is less than 100%). Models for the additional oligonucleotide, GTP molecules and Mg^{2+} ions, have been fitted into electron density maps and refinement of these oligo–Mn^{2+}–polymerase and oligo–GTP–Mg–Mn–polymerase complexes against their data sets, imposing strict threefold NCS constraints, resulted in models with *R* factors of 23.7 and 21.4%, respectively, and good stereochemistry (Table 1).

Figures

Unless otherwise stated figures were drawn using BOBSCRIPT[26] and rendered with RASTER3D[27].

Received 2 August; accepted 28 December 2000.

1. Reinisch, K. M., Nibert, M. L. & Harrison, S. C. Structure of the reovirus core at 3. 6 Å resolution. *Nature* **404,** 960–967 (2000).
2. Grimes, J. M. *et al.* The atomic structure of the bluetongue virus core. *Nature* **395,** 470–478 (1998).
3. Makeyev, E. V. & Bamford, D. H. Replicase activity of purified recombinant protein P2 of double-stranded RNA bacteriophage ϕ6. *EMBO J.* **19,** 124–133 (2000).
4. Butcher, S. J., Makeyev, E. V., Grimes, J. M., Stuart, D. I. & Bamford, D. H. Crystallization and preliminary X-ray crystallographic studies on the bacteriophage ϕ6 RNA-dependent RNA polymerase. *Acta Crystallogr. D* **56,** 1473–1475 (2000).
5. Mindich, L. Reverse genetics of dsRNA bacteriophage ϕ6. *Adv. Virus Res.* **53,** 341–353 (1999).
6. Gottlieb, P., Strassman, J., Quao, X., Frucht, A. & Mindich, L. In vitro replication, packaging, and transcription of the segmented, double-stranded RNA genome of bacteriophage ϕ6: studies with procapsids assembled from plasmid-encoded proteins. *J. Bacteriol.* **172,** 5774–5782 (1990).
7. Mindich, L. Precise packaging of the three genomic segments of the double-stranded-RNA bacteriophage ϕ6. *Microbiol. Mol. Biol. Rev.* **63,** 149–160 (1999).
8. Makeyev, E. V. & Bamford, D. H. The polymerase subunit of a dsRNA virus plays a central role in the regulation of viral RNA metabolism. *EMBO J.* **19,** 124–133 (2000).
9. Ollis, D. L., Kline, C. & Steitz, T. A. Domain of E. coli DNA polymerase I showing sequence homology to T7 DNA polymerase. *Nature* **313,** 818–819 (1985).
10. Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. An attempt to unify the structure of polymerases. *Protein Eng.* **3,** 461–467 (1990).
11. Lesburg, C. A. *et al.* Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nature Struct. Biol.* **6,** 937–943 (1999).
12. Ago, H. *et al.* Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Struct. Fold. Des.* **7,** 1417–1426 (1999).
13. Bressanelli, S. *et al.* Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc. Natl Acad. Sci. USA* **96,** 13034–13039 (1999).
14. Stuart, D. I., Levine, M., Muirhead, H. & Stammers, D. K. Crystal structure of cat muscle pyruvate kinase at resolution of 2. 6Å. *J. Mol. Biol.* **134,** 109–142 (1979).
15. Oh, J. W., Ito, T. & Lai, M. M. A recombinant hepatitis C virus RNA-dependent RNA polymerase capable of copying the full-length viral RNA. *J. Virol.* **73,** 7694–7702 (1999).
16. Lohmann, V., Overton, H. & Bartenschlager, R. Selective stimulation of hepatitis C virus and pestivirus NS5B RNA polymerase activity by GTP. *J. Biol. Chem.* **274,** 10807–10815 (1999).
17. Frilander, M., Poranen, M. & Bamford, D. H. The large genome segment of dsRNA bacteriophage ϕ6 is the key regulator in the in vitro minus and plus strand synthesis. *RNA* **1,** 510–518 (1995).
18. van Dijk, A. A., Frilander, M. & Bamford, D. H. Differentitation between minus- and plus-strand synthesis: polymerase activity of dsRNA bacteriophage ϕ6 in an in vitro packaging and replication system. *Virology* **211,** 320–323 (1995).
19. Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282,** 1669–1675 (1998).
20. Zhong, W., Uss, A. S., Ferrari, E., Lau, J. Y. & Hong, Z. De novo initiation of RNA synthesis by hepatitis C virus nonstructural protein 5B polymerase. *J. Virol.* **74,** 2017–2022 (2000).
21. Yazaki, K. & Miura, K. Relation of the structure of cytoplasmic polyhedrosis virus and the synthesis of its messenger RNA. *Virology* **105,** 467–479 (1980).
22. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254,** 51–58 (1991).
23. Brunger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
24. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50,** 164–182 (1994).
25. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283–291 (1993).
26. Esnouf, R. M. An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. *J. Mol. Graph.* **15,** 132–134 (1997).
27. Merritt, E. A. & Bacon, D. J. in *Macromolecular Crystallography* (eds Carter, J. W. Jr & Sweet, R. M.) 505–524 (Academic, San Diego, 1997).
28. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11,** 281–296 (1991).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

J. Diprose and G. Sutton helped with synchrotron data collection; J. Diprose and S. Ikemizu with calculations; and R. Esnouf and K. Harlos with computing and in-house data collection. We thank the staff at the beamlines of the ESRF, SRS and APS, in particular Sergey Korolev at the APS for help with the MAD experiment. S.J.B. is a Marie Curie Fellow. J.M.G. is funded by the Royal Society and D.I.S. by the Medical Research Council. The work was supported by the Academy of Finland, the Medical Research Council and the European Union.

Correspondence and requests for materials should be addressed to D.I.S. (e-mail: dave@strubi.ox.ac.uk). Coordinates have been deposited in the RCSB Protein database under accesssion codes: 1HHS, 1HHT, 1HI0, 1HI1, 1HI8.

correction

Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data

Alexandra Ganachaud & Carl Wunsch

Nature **408,** 453–457 (2000).

In this first paragraph of this paper, the uncertainty on the net deep-water production rates in the North Atlantic Ocean was given incorrectly. The correct value should have been $(15 \pm 2) \times 10^6\ m^3\ s^{-1}$. □

errata

Changes in Greenland ice sheet elevation attributed primarily to snow accumulation variability

J. R. McConnell, R. J. Arthern, E. Mosley-Thompson, C. H. Davis, R. C. Bales, R. Thomas, J. F. Burkhard & J. D. Kyne

Nature **406,** 877–879 (2000).

As the result of an editing error, the 1993–1998 aircraft-based altimetry surveys of the southern Greenland ice sheet reported by Krabill *et al.* (1999) were erroneously described as satellite-based. □

Genome sequence of enterohaemorrhagic *Escherichia coli* 0157:H7

Nicole T. Perna, Guy Plunkett III, Valerie Burland, Bob Mau, Jeremy D. Glasner, Debra J. Rose, George F. Mayhew, Peter S. Evans, Jason Gregor, Heather A. Kirkpatrick, György Pósfai, Jeremiah Hackett, Sara Klink, Adam Boutin, Ying Shao, Leslie Miller, Erik J. Grotbeck, N. Wayne Davis, Alex Lim, Eileen T. Dimalanta, Konstantinos D. Potamousis, Jennifer Apodaca, Thomas S. Anantharaman, Jieyi Lin, Galex Yen, David C. Schwartz, Rodney A. Welch & Frederick R. Blattner

Nature **409,** 529–533 (2001).

The Genbank accession number for the annotated sequence given in this paper was typeset incorrectly. The correct accession number is AE005174. □

of 3D periodic mesostructured materials without assuming any structural models. The resolution for the structure is primarily limited by the quality of the HREM images, which depends on the long-range mesoscale ordering. Therefore, although further progress may give better resolution, we expect no future change to the present conclusions about the structures of SBA-1, SBA-6 and SBA-16, because the validity of the solutions does not depend on the resolution. This is a characteristic of our method that makes it different from other approaches. We also suggest that the results presented here provide a quantitative topological description of ordered mesostructured composites, and that such descriptions are essential in understanding the properties and possible applications of the composites. The resolution of periodically ordered, 3D arrangements of bimodal (meso-micro) pores in SBA-1 and SBA-6 makes it possible to consider the detailed characterization of the range of complicated porous phases that are now synthetically achievable. □

Methods

Synthesis of SBA-6

3.75 g of tetraethoxysilane (TEOS) was added with magnetic stirring to a clear solution containing 0.5 g of the gemini surfactant $18B_{4-3-1}$ (N,N,N,N′N′-pentamethyl-N′-[4-(4-octadecyloxyphenoxy)-butyl]-propane-1,3-diammonium dibromide, $C_{18}H_{37}OC_6H_4OC_4H_8N(CH_3)_2C_3H_6N(CH_3)_3Br_2$), 45.4 g of doubly distilled water, and 3.69 g of benzyltrimethylammonium hydroxide at room temperature. Stirring was continued for 20 h after the addition of TEOS at room temperature. The reaction gel mixture was heated for 2 d at 80 °C without stirring. The precipitate was filtered and dried in air at room temperature.

Determination of properties

Ar adsorption and desorption isotherms were measured at 87 K. Pore volumes ($cm^3\,g^{-1}$) for SBA-1, SBA-6 and SBA-16 are 0.6, 0.86 and 0.45, respectively, and the ratios of the pore volume to unit cell are respectively 0.57, 0.65 and 0.47. The surface-area/pore-volume ratio ($2.26 \times 10^9\ m^{-1}$) for SBA-1 is nearly three times that of SBA-6 ($7.93 \times 10^8\ m^{-1}$). The silica wall densities determined with an AccPyc 1300 helium pycnometer are also substantially different for SBA-1 ($2.00\ g\,cm^{-3}$) and SBA-6 ($2.20\ g\,cm^{-3}$).

Received 23 May; accepted 6 October 2000.

1. Zhao, D. *et al.* Triblock copolymer syntheses of mesoporous silica with periodic 50 to 300 Ångstrom pores. *Science* **279,** 548–552 (1998).
2. Zhao, D., Huo, Q., Feng, J., Chmelka, B. F. & Stucky, G. D. Nonionic triblock and star diblock copolymer and oligomeric surfactant syntheses of highly ordered, hydrothermally stable, mesoporous silica structures. *J. Am. Chem. Soc.* **120,** 6024–6036 (1998).
3. Alfredsson, V. & Anderson, M. W. Structure of MCM-48 revealed by transmission electron microscopy. *Chem. Mater.* **8,** 1141–1146 (1996).
4. Monnier, A. *et al.* Cooperative formation of inorganic-organic interfaces in the synthesis of silicate mesostructures. *Science* **261,** 1299–1303 (1993).
5. Schacht, S., Janicke, M. & Schüth, F. Modeling X-ray patterns and TEM images of MCM-41. *Microporous Mesoporous Mater.* **22,** 485–493 (1998).
6. Huo, Q. *et al.* Generalized syntheses of periodic surfactant/inorganic composite materials. *Nature* **368,** 317–321 (1994).
7. Huo, Q. *et al.* Organization of organic molecules with inorganic molecular species into nanocomposite biphase arrays. *Chem. Mater.* **6,** 1176–1191 (1994).
8. Auvray, X. *et al.* X-ray diffraction and freeze-fracture electron microscopy study of the cubic phase in the cetylpyridinium chloride formamide and cetyltrimethylammonium chloride formamide systems. *Langmuir* **9,** 444–448 (1993).
9. Charvolin, J. & Sadoc, J. F. Periodic systems of frustrated fluid films and "micellar" cubic structures in liquid crystals. *J. Phys. France* **49,** 521–526 (1988).
10. Ryoo, R., Kim, J. M. & Ko, C. H. in *Studies in Surface Science and Catalysis* Vol. 117 (eds Bonneviot, L., Béland, F., Danumah, C., Giasson, S. & Kaliaguine, S.) 151–158 (Elsevier, Amsterdam, 1998).
11. Nakanishi, K. Pore structure control of silica gels based on phase separation. *J. Porous Mater.* **4,** 67–112 (1997).
12. Geis, H. Studies on clathrasils. III. Crystal structure of melanophlogite, a natural clathrate compound of silica. *Z. Kristallogr.* **164,** 247–257 (1983).

Supplementary Information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

This work was supported in part by CREST, Japan Science and Technology Corporation (O.T.), by the National Research Laboratory Program of Korea (R.R.), and by the National Science Foundation (G.D.S.) and the Army Research Office (G.D.S.). O.T. thanks S. Andersson for encouragement and support. Y.S. thanks the Japan Society for the Promotion of Science.

Correspondence and requests for materials should be addressed to O.M. (e-mail: terasaki@msp.phys.tohoku.ac.jp) or R.R. (e-mail: r.ryoo@mail.kaist.ac.kr).

Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data

Alexandre Ganachaud* & Carl Wunsch

MIT 54-1517, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

Through its ability to transport large amounts of heat, fresh water and nutrients, the ocean is an essential regulator of climate[1,2]. The pathways and mechanisms of this transport and its stability are critical issues in understanding the present state of climate and the possibilities of future changes. Recently, global high-quality hydrographic data have been gathered in the World Ocean Circulation Experiment (WOCE), to obtain an accurate picture of the present circulation. Here we combine the new data from high-resolution trans-oceanic sections and current meters with climatological wind fields, biogeochemical balances and improved a priori error estimates in an inverse model, to improve estimates of the global circulation and heat fluxes. Our solution resolves globally vertical mixing across surfaces of equal density, with coefficients in the range $(3–12) \times 10^{-4}\ m^2\,s^{-1}$. Net deep-water production rates amount to $(15 \pm 12) \times 10^6\ m^3\,s^{-1}$ in the North Atlantic Ocean and $(21 \pm 6) \times 10^6\ m^3\,s^{-1}$ in the Southern Ocean. Our estimates provide a new reference state for future climate studies with rigorous estimates of the uncertainties.

