientific career.

my IDP Science Careers

In collaboration with FASEB, UCSF, and the Medical College of Wisconsin and with support from the Burroughs Wellcome Fund, AAAS and *Science* Careers present the first and only online app that helps scientists prepare their very own individual development plan.

Visit the website and start planning today!
myIDP.sciencecareers.org

In partnership with:

advancing the life sciences 1912-2012 FASEB Federation of American Societies for Experimental Biology

UCSF University of California San Francisco

Medical College of Wisconsin

Burroughs Wellcome Fund

POSITIONS OPEN

USDA ARS Agricultural Research Service
www.ars.usda.gov

The USDA, Agricultural Research Service, Invasive Insect and Behavior Laboratory in Beltsville Agricultural Research Center, Beltsville, Maryland is seeking a **POSTDOCTORAL RESEARCH ASSOCIATE**, Research Chemist for a Two Year Appointment. Ph.D. is required. Salary is commensurate with experience $63,091 to $82,019 per annum, plus benefits. Citizenship restrictions apply. The incumbent will work in the Crop Protection and Quarantine Program and will conduct research with the long-term goal of developing unique behavioral and biological pest management strategies to manage insect pest and other arthropod populations below economic injury levels using environmentally sound methods. The incumbent will conduct semiochemical identification and chemical synthesis to develop efficient crop protection tools for the brown marmorated stinkbug management. The position requires a recent degree in chemistry as well as knowledge of organic synthesis, chemical ecology, analytical chemistry, entomology, biology, and skill in the use of statistical analysis software. Refer to **website: http://www.afm.ars.usda.gov/divisions/hrd/hrdhomepage/vacancy/pd962.html** for further information on Postdoctoral Research Associate Jobs. For complete application instructions, and the full text announcement, refer to RA13-087-H on **website: https://www.usajobs.gov/GetJob/ViewDetails/359234000**. Send application materials and references to **Dr. Aijun Zhang**, USDA, Agricultural Research Service, **Invasive Insect Biocontrol and Behavior Laboratory, Bldg. 007, Rm. 312, BARC-West, 10300 Baltimore Avenue, Beltsville, M.D. 20705-2350. Telephone: 301-504-5223** (Office), **fax: 301-504-6580, e-mail: aijun.zhang@ars.usda.gov**. *USDA/ARS is an Equal Opportunity Employer and Provider.*