Obtaining a consistent picture of the oceanic circulation requires adjusting thousands of parameters consistently with a priori error estimates. We present here our best estimate from selected hydrographic data (Fig. 1), which will improve with the appearance of new data. Mass flux is the most basic element of the circulation and Fig. 2 shows the best-estimate coast-to-coast integrated water mass transports for selected density classes. A volume of 15 ± 2 Sv ($1\ \text{sverdrup} = 1 \times 10^6\ m^3\,s^{-1}$) of North Atlantic Deep Water (NADW) is produced in the northern North Atlantic Ocean and moves southward, entraining Antarctic Bottom Water (AABW) from below, and Antarctic Intermediate Water (AAIW) from above. As a result, the NADW is increased to 23 ± 3 Sv as it exits the South Atlantic at 30° S. In the Southern Ocean, a total of 21 ± 6 Sv of bottom water is formed from lower Circumpolar Deep Water (CDW)—which corresponds approximately to the lower NADW density range. Bottom water inflows (NADW + AABW mixture) to the Atlantic, Indian and Pacific oceans are 6 ± 1.3 Sv, 11 ± 4 Sv and 7 ± 2 Sv, respectively. In the Indian and Pacific oceans, most of this water returns southward at deep and intermediate levels. These net values are the sums of large, strongly spatially varying, flows of opposing sign, and thus oversimplify the actual circulation; a detailed description of the circulation within each ocean basin will be published elsewhere[3,4]. Our standard model estimate of the inflow in the South Pacific Ocean is in the lower range of previously published values, but it depends directly upon the weight given to the "PO" phosphate–oxygen combination (see Methods[4,5]) conservation constraints relative to mass conservation[3]. The deep inflow to the North Pacific Ocean is also weaker than previously found[5], as a consequence of our consideration of heat and salt conservation in the northern parts of those basins.

No definition of bottom-water formation can be completely unambiguous because of the entrainment of ambient fluid during the sinking process. In our Southern Ocean definition, the bottom-

* Present address: Laboratoire de Physique des Océans, IFREMER, 29280 Plouzané, France.

water formation rate corresponds to the total amount of water crossing the neutral surface, $\gamma^n = 28.11\,\mathrm{kg\,m^{-3}}$ downwards, and subsequently ventilating the deep Atlantic, Indian and Pacific basins. This water is provided primarily from the density range of lower NADW, so that the two sources of deep water are not additive, as is often assumed in qualitative calculations[6]. Consistent with conventional wisdom, no very large amount of deep water forms in the North Pacific or Indian oceans, although intermediate waters are produced there. Transient tracer measurements suggest AABW formation rates of water denser than $\gamma^n = 28.27\,\mathrm{kg\,m^{-3}}$ at 5 to 8 Sv (ref. 7). Because this density class is almost immediately mixed with CDW above it when exiting the Southern Ocean, it is not separately resolved in our model, and there is no inconsistency. However, there is a contradiction with the suggestion that the current production rate of AABW recently decreased[6], as we find no inconsistency between the deep, contemporary, geostrophic circulation and the interior tracer field which reflects the circulation integrated over hundreds of years. We believe that the origin of the disagreement is in those authors' *ad hoc* assumptions[6] about the originating water masses.

The integrated circulation in the upper layers is more uncertain (Fig. 2, red arrows) owing to the enhanced temporal variability known to be present there[3]. A northward flow of 16 ± 3 Sv of thermocline water from the South Atlantic Ocean balances the NADW southward flow; the strength of the Pacific-Indonesian throughflow is estimated at 16 ± 5 Sv, consistent with the most recent current-meter measurements[8]. This estimate is sensitive to the use of hydrography at very low latitudes where noise susceptibility and seasonal variability are large.

Vertical exchanges (advection, denoted $w^\star$, and mixing, measured by a diffusion coefficient $\kappa^\star$ across the neutral surfaces[9] defining the layers) are associated with the general circulation. Both parameters are now the focus of intense research[10,11] because numerical models are known to be very sensitive to $\kappa^\star$. Direct measurements have shown little mixing in the ocean interior, while enhanced mixing of order $1{,}000\,\mathrm{cm^2\,s^{-1}}$ has been observed in oceanic regions with rough topography[11]. Table 1 summarizes the abyssal dianeutral (near vertical) exchanges by basin and depth range. Let an overbar denote a basin average. Then, in the deep range (≃2,000–3,500 m), dianeutral velocities are $\overline{w^\star} = (0.1\text{–}0.6) \times 10^{-6}\,\mathrm{m\,s^{-1}}$, and the diffusivities are $\overline{\kappa^\star} = (3\text{–}4) \times 10^{-4}\,\mathrm{m^2\,s^{-1}}$ with global averages of $\overline{w^\star} = (0.13 \pm 0.03) \times 10^{-6}\,\mathrm{m\,s^{-1}}$ and $\overline{\kappa^\star} = (3.7 \pm 0.7) \times 10^{-4}\,\mathrm{m^2\,s^{-1}}$ between 30° S and 47° N. In the bottom layers (≃3,800 m to the bottom), both means are larger: $\overline{w^\star} = (0.4 \pm 0.1) \times 10^{-6}\,\mathrm{m\,s^{-1}}$ and $\overline{\kappa^\star} = (9 \pm 2) \times 10^{-4}\,\mathrm{m^2\,s^{-1}}$. Diffusivity values in the deep range are consistent with previous calculations from a crude one-dimensional global balance[10]. The largest values are found in the bottom layers although they are correspondingly more uncertain[3]. The spatial average values that are required by our tracer balance result from all mixing processes, including probable strong mixing generated near topography[10,11]. We interpret those values[10] as being the basin-average value of a process strongly mixing the ocean at highly localized regions.

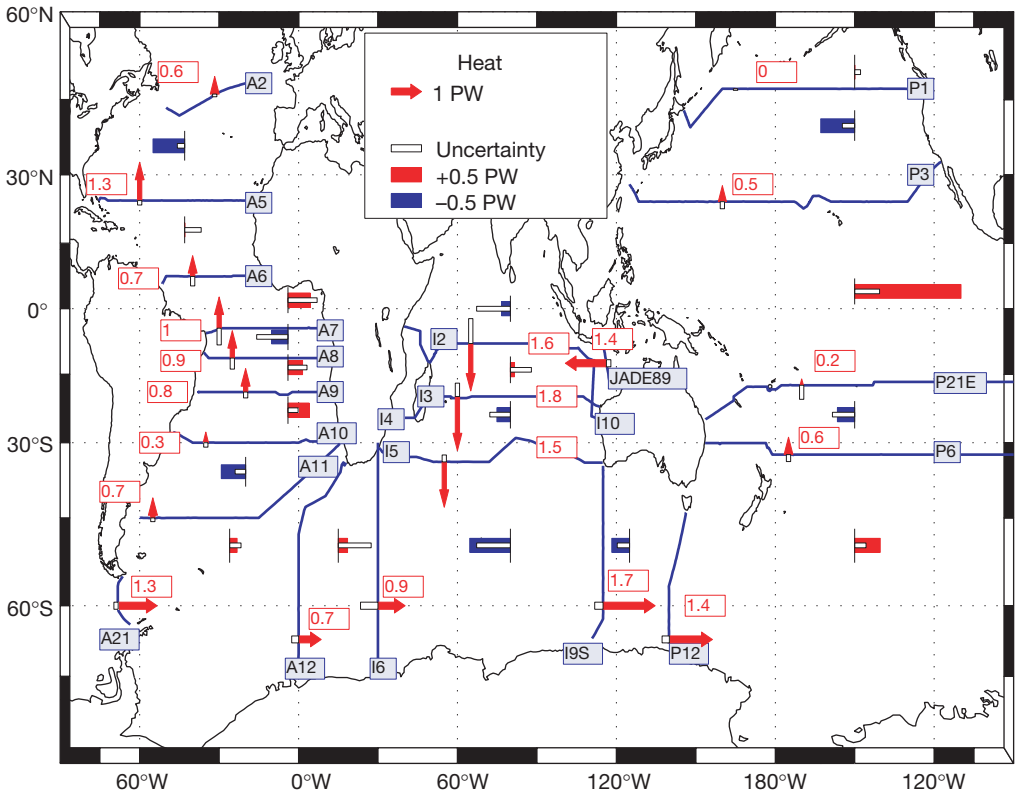


Figure 1 Hydrographic sections and heat fluxes. Transoceanic sections from the WOCE program were selected to ensure a reasonable temporal consistency (1990–1996) and to avoid crossing sections. Between Jakarta and Australia, the 1989 section from the Franco-Indonesian JADE program was used while pre-WOCE sections (1985, 1987)[3] were used in the North Pacific and at 32° S in the Indian Ocean. Each section is a collection of high-density temperature, salinity, oxygen and nutrient measurements. From temperature and salinity, a geostrophic velocity field is calculated and adjusted so that mass and other conservative tracers (see Methods) are conserved between sections. The resulting heat (or 'enthalpy', where the net mass flux is non-zero) transports are indicated by the arrows and red numbers (positive northward/eastward). The white box at the tail end of each arrow is the one-standard-deviation uncertainty. Between sections, ocean–atmosphere heat transfers are indicated by the zonal length of the coloured boxes (blue for ocean cooling; red for ocean heating), with the length of the white box inside indicating the uncertainty. (Because the ocean–atmosphere heat transfers are anomaly residuals, that is, corrected for residual mass imbalances, they do not correspond exactly to the differences between net fluxes across sections, for example, in the North Indian Ocean. But this discrepancy is much less than the uncertainties.)

Diffusivities could not be resolved in the Southern Ocean, where many neutral surfaces outcrop. The improved inverse model method has produced the first near-global, resolved estimates of the dianeutral transfers. The overall results are inconsistent with recent suggestions that the ocean mixes primarily at near-surface outcrops of the neutral surfaces, that is, primarily in the Southern Ocean[12]. Strong abyssal mixing is required by the observed geostrophically balanced circulation, and its absence is incompatible with the observed property distributions.

Figure 1 shows the heat (actually, enthalpy) transports, across each hydrographic section (arrows) along with the residuals reflecting atmospheric heat exchanges (boxes). Residuals are accurately determined at middle and high latitudes, but are more uncertain at lower latitudes (for example, in the Atlantic Ocean) owing to an enhancement of the geostrophic noise there[3]. Nevertheless, the total heating over the tropical Atlantic and Pacific oceans are well-determined, respectively 0.7 ± 0.2 PW ($1\,\mathrm{PW} = 10^{15}$ W) and 1.6 ± 0.4 PW. No significant heat transfers are found in the Indian Ocean because of the large, uncertain, warm water inflow from the Pacific Ocean. This large warm water flux is the main heat escape from the Pacific Ocean, resulting in a northward heat flux in the South Pacific. In the southern Pacific sector, significant heating is found, in contrast with the sparse *in situ* observations[13], but in qualitative agreement with the recent re-analysis of the European Centre for Medium Range Weather Forecasts[14]. Figure 3 shows the globally integrated heat fluxes compared to independent estimates. Most of the cooling occurs in the Northern Hemisphere, at a rate of -1.7 ± 0.2 PW, in balance with the 2.3 ± 0.4 PW heating in the tropical band and the -0.7 ± 0.3 PW cooling in the Southern Ocean.

Changes in the oceanic heat transport can have a large impact on atmospheric temperature gradients[15,16] and thus on climate. Previous estimates of the ocean–atmosphere heat exchanges that are based upon purely ocean surface observations are highly uncertain[17,18]. Analyses from numerical weather prediction centres provide oceanic surface fluxes that are often used as boundary conditions for driving ocean models, but associated uncertainty estimates are not provided. Heuristic calculations suggest uncertainties in their estimates of at least ± 0.6 PW for the meridional oceanic heat transport at most latitudes[19]. The present inversion indicates uncertainties that depend on latitude, with a high accuracy of globally integrated heat transfers (Fig. 3). Similar budgets, to be

Table 1 Basin-averaged dianeutral velocities and diffusivities

	$\overline{w*}$ (10^{-6} m s^{-1})	$\overline{\kappa*}$ (10^{-4} m^2 s^{-1})
Atlantic bottom	0.5 ± 0.2	9 ± 4
Indian bottom	0.6 ± 0.3	12 ± 7
Pacific bottom	0.4 ± 0.1	9 ± 2
Southern bottom	− 0.25 ± 0.1	–
Atlantic deep	0.1 ± 0.05	3 ± 1.5
Indian deep	0.3 ± 0.15	4 ± 2
Pacific deep	0.1 ± 0.03	4 ± 1
Southern deep	0.1 ± 0.1	–

The average is calculated on neutral surfaces from $\gamma^n = 28.1$ kg m^{-3} to the bottom (generally 3,800 decibars (or metres) to the bottom) for the 'bottom layers' and from $\gamma^n = 27.96$ to $\gamma^n = 28.07$ for the 'deep layers' (generally 2,000 m to 3,500 m).

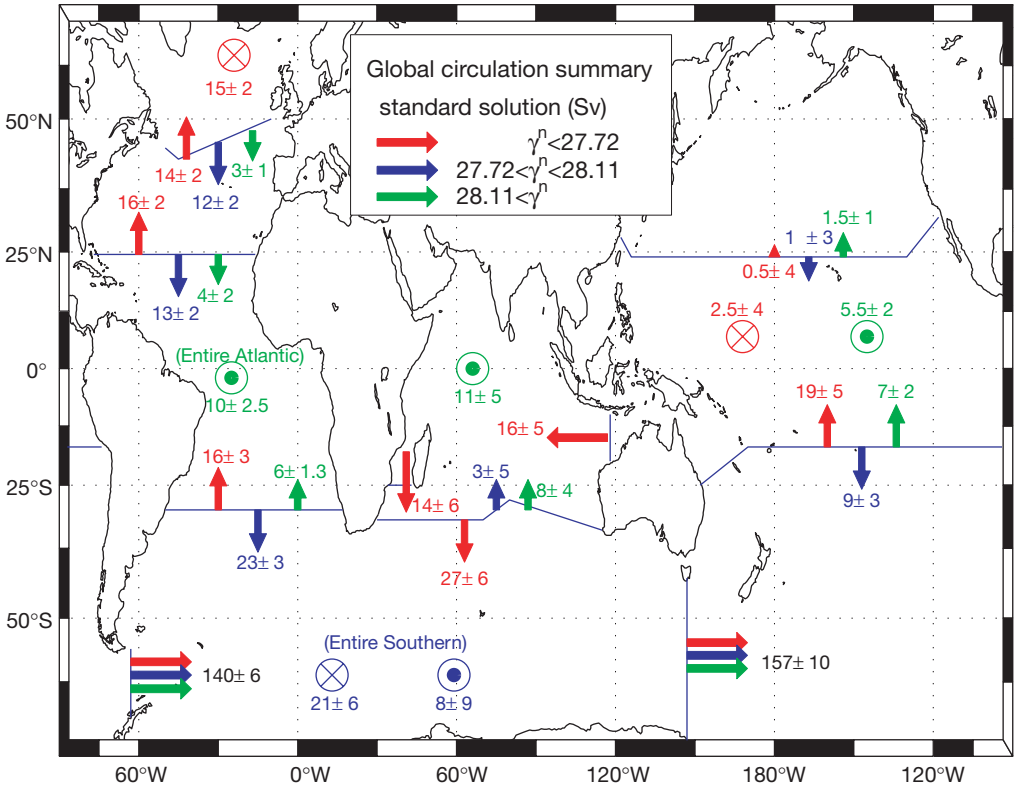


Figure 2 Zonally integrated layer mass transports. The estimated water transports are indicated for the different density classes bounded by neutral surfaces (γ^n, in kg m^{-3}) and across selected hydrographic sections. Neutral surfaces are close to surfaces of constant density, but are chosen so that movement along them minimizes the work done against gravity. The colour of the upwelling or down-welling arrows indicates the layer from which the water is coming. A flux of 0.8 Sv from the Pacific to the Atlantic Ocean through the Bering Strait was taken into account, although it is much smaller than the uncertainties on the net transports. In the Southern Ocean, the bottom water formation takes place mostly in the Weddell Sea, while the upwelling distribution is uncertain. In the Indian Ocean, most of the upwelling takes place north of 7° S. The South Pacific transports are given at 17° S because of the more complicated structure at 32° S (ref. 3). Note the increase in the Southern Ocean transport south of Australia owing to the recirculation of Indonesian throughflow water.

published elsewhere, of oxygen and nutrients, are important for both climate and biogeochemical processes[3,20].

The physical circulation estimated here during WOCE provides a new reference state for climate change studies, and a test—within the estimated uncertainties—of general circulation models of both atmosphere and ocean. The estimated values of $\kappa^\star$ and $w^\star$ represent values that models will need to simulate. The high-resolution WOCE data set, the confinement in time, the use of current-meter data as constraints in boundary currents, and the detailed analysis of the uncertainties associated with our method are great improvements in reliability over previous circulation estimates. Within the error estimates, there is no conflict in either transports or heat fluxes with a previous global inversion that was based on a nearly independent and coarser data set[1,21], and no statistically significant change in integrated mass transports over the past 30 years was found[1,22]. This is an important conclusion in its own right.

Improving the accuracy of the estimates made from the present data sets, if interpreted as climatological averages, will not be easy. Although the addition of new data, for example, existing or future meridional sections, will better the spatial resolution, limitations now lie primarily in the uncertainty introduced by true oceanic variability from the daily to the interannual. Significant improvements in the present numbers will occur only through the use of data sets permitting true temporal averaging of the oceanic circulation. For instance, the present solution has large uncertainties due to undersampling of the highly variable Brazil current and Pacific–Indian throughflow. Additional observations there would greatly improve accuracy. Further refinement of the a priori error estimates by using improved numerical models and other data sets (such as altimetry) is also possible. □

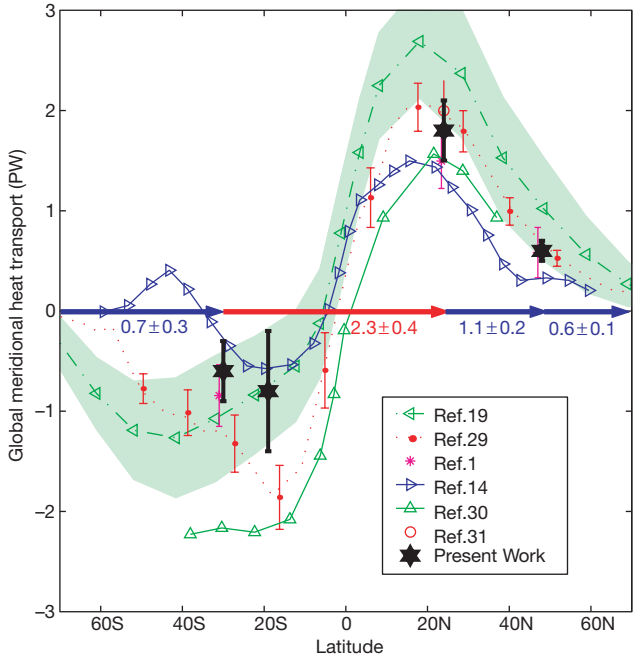


Figure 3 Global meridional heat transports. The black stars with thick error bars indicate the global, zonally integrated heat transports. Ocean–atmosphere heat exchanges between selected latitudes are given by numbers above the northward pointing arrows (positive for oceanic heating). The uncertainties from ref. 19 are indicated by the green band. The previous estimate of ref. 1 is different, but statistically consistent. No incompatibility is found between the estimates by the European Centre for Medium Range Weather Forecasts and US National Center for Environmental Prediction models[14,19,29] and our new values, owing to their large uncertainties[19]. Other climatological estimates[30,31] were published without error estimates, thus precluding any quantitative comparison. The estimate of ref. 30 is taken from their experiment B. (This figure is adapted and completed from refs 1 and 30.)

Methods

The method is that of hydrographic inverse box models[23], where the relative, geostrophic velocity field is obtained from temperature and salinity measurements across the sections of Fig. 1. This initial flow is uniformly adjusted at each location so that the flow satisfies near-conservation of mass, anomalies of salt, heat, and the phosphate-oxygen combination ("PO" $= 170[PO_4] + [O_2]$)[24,25] within oceanic layers defined by neutral surfaces (neutral surfaces are densities chosen so that work against gravity is minimized when following the surface, permitting global use of a single density variable[26]). Silica conservation is also required, top-to-bottom, while heat and PO are conserved only in layers that are not in contact with the surface. The surface layer is directly wind-driven (Ekman transport). Diffusive and advective exchanges between layers are determined by the model. The use of anomaly equations[27] permits, through noise subtraction, determination of diffusivities across each interface. Substantial improvements were made to the method, compared to that used in ref. 1, with, in particular, rigorous estimates of uncertainties through a determination of the a priori model error based on a simulation from the output of a quasi eddy-resolving (1/4°) ocean general circulation model[28].

Received 24 February; accepted 4 September 2000.

1. Macdonald, A. M. & Wunsch, C. An estimate of global ocean circulation and heat fluxes. *Nature* **382,** 436–439 (1996).
2. Houghton, J. T. *et al.* (eds) *Climate Change 1995* (Cambridge Univ. Press, 1996).
3. Ganachaud, A. *Large Scale Oceanic Circulation and Fluxes of Freshwater Heat, Nutrients and Oxygen.* Thesis, Massachusetts Institute of Technology, Woods Hole Oceanographic Institution, (1999).
4. Ganachaud, A., Wunsch, C., Marotzke, J. & Toole, J. The meridional overturning and large-scale circulation of the Indian Ocean. *J. Geophys. Res.* (in the press).
5. Roemmich, D. & McCallister, T. Large scale circulation of the North Pacific Ocean. *Prog. Oceanogr.* **22,** 171–204 (1989).
6. Broecker, W. S., Sutherland, S. & Peng, T.-H. A possible 20th-century slowdown of Southern Ocean deep water formation. *Science* **286,** 1132–1135 (1999).
7. Orsi, A. H., Johnson, G. C. & Bullister, J. L. Circulation, mixing and production of Antarctic Bottom Water. *Prog. Oceanogr.* **43,** 55–109 (1999).
8. Gordon, A. L., Susanto, R. D. & Ffield, A. Throughflow within the Makassar Strait. *Geophys. Res. Lett.* **26,** 3325–3328 (1999).
9. Jackett, D. R. & McDougall, T. J. A neutral density variable for the world's oceans. *J. Phys. Oceanogr.* **27,** 237–263 (1997).
10. Munk, W. & Wunsch, C. The Moon and mixing: abyssal recipes II. *Deep-Sea Res. I* **45,** 1977–2010 (1998).
11. Polzin, K. L., Toole, J. M., Ledwell, G. R. & Schmitt, R. W. Spatial variability of turbulent mixing in the abyssal ocean. *Science* **276,** 93–96 (1997).
12. Toggweiler, J. R. & Samuels, B. On the ocean's large-scale circulation near the limit of no vertical mixing. *J. Phys. Oceanogr.* **28,** 1832–1852 (1998).
13. da Silva, A., Young, C. & Levitus, S. *Atlas of Surface Marine Data* Vol. 1, *Algorithms and Procedures* (NOAA Atlas NESDIS 6, US Dept of Commerce, 1994).
14. Garnier, E., Barnier, B., Siefridt, L. & Béranger, K. Investigating the 15-year air-sea flux climatology from the ECMWF reanalysis project as a surface boundary condition for ocean models. *Int. J. Climatol.* (in the press).
15. Manabe, S. & Stouffer, R. J. Two stable equilibria of a coupled ocean-atmosphere model. *J. Clim.* **1,** 841–866 (1988).
16. Marotzke, J. in *Decadal Climate Variability, Dynamics and Predictability* (eds Anderson, D. L. T. & Willebrand, J.) Vol. I, 44 (NATO ASI/Springer, 1996).
17. Gleckler, P. J. & Weare, B. C. Uncertainties in global ocean surface heat flux climatologies derived from ship observations. *J. Clim.* **10,** 2764–2781 (1997).
18. Josey, S. A., Kent, E. C. & Taylor, P. K. New insights into the ocean heat budget closure problem from analysis of the SOC air-sea flux climatology. *J. Clim.* **9,** 2856–2880 (1999).
19. Keith, D. W. Meridional energy transport: uncertainty in zonal means. *Tellus* **47,** 30–44 (1995).
20. Ganachaud, A. & Wunsch, C. Oceanic nutrient and oxygen fluxes during the World Ocean Circulation Experiment and bounds on export production. *Glob. Biogeochem. Cycles* (submitted).
21. Macdonald, A. The global ocean circulation: a hydrographic estimate and regional analysis. *Prog. Oceanogr.* **41,** 281–382 (1998).
22. Schmitz, W. J. Jr & McCartney, M. S. On the North Atlantic circulation. *Rev. Geophys.* **31,** 29–49 (1993).
23. Wunsch, C. *The Ocean Circulation Inverse Problem* 437 (Cambridge Univ. Press, 1996).
24. Anderson, L. A. & Sarmiento, J. L. Redfield ratios of remineralization determined by nutrient data analysis. *Glob. Biogeochem. Cycles* **8,** 65–85 (1994).
25. Broecker, W. S. "NO", a conservative water-mass tracer. *Earth Planet. Sci. Lett.* **23,** 100–107 (1974).
26. McDougall, T. Neutral surfaces. *J. Phys. Oceanogr.* **17,** 1950–1964 (1987).
27. McDougall, T. J. in *Parameterization of Small-Scale Processes, Hawaiian Winter Workshop Aha Hulikoa* 355–386 (Univ. Hawaii, Manoa, 1991).
28. Stammer, D., Tokmakian, R., Semtner, A. & Wunsch, C. How well does a 1/4° global circulation model simulate large scale oceanic observations? *J. Geophys. Res.* **101,** 25779–25811 (1996).
29. Trenberth, K. E. & Solomon, A. The global heat balance; heat transports in the atmosphere and ocean. *Clim. Dyn.* **10,** 107–134 (1994).
30. de las Heras, M. & Schlitzer, R. On the importance of intermediate water flows for the global ocean overturning. *J. Geophys. Res.* **104,** 15515–15536 (1999).
31. Bryden, H. L., Roemmich, D. H. & Church, J. A. Ocean heat transport across 24°N in the Pacific. *Deep-Sea Res.* **38,** 297–324 (1991).

Acknowledgements

This work was completed while A.G. was a graduate student in the MIT/WHOI Joint Program in Physical Oceanography. We thank J. Toole, B. Warren, J. Marotzke, D. Glover and N. Hogg. B. Arbic, G. McKinley, A. Czaja, A. Macdonald, J. Marshall and M. Fieux also provided helpful comments on the manuscript. We are grateful to the principal investigators who provided the data from the World Ocean Circulation Experiment and the Franco-Indonesian Java-Australia Dynamic Experiment. G. Brown and D. Spiegel helped to design the figures. We were supported by the Jet Propulsion Laboratory and by gifts from Ford, General Motors and Daimler-Chrysler to MIT's Climate Modelling Initiative. This work is a contribution to the World Ocean Circulation Experiment.

Correspondence and requests for materials should be addressed to A.G. (e-mail: ganacho@ifremer.fr).

The use of earthquake rate changes as a stress meter at Kilauea volcano

James Dieterich*, Valérie Cayol†* & Paul Okubo‡

* *US Geological Survey, Menlo Park, California 94025, USA*
† *Université B. Pascal, Clermont Ferrand, France*
‡ *Hawaiian Volcano Observatory, US Geological Survey, Hawaii 96718, USA*

Stress changes in the Earth's crust are generally estimated from model calculations that use near-surface deformation as an observational constraint. But the widespread correlation of changes of earthquake activity with stress[1–5] has led to suggestions that stress changes might be calculated from earthquake occurrence rates obtained from seismicity catalogues. Although this possibility has considerable appeal, because seismicity data are routinely collected and have good spatial and temporal resolution, the method has not yet proven successful, owing to the nonlinearity of earthquake rate changes with respect to both stress and time. Here, however, we present two methods for inverting earthquake rate data to infer stress changes, using a formulation for the stress- and time-dependence of earthquake rates[6]. Application of these methods at Kilauea volcano, in Hawaii, yields good agreement with independent estimates, indicating that earthquake rates can provide a practical remote-sensing stress meter.

The inversions use a formulation for earthquake rate changes[6] derived from laboratory observations of rate- and state-dependent fault strength[6–8], which constrain the earthquake nucleation process to be dependent on both time and stress. Previously, this formulation has been applied to model the spatial and temporal characteristics of earthquake clustering phenomena, including foreshocks and aftershocks[6,7], and to evaluate earthquake probabilities following large earthquakes[9]. The effectiveness of the formulation for forward modelling of earthquake phenomena, and its derivation from observed fault properties, provide the basis for its use to estimate stress changes from earthquake rate data. This approach yields stresses that drive the earthquake process. As such, it is distinct from other seismological methods that yield measures of stress changes resulting from earthquakes.

The formulation of Dieterich[6] for rate of earthquake activity R (in a specified magnitude range) can be written in the condensed form

$$R = \frac{r}{\gamma \dot{S}_r}, \text{ where } \mathrm{d}\gamma = \frac{1}{A\sigma}[\mathrm{d}t - \gamma \mathrm{d}S] \quad (1)$$

where γ is a state variable, t is time, and S is a modified Coulomb stress function defined below. The constant r is the steady-state earthquake rate at the reference stressing rate $\dot{S}_r$. A is a dimensionless fault constitutive parameter with values usually in the range 0.005–0.015 (refs 6–8). The modified Coulomb stress function is defined as

$$S = \tau - [\mu - \alpha]\sigma \quad (2)$$

where τ is the shear stress acting across fault planes that generate earthquakes (positive in the slip direction), σ is the normal stress (less pore fluid pressure), μ is the coefficient of fault friction and α is a constitutive parameter[6,10] with an assigned value in this study of 0.25 (refs 6, 10). In equation (1), the term $A\sigma$ is a constant (that is, changes in σ are negligible relative to total σ). For a stress step, equation (1) yields the characteristic aftershock sequence, which consists of an immediate jump of seismicity rate followed by decay that obeys the Omori t^{-1} aftershock decay law with aftershock duration $t_a = A\sigma/\dot{S}$ (ref. 6).

We use two methods to estimate stress changes from earthquake rate data. The first gives stress as a function of time in a specified volume. From equation (1), the observed rate R is used to directly calculate γ as a function of time (that is, $\gamma(t) = r/R(t)\dot{S}_r$). This requires an estimate of $\dot{S}_r$, which can be obtained from independent

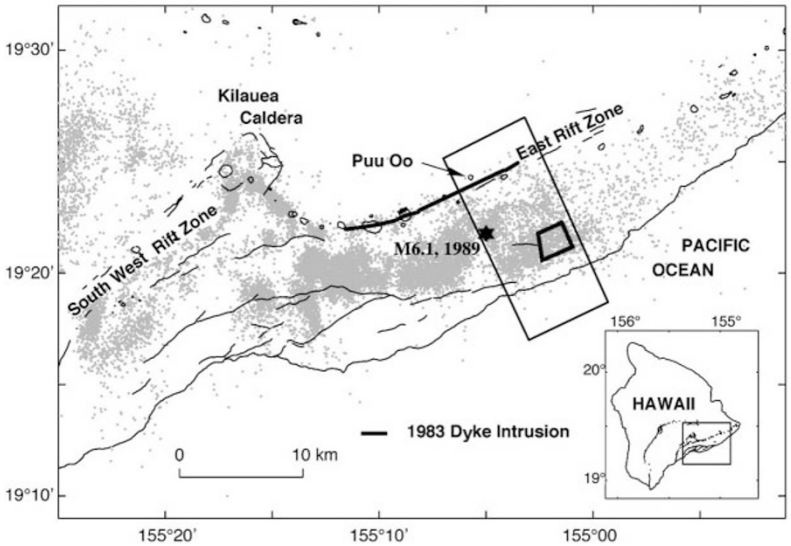


Figure 1 Map of Kilauea volcano showing earthquakes of magnitude $M \approx 1.5$ from 1976 to 1983. Intrusion of magma into the southwest and east rift zones of Kilauea causes rift expansion, and results in motion of the south flank to the SSE. Eruptive fissures initiated the Puu Oo eruption, which started 1 January 1983 and continues to the present. The small polygon is the region of analysis of Fig. 2; the large rectangle gives the region of analysis of Figs 3 and 4.

Changes in Greenland ice sheet elevation attributed primarily to snow accumulation variability

J. R. McConnell*, R. J. Arthern†, E. Mosley-Thompson‡, C. H. Davis§, R. C. Bales‖, R. Thomas¶, J. F. Burkhart‖ & J. D. Kyne#

** Desert Research Institute, University & Community College System of Nevada, Reno, Nevada 89512, USA*
† Applied Physics Laboratory, University of Washington, Seattle, Washington 98105, USA
‡ Byrd Polar Research Centre & Department of Geography, The Ohio State University, Columbus, Ohio 43210, USA
§ Department of Electrical Engineering, University of Missouri-Columbia, Columbia, Missouri 65211, USA
‖ Department of Hydrology & Water Resources, University of Arizona, Tucson, Arizona 85721, USA
¶ EG&G Services, NASA Wallops Flight Facility, Wallops Island, Virginia 23337, USA
 # Snow & Ice Research Group, University of Nebraska, Lincoln, Nebraska 68583, USA

The response of grounded ice sheets to a changing climate critically influences possible future changes in sea level. Recent satellite surveys over southern Greenland show little overall elevation change at higher elevations, but large spatial variability[1–3]. Using satellite studies alone, it is not possible to determine the geophysical processes responsible for the observed elevation changes and to decide if recent rates of change exceed the natural variability. Here we derive changes in ice-sheet elevation in southern Greenland, for the years 1978–88, using a physically based model of firn densification[4] and records of annual snow accumulation reconstructed from 12 ice cores at high elevation. Our patterns of accumulation-driven elevation change agree closely with contemporaneous satellite measurements of ice-sheet elevation change, and we therefore attribute the changes observed in 1978–88 to variability in snow accumulation. Similar analyses of longer ice-core records show that in this decade the Greenland ice sheet exhibited typical variability at high elevations, well within the long-term natural variability. Our results indicate that a better understanding of ice-sheet mass changes will require long-term measurements of both surface elevation and snow accumulation.

The previously reported, 1978–88 elevation-change estimates for the southern Greenland ice sheet[1], derived from analysis of Seasat (6 July to 10 October 1978) and Geosat-Exact Repeat Mission (8 November 1986 to 7 November 1988) satellite radar altimeter data, were improved recently by adding data from the Geosat-Geodetic Mission (1 April 1985 to 30 September 1986). The updated analysis[5] included twice as many elevation-change measurements and expanded the spatial coverage by 50%. In addition, a precise global-ocean reference network created from four years of Topex/Poseidon altimeter data was used to obtain improved corrections for altimeter radial orbit errors and measurement system biases. The updated elevation-change map includes about 90% of the area above the 2,000-m elevation contour south of 72.1° N (maximum latitude of Seasat/Geosat) and about 75% of the area west of the elevation divide between 1,700–2,000 m (Fig. 1). Effectively no coverage is available below 1,700 m because of the poor quality and limited quantity of the radar altimeter data near the ice-sheet edge[5]. The rate of elevation change (dH/dt) varies spatially from −24 to +24 cm yr^{-1} over distances as small as 200 km (note the strong dH/dt gradient around 66° N and the abrupt transition from positive to negative change at the elevation divide).

Between 1997 and 1999, ice cores that penetrated at least to a depth corresponding to 1978 were collected at ten locations around the southern ice sheet. All cores were analysed for a number of seasonally varying chemical species (hydrogen peroxide, calcium, ammonium, nitrate ions) using a continuous ice-core melter analysis system[6]. Two additional seasonally varying constituents, insoluble dust concentration and stable water-isotope ratios, were measured using discrete samples. In combination with field measurements of snow density, these multi-parameter glacio-chemical records were used to determine records of net annual snow accumulation at each site. Because of spatial variability in snow accumulation[7], multiple cores were collected at four locations to identify regional-scale versus local-scale accumulation components[8,9]. Discrete measurements of hydrogen peroxide from a core drilled in 1988 were used to develop a record of annual accumulation at Dye 3 (65.2° N, 43.9° W)[10] and a 1778–1989 accumulation record was developed at Crete (71.1° N, 37.3° W) by combining published accumulation records[11,12].

The density of near-surface firn layers increases with depth and, in response to changing snow accumulation rates and temperature, this increase with depth also varies with time. To derive changing ice-sheet elevation from measured time series of snow accumulation, we used a model that predicts the evolution of the density–depth profile by solving coupled heat- and mass-transfer equations using a one-dimensional, lagrangian scheme, referenced to pressure coordinates[4]. The constitutive relationship for snow was calculated using rate equations for the various sintering mechanisms important in snow compaction[13,14]. The initial depth profiles of temperature and density were obtained from steady-state calculations at the mean temperature, mean accumulation rate and mean surface density. Boundary conditions include the temperature of the ice at depth, and surface histories of temperature[15], accumulation

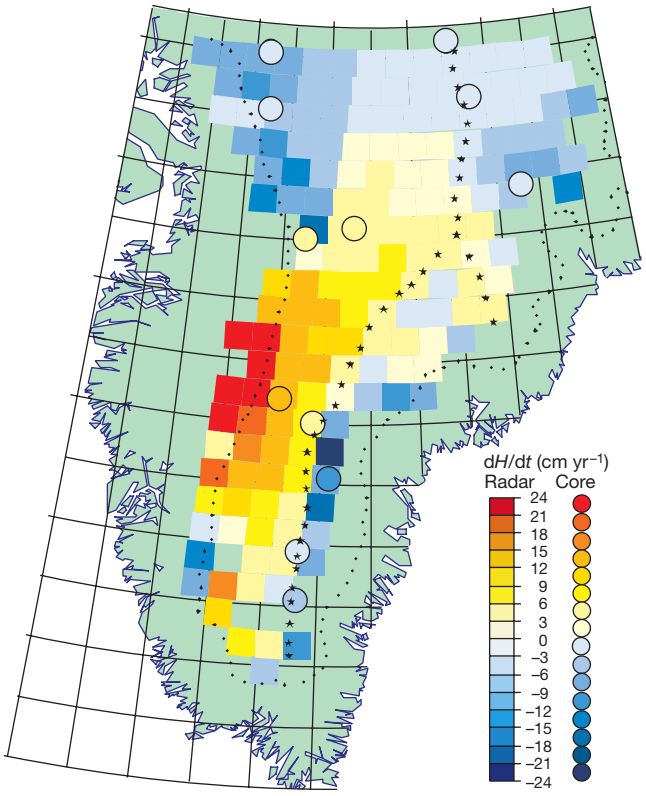


Figure 1 Accumulation-derived and radar-altimetry-derived 1978–88 ice-sheet elevation change. Shown are the approximate locations of the elevation divide (stars), the 2,000-m elevation contour (dots) and the ice-core sites (circles). The radar dH/dt values are shown in 50 × 50 km cells (adapted from ref. 5).

rate and density. Mean surface density was derived from the ice-core data using the zero-depth intercept of a polynomial fit to the measured density–depth profile. Densification rates for the upper two metres were also derived from the measured density profiles rather than from sintering rate equations (at shallow depths the effects of vapour transport, seasonal temperature variations and reworking by wind are important although these were not explicitly included in the sintering model). The annual temperature cycle and surface density (uppermost 1 m) in Greenland change little from year to year. Earlier calculations[4] show that elevation changes resulting from these intra-annual variations are small compared with those resulting from inter-annual variations in snow accumulation.

The modelled change in vertical separation between the surface and the 4-MPa pressure level overburden (deep enough to lie within incompressible ice) defines the elevation change caused by variations in density[4]. The actual elevation change observed at any location is the sum of this component and the change in the local ice-equivalent-thickness (defined as the total mass of ice occupying a unit area of bedrock, divided by the density of pore-free ice). Because the divergence of the ice flow field at each core site is unknown, we cannot directly calculate the change in ice-equivalent thickness over the time period spanned by the ice core records. However, on the assumption that the ice-flow divergence is constant during this period, an estimate for the higher-frequency changes in ice-equivalent thickness can be obtained using the measured accumulation rates and added to the elevation changes predicted by the density model. We refer to the resulting time series as the 'elevation anomaly' at each site. These will be superimposed upon any secular trend in elevation change caused by long-term mass imbalance. Both the observed and accumulation-derived elevation changes at the twelve locations where annual accumulation data are available over the 10-year period are listed in Table 1 and shown in Fig. 1. To facilitate comparison with the radar-derived observations, 1978–88 accumulation-driven elevation change (dH/dt) is expressed in cm yr^{-1}.

We investigated the sensitivity in the calculation of accumulation-derived estimates of ice-sheet elevation change to several sources of error. To examine the sensitivity to errors in the constitutive equation for snow, strain rates in the densification model were artificially increased and decreased by a factor of ten[4]. To determine the sensitivity of our results to fluctuations in accumulation rate before the beginning of the ice-core records, we used an ensemble of six plausible ~50-year accumulation-rate histories having the same mean and inter-annual variance as the observed record in Monte-Carlo simulations. Similarly, we investigated the effects of short-scale (tens of metres) spatial variations of accumulation rate by adding noise to the measured time series and repeating these calculations for an ensemble of such records. The variance of the local noise component was obtained from closely spaced ice cores at four of the ice-core sites[8,9]. Note that kilometre-scale topography may also influence accumulation, although generally at multi-decadal to century, rather than annual, timescales. The 65% confidence intervals in Table 1 indicate the combined contribution from all these sources of error. Sensitivity to the accumulation history before the start of the core decreases rapidly with time under the relatively warm conditions found in Greenland[4]. Many of the core records extended to 1974 and earlier (Table 1) and for these sites, short-scale spatial variability in accumulation during the 1978–88 period provided the largest contribution to the combined uncertainties. At the six sites with shorter accumulation records, approximately half of the total uncertainty was the result of unknown accumulation before the start of the core record. If the previous accumulation were markedly different than during the core-based observation period, this would affect the results at sites with relatively short accumulation records.

The radar dH/dt values reported in Table 1 are the average of all elevation-change measurements within a circular area centred at the ice-core location. A nominal radius of 20 km was used for ten of the core locations, whereas a radius of 30 km was used at two core locations (Summit and Dye 3) where insufficient data were available within the nominal radius. Each set of data was edited by excluding elevation differences greater than two standard deviations from the primary gaussian distribution. The 65% confidence intervals include the combined contribution of random measurement error and estimates of the uncertainties in various corrections applied to the elevation data (isostatic uplift, inter-satellite bias, orbit-error corrections, and environmental corrections)[1,5].

Comparisons between the accumulation-derived and radar-altimeter elevation-change estimates (see Supplementary Information) indicate that the large majority of the measured change in ice-sheet elevation in southern Greenland from 1978 to 1988 is the result of temporal variability in snow accumulation. The accumulation-derived elevation change gives the same general pattern of strong thickening to the west of the elevation divide and thinning to the east. We note the excellent match in the strong radar dH/dt gradient across the ice divide from Dye 2 (66.4° N, 46.2° W) to Dye 3 (65.2° N, 43.9° W). The notable exception is in the western region along the 2,000-m contour between 69° N and 72° N where the accumulation-derived elevation change is generally smaller than the radar-based observations. A significant difference was also found in this area in a comparison between 1978–88 radar dH/dt and 1993–98 laser dH/dt results[16] with no corresponding change in accumulation rate between the two time periods. Previous modelling and analysis of the Seasat and Geosat altimeter waveforms demonstrated that there was a significant temporal change in the geophysical characteristics of the near-surface snow in this particular area of the ice sheet[17]. Indeed, an unusually large number of melt events were found in the 69.0° N, 45.0° W core at depths corresponding to 1977 and 1978, with relatively few melt events in the early to mid-1980s. No such anomalous melt features were observed in the nearby 69.2° N, 43.0° W core where there is good

Table 1 Ice-core accumulation-derived and radar-based estimates of mean dH/dt from 1978 to 1988

Site latitude (° N), longitude (° W)	Initial year in accumulation record	1978—1988 Accumulation dH/dt (cm yr^{-1})	1978—1988 Radar* dH/dt (cm yr^{-1})
72.3, 38.0 (Summit)	1950	–1.1 (±2.9)	–0.1 (±1.5)
71.1, 37.3 (Crete)	1778	–0.7 (±4.2)	–1.9 (±1.5)
71.9, 47.5	1974	–1.9 (±2.1)	–8.0 (±1.5)
71.1, 47.2	1974	–1.1 (±2.0)	–5.1 (±2.1)
69.8, 35.0	1976	–3.8 (±1.5)	–6.2 (±1.5)
69.2, 43.0	1977	+5.8 (±1.6)	+4.5 (±1.7)
69.0, 45.0	1977	+6.9 (±2.3)	–22.3 (±1.7)
66.4, 46.2 (Dye 2)	1900	+12.7 (±2.3)	+14.9 (±1.6)
66.0, 44.5	1976	+4.8 (±1.2)	+4.7 (±1.6)
65.2, 43.9 (Dye 3)	1931	–10.3 (±2.3)	–8.6 (±1.8)
63.8, 45.0	1978	–0.4 (±2.0)	+3.7 (±1.6)
63.1, 44.8	1978	–3.2 (±2.0)	–2.3 (±1.8)

Error limits show the 65% confidence interval.
* Includes subtraction of 0.5 cm yr^{-1} uplift from long-term isostatic rebound.

Table 2 Ice core accumulation-derived estimates of the standard deviation in decadal ice-sheet elevation change (dH/dt) over the specified time ranges

Site name	Latitude (°N), longitude (°W)	Time range	Standard deviation in decadal-scale dH/dt (cm yr^{-1})
Humboldt	78.5, 56.8	1690–1994	2.4
Nasa U	73.8, 49.5	1690–1994	4.1
Summit	72.3, 38.0	1950–1998	2.0
Crete	71.1, 37.3	1778–1982	3.7
Milcent	70.3, 45.0	1177–1966	8.4
Dye2	66.4, 46.2	1900–1998	12.7
Dye3	65.2, 43.9	1931–1987	9.4

1978–88 dH/dt values for Summit, Crete, Dye 2, and Dye 3 are given in Table 1.

agreement between the radar and accumulation dH/dt estimates. These observations indicate that a large melt layer existed in 1977 and 1978 which resulted in a different radar signature at the snow–air interface than in subsequent years. This caused Geosat-Seasat dH/dt estimates to be biased towards excess thinning and/or reduced thickening in that region over the 10-year period.

To put these accumulation-derived changes in ice-sheet elevation in historical perspective, we derived estimates of dH/dt over longer time periods using previously published[6,10,12] and new records of annual accumulation. From the annual dH/dt time series, decadal dH/dt (cm yr^{-1}) values were computed. Standard deviations in decadal dH/dt (Table 2) range from 2.0 cm yr^{-1}at the Summit location in central Greenland to 12.7 cm yr^{-1} at the southern Dye 2 site. Comparisons of the historical range of decadal dH/dt (Table 2) with the 1978–88 accumulation-derived and radar-based measurements (Table 1) indicate that recently observed changes in ice-sheet elevation above ~2,000-m elevation are typical. Because snow accumulation at Dye 2 is highly cyclic, even the marked increase in annual accumulation from 1978 to 1988 that has led to the recent large increase in ice-sheet elevation in that region is not unusual over the last century.

These widely distributed accumulation measurements and computed elevation anomalies demonstrate that altimetry-derived estimates of ice-sheet thickening and thinning from 1978–88 over much of the southern Greenland ice sheet above ~2,000-m elevation are consistent with elevation changes caused by temporal variability in snow accumulation. The ice-core accumulation measurements reported here do not address the significance of observed elevation change at lower elevations[2]. Analyses of longer accumulation records indicate that the decadal-scale changes in ice-sheet elevation that occurred during 1978–88 are typical over the last few centuries and well within the natural variability of accumulation-driven elevation change. It is clear that decadal-scale variability in snow accumulation has the potential to mask longer-period changes in surface elevation associated with ice-sheet dynamics. This suggests that accurate detection of any long-term mass imbalance of the ice sheets and assessment of likely causes will require multi-decadal time series of surface elevation in conjunction with widely distributed ice-core-derived accumulation measurements collected over the time period of interest. □

Received 2 February; accepted 12 July 2000.

1. Davis, C. H., Kluever, C. A. & Haines, B. J. Elevation change of the southern Greenland Ice Sheet. *Science* **279,** 2086–2088 (1998).
2. Krabill, W. *et al.* Rapid thinning of parts of the southern Greenland Ice Sheet. *Science* **283,** 1522–1524 (1999).
3. Zwally, H. J., Brenner, A. C. & DiMarzio, J. P. Comment: Growth of the southern Greenland Ice Sheet. *Science* **281,** 1251 (1998).
4. Arthern, R. J. & Wingham, D. J. The natural fluctuations of firn densification and their effect on the geodetic determination of ice sheet mass balance. *Clim. Change* **40,** 605–624 (1998).
5. Davis, C. H., Kluever, C. A., Haines, B. J., Perez, C. & Yoon, Y. T. Improved elevation change measurement of the southern Greenland Ice Sheet from satellite radar altimetry. *IEEE Trans. Geosci. Remote Sens.* **38,** 1367–1378 (2000).
6. Anklin, M., Bales, R. C., Mosley-Thompson, E. & Steffen, K. Annual accumulation at two sites in northwest Greenland during recent centuries. *J. Geophys. Res.* **103,** 28775–28783 (1998).
7. Van der Veen, C. J. Interpretation of short-term ice-sheet elevation change inferred from satellite altimetry. *Clim. Change* **23,** 383–405 (1993).
8. McConnell, J. R., Mosley-Thompson, E., Bromwich, D. H., Bales, R. C. & Kyne, J. D. Interannual variations in snow accumulation on the Greenland Ice Sheet (1985–1996): New observations versus model predictions. *J. Geophys. Res.* **105,** 4039–4046 (2000).
9. Fisher, D. A., Reeh, N. & Clausen, H. B. Stratigraphic noise in time series derived from ice cores. *Ann. Glaciol.* **7,** 76–83 (1985).
10. Sigg, A. & Neftel, A. Evidence for a 50% increase in H_2O_2 over the past 200 years from a Greenland ice core. *Nature* **351,** 557–559 (1991).
11. Anklin, M., Stauffer, B., Geis, K. & Wagenbach, D. Pattern of annual snow accumulation along the west Greenland flow line: no significant change observed during recent decades. *Tellus B* **46,** 294–303 (1994).
12. Clausen, H. B., Gundestrup, N. S., Johnsen, S. J., Bindschadler, R. & Zwally, J. Glaciological investigations in the Crete area, central Greenland: A search for a new deep-drilling site. *Ann. Glaciol.* **10,** 10–15 (1988).
13. Alley, R. B. Firn densification by grain-boundary sliding: A first model. *J. Phys. (Paris)* **48,** 249–256 (1987).
14. Maeno, N. & Ebinuma, T. Pressure sintering of ice and its implication to the densification of snow at polar glaciers and ice sheets. *J. Phys. Chem.* **87,** 4103–4110 (1983).
15. Ohmura, A. New temperature distribution maps for Greenland. *Z. Gletscherkunde Glazialgeol.* **23,** 1–45 (1987).
16. Thomas, R. *et al.* 20-year time series of Greenland Ice-Sheet thickness change from radar and laser altimetry. *Polar Geogr.* (submitted).
17. Davis, C. H. Temporal change in the extinction coefficient of snow on the Greenland ice sheet from an analysis of Seasat and Geosat altimeter data. *IEEE Trans. Geosci. Remote Sens.* **34,** 1066–1073 (1996).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We acknowledge D. Belle-Oudry, M. Hutterli, B. Snider, E. Kline, Z. Li, P.-N. Lin, V. Zagorodnov, E. Ramsey, the Twin Otter and LC-130 aircraft crews and others for assistance in the laboratory and the field. This work was supported by grants from NASA and the US NSF.

Correspondence and requests for materials should be addressed to J.M. (email: jmcconn@dri.edu).

Evidence for iron, copper and zinc complexation as multinuclear sulphide clusters in oxic rivers

Tim F. Rozan*, Michael E. Lassman†, Douglas P. Ridge† & George W. Luther III*

**College of Marine Studies, University of Delaware, 700 Pilottown Road, Lewes, Delaware 19958, USA*
† *Department of Chemistry and Biochemistry, Lamont duPont Laboratory, University of Delaware, Newark, Delaware 19716, USA*

The availability and toxicity of trace metals in fresh water are known to be regulated by the complexation of free metal ions with dissolved organic matter[1–3]. The potential role of inorganic sulphides in binding trace metals has been largely ignored because of the reduced persistence of sulphides in these oxic waters. However, nanomolar concentrations of copper and zinc sulphides have been observed in four rivers in Connecticut and Maryland[4,5]. Here we report dissolved (< 0.2 μm particle diameter) sulphide concentrations ranging up to 600 nM, with more than 90% being complexed by copper, iron and zinc. These complexes account for up to 20% of the total dissolved Fe and Zn and 45% of the total dissolved Cu. Fourier transform mass spectrometry reveals that these complexes are not simple $M(HS)^+$ protonated species[6,7] but are higher-order unprotonated clusters (M_3S_3, M_4S_6, M_2S_4), similar to those found in laboratory solutions[8–10] and bio-inorganic molecules[11]. These extended structures have high stability constants[8,10] and are resistant to oxidation and dissociation[10,12], which may help control the toxicity of these and other less abundant, but more toxic, trace metals, such as silver, cadmium and mercury.

We found the total dissolved sulphide (DS_T) concentrations and metal sulphide speciation in river waters to be strongly correlated with the extent of watershed development. DS_T, defined as Σ Fe sulphides + Cu sulphides + Zn sulphides + other soft class B metal sulphides (Ag, Cd, Hg, Pb) + HS^- + S_x^{2-} + S^0, was found in concentrations ranging between 20 and 580 nM (Table 1). High DS_T (> 200 nM) was found in rivers that drained urbanized areas that received inputs of treated sewage effluent. In these rivers, the metal sulphide speciation, as determined by electrochemical analysis[4,5,13,14], consisted of metal sulphide complexes in the following order of concentration abundance: Fe sulphides (Σ FeS + $FeSH^+$) ≫ Zn sulphides > Cu sulphides > (polysulphides). Polysulphides ($S_{4,5}^{2-}$), which are partly oxidized sulphide species that are reduced by Cr(II), were only electrochemically measured[15] in low

examine the quaternary complex, *LexA–AP3* and *PI* were expressed on the bait vector, and *GAL4 AD-AG* and/or *SEP3-MIK* were expressed on the prey vector. When two genes were expressed on the same vector, they were both driven by ADH1 promoters. Amino-acid residues 1–167 and 1–171 were used for the truncated AP1-MIK and SEP3-MIK proteins, respectively. Other processes and the colony-lift β-gal assays were performed in accordance with the manufacturer's instructions (Clontech).

Immunoprecipitation

For immunoprecipitation experiments, radiolabelled AP1 or SEP3 were mixed with haemagglutinin (HA)-tagged proteins and precipitated with anti-HA antibody. Precipitated AP1 and SEP3 were separated by SDS–PAGE and detected by radio-imaging analyser, BAS2000 (Fujifilm). Other procedures were done as described[7,12].

Transactivation assay

For yeast, MADS proteins cDNAs were fused in-frame to GAL4 DNA-binding domain on pAS2-1 (Clontech) and transformed into the yeast strain YRG-2 (*UAS::lacZ*, Stratagene). AP1-K2C (residues 125–256) and SEP3-K2C (128–257) were used as truncated MADS proteins. Yeast cells were grown at 22 °C overnight, and the β-gal activity was assayed at 30 °C using *o*-nitrophenyl-β-D-galactopyranoside.

For onion epidermal cells, 35S promoter-driven MADS cDNAs that express native MADS proteins (effector) and *CArG::LUC* (reporter) were co-transfected into onion epidermal cells by using a particle delivery system (Bio-Rad). *CArG::LUC* has seven repeats of MADS protein binding consensus sequence[29], 5′-GGGGTGGCTTTCCTTTTTTGG TAAATTTTGGATCC-3′ (CArG box is underlined), upstream of the 35S minimal promoter (−30). 35S::*Renilla* luciferase (RLUC) was used for the internal control. LUC assays were conducted using Dual-luciferase reporter system (Promega). Other procedures were done as described[30].

Plant material

Arabidopsis Columbia ecotype was used for *Agrobacterium*-mediated vacuum transformation[31]. Plant crossing was carried out by manual cross-pollination. The presence of the transgenes was confirmed by PCR. *AP3::GUS* plants have a 600-base-pair region of the *AP3* promoter[16]. Staining for GUS activity was done as described[16].

Cryo-scanning electron micrograph

We used a Hitachi S-3500N scanning electron microscope equipped with a cryo-stage. For observation and photography, the stage was chilled at −20 °C and the natural scanning electron microscopy (SEM) mode (70 Pa) was used with a 25-kV accelerating voltage.

Received 2 October; accepted 6 November 2000.

1. Coen, E. S. & Meyerowitz, E. M. The war of the whorls: genetic interactions controlling flower development. *Nature* **353,** 31–37 (1991).
2. Bowman, J. L., Smyth, D. R. & Meyerowitz, E. M. Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112,** 1–20 (1991).
3. Mizukami, Y. & Ma, H. Ectopic expression of the floral homeotic gene AGAMOUS in transgenic *Arabidopsis* plants alters floral organ identity. *Cell* **71,** 119-131 (1992).
4. Krizek, B. A. & Meyerowitz, E. M. The *Arabidopsis* homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. *Development* **122,** 11–22 (1996).
5. Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E. & Yanofsky, M. F. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **405,** 200–203 (2000).
6. Mandel, M. A., Gustafson-Brown, C., Savidge, B. & Yanofsky, M. F. Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* **360,** 273–277 (1992).
7. Goto, K. & Meyerowitz, E. M. Function and regulation of the *Arabidopsis* floral homeotic gene PISTILLATA. *Genes Dev.* **8,** 1548–1560 (1994).
8. Jack, T., Brockman, L. L. & Meyerowitz, E. M. The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell* **68,** 683-697 (1992).
9. Yanofsky, M. F. *et al.* The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature* **346,** 35–39 (1990).
10. Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H. & Sommer, H. Genetic control of flower development: homeotic genes in *Antirrhinum majus*. *Science* **250,** 931–936 (1990).
11. Ma, H., Yanofsky, M. F. & Meyerowitz, E. M. *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* **5,** 484–495 (1991).
12. Riechmann, J. L., Krizek, B. A. & Meyerowitz, E. M. Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc. Natl Acad. Sci. USA* **93,** 4793–4798 (1996).
13. Herskowitz, I. A regulatory hierarchy for cell specialization in yeast. *Nature* **342,** 749–757 (1989).
14. Tilly, J. J., Allen, D. W. & Jack, T. The CArG boxes in the promoter of the *Arabidopsis* floral organ identity gene APETALA3 mediate diverse regulatory effects. *Development* **125,** 1647–1657 (1998).
15. Hill, T. A., Day, C. D., Zondlo, S. C., Thackeray, A. G. & Irish, V. F. Discrete spatial and temporal cis-acting elements regulate transcription of the *Arabidopsis* floral homeotic gene APETALA3. *Development* **125,** 1711–1721 (1998).
16. Honma, T. & Goto, K. The *Arabidopsis* floral homeotic gene PISTILLATA is regulated by discrete cis-elements responsive to induction and maintenance signals. *Development* **127,** 2021–2030 (2000).
17. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335,** 563–564 (1988).
18. Mandel, M. A. & Yanofsky, M. F. The *Arabidopsis AGL9* MADS box gene is expressed in young flower primordia. *Sex Plant Reprod.* **11,** 22–28 (1998).
19. Rubinelli, P., Hu, Y. & Ma, H. Identification, sequence analysis and expression studies of novel anther-specific genes of *Arabidopsis thaliana*. *Plant Mol. Biol.* **37,** 607–619 (1998).
20. Fan, H. -Y., Hu, Y., Tudor, M. & Ma, H. Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. *Plant J.* **12,** 999–1010 (1997).
21. Cho, S. *et al.* Analysis of the C-terminal region of *Arabidopsis thaliana* APETALA1 as a transcription activation domain. *Plant Mol. Biol.* **40,** 419–429 (1999).
22. Riechmann, J. L. & Meyerowitz, E. M. MADS domain proteins in plant development. *J. Biol. Chem.* **378,** 1079–1101 (1997).
23. Egea-Cortines, M., Saedler, H. & Sommer, H. Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *EMBO J.* **18,** 5370–5379 (1999).
24. Davies, B., Egea-Cortines, M., de Andrade Silva, E., Saedler, H. & Sommer, H. Multiple interactions amongst floral homeotic MADS box proteins. *EMBO J.* **15,** 4330-4343 (1996).
25. Rounsley, S. D., Ditta, G. S. & Yanofsky, M. F. Diverse roles for MADS box genes in *Arabidopsis* development. *Plant Cell* **7,** 1259–1269 (1995).
26. Smyth, D. A reverse trend—MADS functions revealed. *Trends Plant Sci.* **5,** 315–317 (2000).
27. Parcy, F., Nilsson, O., Busch, M. A., Lee, I. & Weigel, D. A genetic framework for floral patterning. *Nature* **395,** 561–566 (1998).
28. Bartel, P. L., Chien, C., Sternglanz, R. & Fields, S. in *Cellular Interactions in Development: a Practical Approach.* (ed. Hartley, D. A.) 153–179 (IRL Press, Oxford, 1993).
29. Shiraishi, H., Okada, K. & Shimura, Y. Nucleotide sequences recognized by the AGAMOUS MADS domain of *Arabidopsis thaliana* in vitro. *Plant J.* **4,** 385–398 (1993).
30. Pan, S., Sehnke, P. C., Ferl, R. J. & Gurley, W. B. Specific interactions with TBP and TFIIB in vitro suggest that 14-3-3 proteins may participate in the regulation of transcription when part of a DNA binding complex. *Plant Cell* **11,** 1591–1602 (1999).
31. Bechtold, N., Ellis, J. & Pelletier, G. *In planta Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis* plants. *C. R. Acad. Sci. Paris* **316,** 1194–1199 (1993).

Acknowledgements

We are grateful to M. Yanofsky for communicating data before publication, and to D. Weigel for providing the cDNA library. We also thank J. Bowman, T. Ito and H. Tsukaya for critical reading of the manuscript. This work was supported by grants from the Monbusho and JSPS.

Correspondence and requests for materials should be addressed to K.G. (e-mail: kgoto@v004.vaio.ne.jp).

Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7

Nicole T. Perna*†, Guy Plunkett III‡, Valerie Burland‡, Bob Mau‡, Jeremy D. Glasner‡, Debra J. Rose‡, George F. Mayhew‡, Peter S. Evans‡, Jason Gregor‡, Heather A. Kirkpatrick‡, György Pósfai§, Jeremiah Hackett‡, Sara Klink‡, Adam Boutin‡, Ying Shao‡, Leslie Miller‡, Erik J. Grotbeck‡, N. Wayne Davis‡, Alex Lim‖, Eileen T. Dimalanta‖, Konstantinos D. Potamousis‡‖, Jennifer Apodaca‡‖, Thomas S. Anantharaman¶, Jieyi Lin#, Galex Yen*, David C. Schwartz*‡‖, Rodney A. Welch☆ & Frederick R. Blattner*‡

** Genome Center of Wisconsin, † Department of Animal Health and Biomedical Sciences, ‡ Laboratory of Genetics, ‖ Department of Chemistry, ¶ Department of Biostatistics, and ☆ Department of Medical Microbiology and Immunology, University of Wisconsin, Madison, Wisconsin 53706, USA*
§ Institute of Biochemistry, Biological Research Center, H-6701 Szeged, Hungary
 # Cereon Genomics, LLC, 45 Sidney Street, Cambridge, Massachusetts 02139, USA

The bacterium *Escherichia coli* O157:H7 is a worldwide threat to public health and has been implicated in many outbreaks of haemorrhagic colitis, some of which included fatalities caused by haemolytic uraemic syndrome[1,2]. Close to 75,000 cases of O157:H7 infection are now estimated to occur annually in the United States[3]. The severity of disease, the lack of effective treatment and the potential for large-scale outbreaks from contaminated food supplies have propelled intensive research on the pathogenesis and detection of *E. coli* O157:H7 (ref. 4). Here we have sequenced the genome of *E. coli* O157:H7 to identify candidate genes responsible for pathogenesis, to develop better methods of strain detection and to advance our understanding of

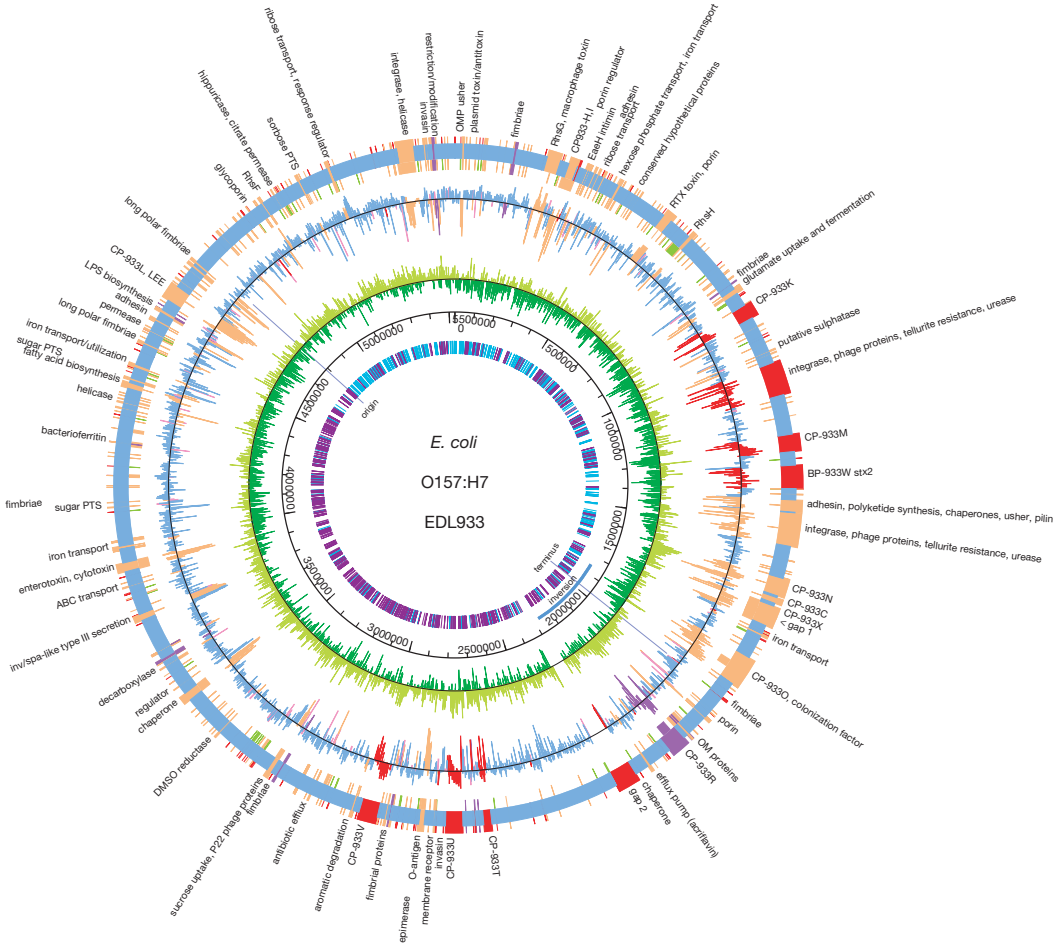


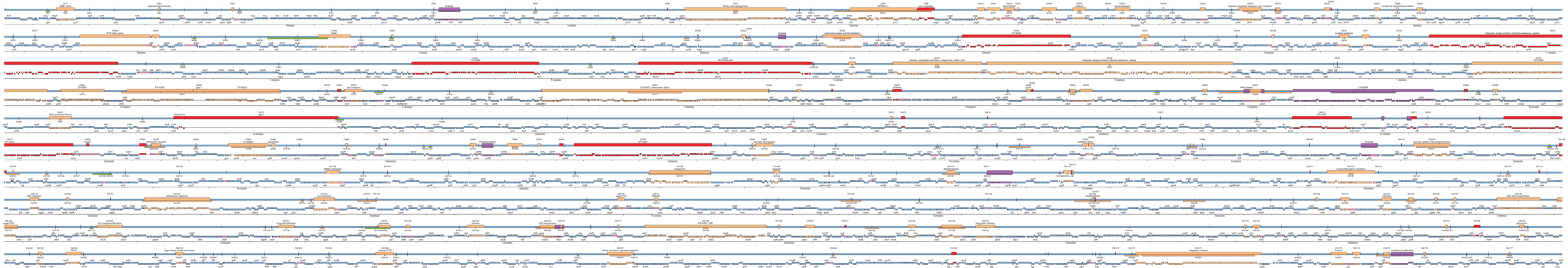
Figure 1 Circular genome map of EDL933 compared with MG1655. Outer circle shows the distribution of islands: shared co-linear backbone (blue); position of EDL933-specific sequences (O-islands) (red); MG1655-specific sequences (K-islands) (green); O-islands and K-islands at the same locations in the backbone (tan); hypervariable (purple). Second circle shows the G+C content calculated for each gene longer than 100 amino acids, plotted around the mean value for the whole genome, colour-coded like outer circle. Third circle shows the GC skew for third-codon position, calculated for each gene longer than 100 amino acids: positive values, lime; negative values, dark green. Fourth circle gives the scale in base pairs. Fifth circle shows the distribution of the highly skewed octamer Chi (GCTGGTGG), where bright blue and purple indicate the two DNA strands. The origin and terminus of replication, the chromosomal inversion and the locations of the sequence gaps are indicated. Figure created by Genvision from DNASTAR.

the evolution of *E. coli*, through comparison with the genome of the non-pathogenic laboratory strain *E. coli* K-12 (ref. 5). We find that lateral gene transfer is far more extensive than previously anticipated. In fact, 1,387 new genes encoded in strain-specific clusters of diverse sizes were found in O157:H7. These include candidate virulence factors, alternative metabolic capacities, several prophages and other new functions—all of which could be targets for surveillance.

Escherichia coli O157:H7 was first associated with human disease after a multi-state outbreak in 1982 involving contaminated hamburgers[1]. The strain EDL933 that we sequenced was isolated from Michigan ground beef linked to this incident, and has been studied as a reference strain for O157:H7. Figures 1 and 2 show the gene content and organization of the EDL933 genome, and compare it with the chromosome of the K-12 laboratory strain MG1655 (ref. 5). These strains last shared a common ancestor about 4.5 million years ago[6]. The two *E. coli* genomes revealed an unexpectedly complex segmented relationship, even in a preliminary examination[7]. They share a common 'backbone' sequence which is co-linear except for one 422-kilobase (kb) inversion spanning the replication terminus (Fig. 1). Homology is punctuated by hundreds of islands of apparently introgressed DNA—numbered and designated 'K-islands' (KI) or 'O-islands' (OI) in Fig. 2, where K-islands are DNA segments present in MG1655 but not in EDL933, and O-islands are unique segments present in EDL933.

The backbone comprises 4.1 megabases (Mb), which are clearly homologous between the two *E. coli* genomes. O-islands total 1.34 Mb of DNA and K-islands total 0.53 Mb. These lineage-specific segments are found throughout both genomes in clusters of up to 88 kb. There are 177 O-islands and 234 K-islands greater than 50 bp in length. Histograms (Fig. 3) show more intermediate and large islands in EDL933 than in MG1655. Only 14.7% (26/177) of the O-islands correspond entirely to intergene regions. The two largest are identical copies of a 106-gene island, both in the same orientation and adjacent to genes encoding identical transfer RNAs.

Figure 2 Detailed comparative map of the EDL933 and MG1655 genomes. The upper double bar in each tier shows the genome comparison in EDL933 coordinates, with segments shown in detail and colour coded as in Fig. 1. Segments shown below the blue bar represent K-islands (MG1655-specific sequence). Segments extending above the blue backbone bar represent O-islands (EDL933-specific sequence). Unique identifying names (KI and OI numbers) were assigned to all segments of more than 50 bp. Unnamed vertical black lines across the blue bar indicate segments of less than 50 bp. In the lower line of each tier, EDL933 genes are presented showing orientation, and are coloured by segment type. Genes spanning segment junctions are shown in pink. Some gene names are given to provide landmarks in the backbone regions, and the sequence gaps are indicated. The scale in base pairs marks the base of each tier. Map created by Genvision from DNASTAR.


Chr1
CRISPR-cas/Type-I-E
Chr2
plasmid maintenance
Chr3
Chr4
fimbriae
Chr5
Chr6
RNA/DNA exchange fuco
Chr7
CPS/ECA1
Chr8
cross-reaction
Chr9
Chr10
Chr11
Chr12
Chr13
EcpR transcription
Chr14
Chr15
adhesin
Chr16
sugar transport
Chr17
Chr18
fimbrial chaperone, tpp transport
Chr19
Chr20
Chr21
Chr22
Chr23
conserved hypothetical proteins
Chr24
Chr25
Chr26
Chr27
RTX toxin, pyuric
Chr28
Chr29
Rhsm
Chr30
Chr31
Chr32
Chr33
Chr34
Chr35
Chr36
Chr37
Chr38
Chr39
Chr40
Chr41
Chr42
Chr43
Chr44
Chr45
Chr46
Chr47
Chr48
Chr49
Chr50
Chr51
Chr52
Chr53
Chr54
Chr55
Chr56
Chr57
Chr58
Chr59
Chr60
Chr61
Chr62
Chr63
Chr64
Chr65
Chr66
Chr67
Chr68
Chr69
Chr70
Chr71
Chr72
Chr73
Chr74
Chr75
Chr76
Chr77
Chr78
Chr79
Chr80
Chr81
Chr82
Chr83
Chr84
Chr85
Chr86
Chr87
Chr88
Chr89
Chr90
Chr91
Chr92
Chr93
Chr94
Chr95
Chr96
Chr97
Chr98
Chr99
Chr100
Chr101
Chr102
Chr103
Chr104
Chr105
Chr106
Chr107
Chr108
Chr109
Chr110
Chr111
Chr112
Chr113
Chr114
Chr115
Chr116
Chr117
Chr118
Chr119
Chr120
Chr121
Chr122
Chr123
Chr124
Chr125
Chr126
Chr127
Chr128
Chr129
Chr130
Chr131
Chr132
Chr133
Chr134
Chr135
Chr136
Chr137
Chr138
Chr139
Chr140
Chr141
Chr142
Chr143
Chr144
Chr145
Chr146
Chr147
Chr148
Chr149
Chr150
Chr151
Chr152
Chr153
Chr154
Chr155
Chr156
Chr157
Chr158
Chr159
Chr160
Chr161
Chr162
Chr163
Chr164
Chr165
Chr166
Chr167
Chr168
Chr169
Chr170
Chr171
Chr172
Chr173
Chr174
Chr175
Chr176
Chr177
T1000000
T2000000
T3000000
T4000000
T5000000
T6000000
T7000000
T8000000
T9000000
T10000000
T11000000
T12000000
T13000000
T14000000
T15000000
T16000000
T17000000
T18000000
T19000000
T20000000
T21000000
T22000000
T23000000
T24000000
T25000000
T26000000
T27000000
T28000000
T29000000
T30000000
T31000000
T32000000
T33000000
T34000000
T35000000
T36000000
T37000000
T38000000
T39000000
T40000000
T41000000
T42000000
T43000000
T44000000
T45000000
T46000000
T47000000
T48000000
T49000000
T50000000
T51000000
T52000000
T53000000
T54000000
T55000000
T56000000
T57000000
T58000000
T59000000
T60000000
T61000000
T62000000
T63000000
T64000000
T65000000
T66000000
T67000000
T68000000
T69000000
T70000000
T71000000
T72000000
T73000000
T74000000
T75000000
T76000000
T77000000
T78000000
T79000000
T80000000
T81000000
T82000000
T83000000
T84000000
T85000000
T86000000
T87000000
T88000000
T89000000
T90000000
T91000000
T92000000
T93000000
T94000000
T95000000
T96000000
T97000000
T98000000
T99000000
T100000000
T101000000
T102000000
T103000000
T104000000
T105000000
T106000000
T107000000
T108000000
T109000000
T110000000
T111000000
T112000000
T113000000
T114000000
T115000000
T116000000
T117000000
T118000000
T119000000
T120000000
T121000000
T122000000
T123000000
T124000000
T125000000
T126000000
T127000000
T128000000
T129000000
T130000000
T131000000
T132000000
T133000000
T134000000
T135000000
T136000000
T137000000
T138000000
T139000000
T140000000
T141000000
T142000000
T143000000
T144000000
T145000000
T146000000
T147000000
T148000000
T149000000
T150000000
T151000000
T152000000
T153000000
T154000000
T155000000
T156000000
T157000000
T158000000
T159000000
T160000000
T161000000
T162000000
T163000000
T164000000
T165000000
T166000000
T167000000
T168000000
T169000000
T170000000
T171000000
T172000000
T173000000
T174000000
T175000000
T176000000
T177000000
T178000000
T179000000
T180000000
T181000000
T182000000
T183000000
T184000000
T185000000
T186000000
T187000000
T188000000
T189000000
T190000000
T191000000
T192000000
T193000000
T194000000
T195000000
T196000000
T197000000
T198000000
T199000000
T200000000
T201000000
T202000000
T203000000
T204000000
T205000000
T206000000
T207000000
T208000000
T209000000
T210000000
T211000000
T212000000
T213000000
T214000000
T215000000
T216000000
T217000000
T218000000
T219000000
T220000000
T221000000
T222000000
T223000000
T224000000
T225000000
T226000000
T227000000
T228000000
T229000000
T230000000
T231000000
T232000000
T233000000
T234000000
T235000000
T236000000
T237000000
T238000000
T239000000
T240000000
T241000000
T242000000
T243000000
T244000000
T245000000
T246000000
T247000000
T248000000
T249000000
T250000000
T251000000
T252000000
T253000000
T254000000
T255000000
T256000000
T257000000
T258000000
T259000000
T260000000
T261000000
T262000000
T263000000
T264000000
T265000000
T266000000
T267000000
T268000000
T269000000
T270000000
T271000000
T272000000
T273000000
T274000000
T275000000
T276000000
T277000000
T278000000
T279000000
T280000000
T281000000
T282000000
T283000000
T284000000
T285000000
T286000000
T287000000
T288000000
T289000000
T290000000
T291000000
T292000000
T293000000
T294000000
T295000000
T296000000
T297000000
T298000000
T299000000
T300000000
T301000000
T302000000
T303000000
T304000000
T305000000
T306000000
T307000000
T308000000
T309000000
T310000000
T311000000
T312000000
T313000000
T314000000
T315000000
T316000000
T317000000
T318000000
T319000000
T320000000
T321000000
T322000000
T323000000
T324000000
T325000000
T326000000
T327000000
T328000000
T329000000
T330000000
T331000000
T332000000
T333000000
T334000000
T335000000
T336000000
T337000000
T338000000
T339000000
T340000000
T341000000
T342000000
T343000000
T344000000
T345000000
T346000000
T347000000
T348000000
T349000000
T350000000
T351000000
T352000000
T353000000
T354000000
T355000000
T356000000
T357000000
T358000000
T359000000
T360000000
T361000000
T362000000
T363000000
T364000000
T365000000
T366000000
T367000000
T368000000
T369000000
T370000000
T371000000
T372000000
T373000000
T374000000
T375000000
T376000000
T377000000
T378000000
T379000000
T380000000
T381000000
T382000000
T383000000
T384000000
T385000000
T386000000
T387000000
T388000000
T389000000
T390000000
T391000000
T392000000
T393000000
T394000000
T395000000
T396000000
T397000000
T398000000
T399000000
T400000000
T401000000
T402000000
T403000000
T404000000
T405000000
T406000000
T407000000
T408000000
T409000000
T410000000
T411000000
T412000000
T413000000
T414000000
T415000000
T416000000
T417000000
T418000000
T419000000
T420000000
T421000000
T422000000
T423000000
T424000000
T425000000
T426000000
T427000000
T428000000
T429000000
T430000000
T431000000
T432000000
T433000000
T434000000
T435000000
T436000000
T437000000
T438000000
T439000000
T440000000
T441000000
T442000000
T443000000
T444000000
T445000000
T446000000
T447000000
T448000000
T449000000
T450000000
T451000000
T452000000
T453000000
T454000000
T455000000
T456000000
T457000000
T458000000
T459000000
T460000000
T461000000
T462000000
T463000000
T464000000
T465000000
T466000000
T467000000
T468000000
T469000000
T470000000
T471000000
T472000000
T473000000
T474000000
T475000000
T476000000
T477000000
T478000000
T479000000
T480000000
T481000000
T482000000
T483000000
T484000000
T485000000
T486000000
T487000000
T488000000
T489000000
T490000000
T491000000
T492000000
T493000000
T494000000
T495000000
T496000000
T497000000
T498000000
T499000000
T500000000
T501000000
T502000000
T503000000
T504000000
T505000000
T506000000
T507000000
T508000000
T509000000
T510000000
T511000000
T512000000
T513000000
T514000000
T515000000
T516000000
T517000000
T518000000
T519000000
T520000000
T521000000
T522000000
T523000000
T524000000
T525000000
T526000000
T527000000
T528000000
T529000000
T530000000
T531000000
T532000000
T533000000
T534000000
T535000000
T536000000
T537000000
T538000000
T539000000
T540000000
T541000000
T542000000
T543000000
T544000000
T545000000
T546000000
T547000000
T548000000
T549000000
T550000000
T551000000
T552000000
T553000000
T554000000
T555000000
T556000000
T557000000
T558000000
T559000000
T560000000
T561000000
T562000000
T563000000
T564000000
T565000000
T566000000
T567000000
T568000000
T569000000
T570000000
T571000000
T572000000
T573000000
T574000000
T575000000
T576000000
T577000000
T578000000
T579000000
T580000000
T581000000
T582000000
T583000000
T584000000
T585000000
T586000000
T587000000
T588000000
T589000000
T590000000
T591000000
T592000000
T593000000
T594000000
T595000000
T596000000
T597000000
T598000000
T599000000
T600000000
T601000000
T602000000
T603000000
T604000000
T605000000
T606000000
T607000000
T608000000
T609000000
T610000000
T611000000
T612000000
T613000000
T614000000
T615000000
T616000000
T617000000
T618000000
T619000000
T620000000
T621000000
T622000000
T623000000
T624000000
T625000000
T626000000
T627000000
T628000000
T629000000
T630000000
T631000000
T632000000
T633000000
T634000000
T635000000
T636000000
T637000000
T638000000
T639000000
T640000000
T641000000
T642000000
T643000000
T644000000
T645000000
T646000000
T647000000
T648000000
T649000000
T650000000
T651000000
T652000000
T653000000
T654000000
T655000000
T656000000
T657000000
T658000000
T659000000
T660000000
T661000000
T662000000
T663000000
T664000000
T665000000
T666000000
T667000000
T668000000
T669000000
T670000000
T671000000
T672000000
T673000000
T674000000
T675000000
T676000000
T677000000
T678000000
T679000000
T680000000
T681000000
T682000000
T683000000
T684000000
T685000000
T686000000
T687000000
T688000000
T689000000
T690000000
T691000000
T692000000
T693000000
T694000000
T695000000
T696000000
T697000000
T698000000
T699000000
T700000000
T701000000
T702000000
T703000000
T704000000
T705000000
T706000000
T707000000
T708000000
T709000000
T710000000
T711000000
T712000000
T713000000
T714000000
T715000000
T716000000
T717000000
T718000000
T719000000
T720000000
T721000000
T722000000
T723000000
T724000000
T725000000
T726000000
T727000000
T728000000
T729000000
T730000000
T731000000
T732000000
T733000000
T734000000
T735000000
T736000000
T737000000
T738000000
T739000000
T740000000
T741000000
T742000000
T743000000
T744000000
T745000000
T746000000
T747000000
T748000000
T749000000
T750000000
T751000000
T752000000
T753000000
T754000000
T755000000
T756000000
T757000000
T758000000
T759000000
T760000000
T761000000
T762000000
T763000000
T764000000
T765000000
T766000000
T767000000
T768000000
T769000000
T770000000
T771000000
T772000000
T773000000
T774000000
T775000000
T776000000
T777000000
T778000000
T779000000
T780000000
T781000000
T782000000
T783000000
T784000000
T785000000
T786000000
T787000000
T788000000
T789000000
T790000000
T791000000
T792000000
T793000000
T794000000
T795000000
T796000000
T797000000
T798000000
T799000000
T800000000
T801000000
T802000000
T803000000
T804000000
T805000000
T806000000
T807000000
T808000000
T809000000
T810000000
T811000000
T812000000
T813000000
T814000000
T815000000
T816000000
T817000000
T818000000
T819000000
T820000000
T821000000
T822000000
T823000000
T824000000
T825000000
T826000000
T827000000
T828000000
T829000000
T830000000
T831000000
T832000000
T833000000
T834000000
T835000000
T836000000
T837000000
T838000000
T839000000
T840000000
T841000000
T842000000
T843000000
T844000000
T845000000
T846000000
T847000000
T848000000
T849000000
T850000000
T851000000
T852000000
T853000000
T854000000
T855000000
T856000000
T857000000
T858000000
T859000000
T860000000
T861000000
T862000000
T863000000
T864000000
T865000000
T866000000
T867000000
T868000000
T869000000
T870000000
T871000000
T872000000
T873000000
T874000000
T875000000
T876000000
T877000000
T878000000
T879000000
T880000000
T881000000
T882000000
T883000000
T884000000
T885000000
T886000000
T887000000
T888000000
T889000000
T890000000
T891000000
T892000000
T893000000
T894000000
T895000000
T896000000
T897000000
T898000000
T899000000
T900000000
T901000000
T902000000
T903000000
T904000000
T905000000
T906000000
T907000000
T908000000
T909000000
T910000000
T911000000
T912000000
T913000000
T914000000
T915000000
T916000000
T917000000
T918000000
T919000000
T920000000
T921000000
T922000000
T923000000
T924000000
T925000000
T926000000
T927000000
T928000000
T929000000
T930000000
T931000000
T932000000
T933000000
T934000000
T935000000
T936000000
T937000000
T938000000
T939000000
T940000000
T941000000
T942000000
T943000000
T944000000
T945000000
T946000000
T947000000
T948000000
T949000000
T950000000
T951000000
T952000000
T953000000
T954000000
T955000000
T956000000
T957000000
T958000000
T959000000
T960000000
T961000000
T962000000
T963000000
T964000000
T965000000
T966000000
T967000000
T968000000
T969000000
T970000000
T971000000
T972000000
T973000000
T974000000
T975000000
T976000000
T977000000
T978000000
T979000000
T980000000
T981000000
T982000000
T983000000
T984000000
T985000000
T986000000
T987000000
T988000000
T989000000
T990000000
T991000000
T992000000
T993000000
T994000000
T995000000
T996000000
T997000000
T998000000
T999000000
T1000000000

Labelling lineage-specific segments 'islands' is an extension of the term 'pathogenicity island' now in common, albeit varied, use. The original term arose from observations that virulence determinants are often clustered in large genomic segments showing hallmarks of horizontal transfer[8]. However, we found K- and O-islands of all sizes with no obvious association with pathogenicity; conversely, genes probably associated with virulence are not limited to the largest islands.

Roughly 26% of the EDL933 genes (1,387/5,416) lie completely within O-islands. In 189 cases, backbone-island junctions are within predicted genes. We classified the EDL933 genes into the functional groups reported for the MG1655 genome[5] and this is included in the annotation. Of the O-island genes, 40% (561) can be assigned a function. Another 338 EDL933 genes marked as unknowns lie within phage-related clusters and are probably remnants of phage genomes. About 33% (59/177) of the O-islands contain only genes of unknown function. Many classifiable proteins are related to known virulence-associated proteins from other *E. coli* strains or related enterobacteria.

Nine large O-islands (>15 kb) encode putative virulence factors: a macrophage toxin and ClpB-like chaperone (OI#7); a RTX-toxin-like exoprotein and transport system (OI#28); two urease gene clusters (OI#43 and #48); an adhesin and polyketide or fatty-acid biosynthesis system (OI#47); a type III secretion system and secreted proteins similar to the *Salmonella–Shigella inv–spa* host-cell invasion genes (OI#115); two toxins and a PagC-like virulence factor (OI#122); a fatty-acid biosynthesis system (OI#138); and the previously described locus of enterocyte effacement (OI#148)[9]. Among the large islands, four include a P4-family integrase and are directly adjacent to tRNAs (OI#43-*serW*, #48-*serX*, #122-*pheV* and #148-*selC*). Only the locus of enterocyte effacement and two of the lambdoid phages (see below) have as yet been experimentally associated with virulence in animal models.

Smaller islands that may be involved in virulence contain fimbrial biosynthesis systems, iron uptake and utilization clusters, and putative non-fimbrial adhesins. Many clusters have no obvious role in virulence, but may confer strain-specific abilities to survive in different niches. Examples include candidates for transporting diverse carbohydrates, antibiotic efflux, aromatic compound degradation, tellurite resistance and glutamate fermentation. Not all islands are expected to be adaptive. Some may represent neutral variation between strains. Still others may be deleterious but either have not yet been eliminated by selection or cannot be eliminated because of linkage constraints.

We identified 18 multigenic regions of the EDL933 chromosome related to known bacteriophages. Only one, the Stx2 Shiga toxin-converting phage BP-933W, is known to be capable of lytic growth and production of infectious particles[10]. We named the other EDL933 prophages cryptic prophage (CP) to indicate that they probably lack a full complement of functional phage genes. They vary in size from 7.5 kb (CP-933L) to 61.6 kb (BP-933W) and consist of a mosaic of segments similar to various bacteriophages, recalling the 'modular' phage genome hypothesis[11]. The two remaining physical gaps in the genome sequence correspond to prophage-related regions, and resolution of the sequence is complicated by extensive similarities to other prophage within this genome. The gap sizes and positions (4 kb and 54 kb) were determined from optical restriction maps. With only one exception, the EDL933 prophages and the eight cryptic prophages of MG1655 are all lineage-specific. Prophage Rac (MG1655) and CP-933R are similarly located in the backbone, and are sufficiently related to suggest a common prophage ancestor at the time that the strains diverged.

Subunits Stx1A and Stx1B of the second Shiga toxin of EDL933 (ref. 12) are encoded in the newly identified CP-933V. The position of the *stx1* genes in a putative Q antiterminator-dependent transcript is analogous to the placement of the *stx2* genes in BP-933W, although there are no tRNAs adjacent to *stx1AB*. Genes in this position should be expressed maximally during lytic growth. The relationship between Stx toxin expression and phage induction is important, because treatment of O157:H7 with macrolide and quinolone antibiotics increase expression of the toxins[13,14]. Clinical decisions regarding drug therapy are complicated by strain-specific variation in this response[15], and reports in the literature (for example, refs 6 and 12) taken together suggest that the Stx phage status is variable among O157:H7 strains. Given the potential for recombination among the prophage reported here, this does not seem surprising. In addition, the *stx* locus in *Shigella* is known to lie within a cryptic prophage, inserted at a site different from either *stx* phage of EDL933 (ref. 16).

The MG1655 genome contains 528 genes (528/4,405 = 12%) not found in EDL933. About 57% (303) of these were classified into known functional groups and include genes, such as for ferric citrate utilization, that would suggest a role in virulence if identified in a pathogen. It is unclear whether these are remnants of a recent pathogenic ancestor, steps along a path leading to evolution of a new pathogen, indicators that K-12 strains may be pathogenic for non-human hosts, or completely unrelated to pathogenicity. There are 106 examples of O-islands and K-islands present at the same locations relative to the conserved chromosomal backbone. The two replichores in each strain are nearly equal in length despite the large number of insertion/deletion events necessary to generate the observed segmented structure between strains. Only a subset of islands is associated with elements likely to be autonomously mobile.

Each island might be ancestral and lost from the reciprocal genome; however, atypical base composition suggests that most islands are horizontal transfers of relatively recent origin from a donor species with a different intrinsic base composition. Restricting analysis to the 108 O-islands greater than 1 kb, 94% (101/108) are significantly different ($\chi^2 > 7.815$, $P < 0.05$) from the average base composition of shared backbone regions in the same replichore. The percentage drops very little with a Bonferroni correction for multiple tests (91/108; $\chi^2 > 17.892$, $P < 0.05$). Similar results are obtained for analysis of the third-codon position composition (Fig. 1). Still more islands may have originated as horizontal transfers but have been resident in genomes with a spectrum of mutation similar enough to *E. coli* to have obtained equilibrium

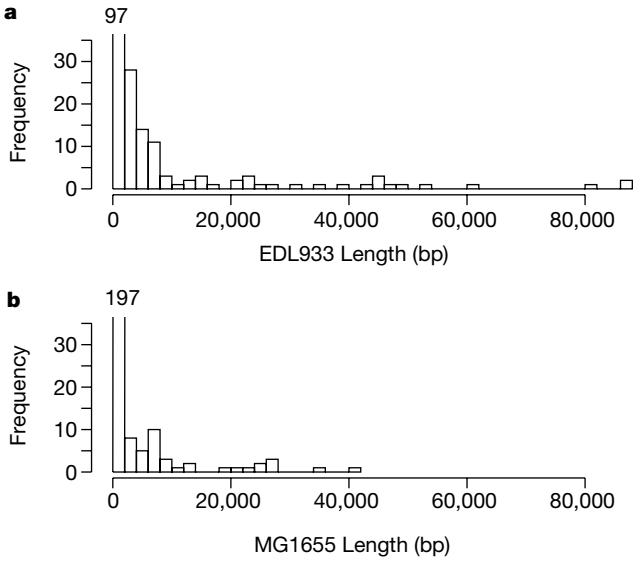


Figure 3 Histograms of lineage-specific segment lengths. **a**, EDL933; **b**, MG1655. The frequencies for the smallest length class are truncated to emphasize the distribution of longer clusters.

nucleotide frequencies or at least obscure statistical significance[17]. Still other gene clusters may be horizontal transfers that predate the divergence of MG1655 and EDL933.

Single nucleotide polymorphisms (75,168 differences) are distributed throughout the homologous backbone. There are 3,574 protein-coding genes encoded in backbone, and the average nucleotide identity for orthologous genes is 98.4%. Many orthologues (3,181/3,574 = 89%) are of equal length in the two genomes, but only 25% (911) encode identical proteins. Table 1 shows the number of each type of polymorphism observed by codon position. As expected, most differences are synonymous changes at third-codon positions. Multiple mutations at the same site should be infrequent at this low level of divergence. Thus the co-occurrence matrix provides insight into the substitution pattern, despite uncertainty of the ancestral state. The overall ratio of transitions to transversions is close to 3:1. A bias towards a greater number of T↔C than A↔G transitions on the coding strand previously attributed to transcription-coupled repair is evident[18]. An additional bias was observed at third-codon positions. Thymidines are more frequently involved in tranversions than cytosines, and G↔T are the most frequent transversions for the coding strand. The reciprocal polymorphisms, C↔A, are not over-represented. This bias is consistent for genes on both the leading and lagging strands (data not shown) and is therefore not related to asymmetries in the replication process. One possible explanation is transcription-coupled repair of damage associated with oxidative stress. Oxidized products of guanine (2,2,4-triaminooxazolone and 7,8-dihydro-8-oxoguanine) lead to G→T transversions by mispairing with A, and two DNA glycosylases (MutY and Fpg) are responsible for mismatch resolution[19]. Preferential repair of these lesions on the transcribed strand has been observed in humans[20], and a similar mechanism could account for the observed transversion bias on the coding strand in *E. coli*.

Some chromosomal regions are more divergent ('hypervariable') than the average homologous segment but encode a comparable set of proteins at the same relative chromosomal position. In the most extreme case (YadC), the MG1655 and EDL933 proteins exhibit only 34% identity. Four such loci encode known or putative fimbrial biosynthesis operons. Another encodes a restriction/modification system. Elevated divergence has been associated with positive selection at both these types of loci and among proteins that interact directly with the host[9,21,22]. Alternatively, hypervariable genes may result from locally elevated mutation rates or differential paralogue retention from an ancient tandem duplication.

Comparison of our observations with other genome-scale analyses of closely related strains or species supports the idea that enterobacterial genomes are particularly subject to recombinational evolution. Two *Helicobacter pylori* strains exhibited only 6–7% differential coding capacity despite showing less identity among orthologues (92.6%) than observed among these *E. coli*. Furthermore, almost half of the lineage-specific *Helicobacter* genes are clustered in a single region referred to as the plasticity zone[23]. Analyses of four *Chlamydia* genomes with orthologues that differ by as much as 19.5% show little evidence of horizontal transfer, and this is attributed to the inherent isolation of an obligate intracellular parasite[24]. Most lineage-specific genes are expansions of paralogous gene families. As in *Helicobacter*, many of the *Chlamydia* lineage-specific elements are clustered in a plasticity zone. Continuing genome projects will elucidate the generality of observations made from these comparisons of closely related organisms.

Together, our findings reveal a surprising level of diversity between two members of the species *E. coli*. Most differences in overall gene content are attributable to horizontal transfer, and offer a wealth of candidate genes that may be involved in pathogenesis. Base substitution has introduced variation into most gene products even among conserved regions of the two strains. Many of these differences can be exploited for development of highly sensitive diagnostic tools; but diagnostic utility will require a clearer understanding of the distribution of genetic elements in *E. coli* species as a whole. An independently isolated O157:H7 strain showed differences from EDL933 by restriction mapping[25]. Additional genome sequence data from other *E. coli* strains as well as functional characterization of gene products is necessary before the complex relationship between *E. coli* genotypes and phenotypes can be understood. Showing that disease-related traits are associated with predicted genes will require many areas of study including extensive testing in animal models that mimic symptoms of human infections, but the genome sequence offers a unique resource to help meet the challenge. □

Table 1 Frequency of each type of single nucleotide polymorphism by codon position for 3,181 genes of equal length in EDL933 and MG1655

First-codon position

	Base in MG1655				EDL933 totals
Base in EDL933	G	A	T	C	
G	–	865	154	137	1,156
A	924	–	129	286	1,339
T	170	143	–	1,260	1,573
C	124	284	1,156	–	1,564
MG1655 totals	1,218	1,292	1,439	1,683	5,632

Second-codon position

	Base in MG1655				EDL933 totals
Base in EDL933	G	A	T	C	
G	–	410	66	118	594
A	393	–	159	166	718
T	67	147	–	464	678
C	107	176	394	–	677
MG1655 totals	567	733	619	748	2,667

Third-codon position

	Base in MG1655				EDL933 totals
Base in EDL933	G	A	T	C	
G	–	6,021	1,562	1,024	8,607
A	6,107	–	1,242	1,124	8,473
T	1,619	1,228	–	8,538	11,385
C	1,049	1,010	8,307	–	10,366
MG1655 totals	8,775	8,259	11,111	10,686	38,831

Methods

Clones and sequencing

EDL933 was kindly provided by C. Kaspar, who obtained it from the American Type Culture Collection (ATCC 43895). The sequenced isolate has been redeposited at the ATCC and is available as ATCC 700927. Whole-genome libraries in M13Janus and pBluescript were prepared from genomic DNA as described for genome segments used in the K-12 genome project[26]. Random clones were sequenced using dye-terminator chemistry and data were collected on ABI377 and 3700 automated sequencers. Sequence data were assembled by Seqman II (DNASTAR). Finishing used sequencing of opposite ends of linking clones, several PCR-based techniques and primer walking. Whole-genome optical maps for restriction enzymes *Nhe*I and *Xho*I were prepared[27] so that the ordering of contigs during assembly could be confirmed. Two gaps remain in the genome sequence. Extended exact matches pose a significant assembly challenge. The final determination of sequence for the 100-kb duplicated region was based on clones that span the junction between unique flanking sequences and the ends of the duplicated island, concordance of the two regions in optical restriction maps, excess random sequence coverage in the duplicated region, lack of polymorphism and confirmation of duplication of an internal segment by Southern blotting (data not shown).

Sequence features and database searches

Potential open reading frames (ORFs) were defined by GeneMark.hmm[28]. The GenPept118 protein and MG1655 protein and DNA databases were searched by each ORF using BLAST[29]. Annotations were created from the search output in which each gene was inspected, assigned a unique identifier, and its product classified by functional group[5]. Alternative start sites were chosen to conform to the annotated MG1655 sequence. Orthology was inferred when matches for EDL933 genes in the MG1655 database exceeded 90% nucleotide identity, alignments included at least 90% of both genes, and the MG1655 gene did not have an equivalent match elsewhere in the EDL933 genome. This list

was supplemented by manual inspection of the protein-level matches in the complete GenPept database to include genes with lower similarities if they occurred within co-linear regions of the genomes. The genome sequence was compared with that of MG1655 by the maximal exact match (MEM) alignment utility, (B.M., manuscript in preparation) an adaptation of MUMmer[30]. This program was based on suffix arrays rather than suffix trees, and exact rather than unique matches, coupled with a custom anchored-alignment algorithm that extends sequence homology into the regions separating contiguous co-linear exact matches. Inferences on biases in polymorphism patterns are based on χ^2 goodness-of-fit tests of a nested sequence of multinomial log–linear models. These predict symmetric elevated levels of A↔G, T↔C and G↔T polymorphisms, above a quasi-independent baseline generated from marginal frequencies in the co-occurrence matrix of synonymous third-codon differences. Further information may be found at our Website http://www.genome.wisc.edu/, including a Genome Browser displaying a comparative map of EDL933 and K-12.

Received 24 July; accepted 6 November 2000.

1. Riley, L. W. *et al.* Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N. Engl. J. Med.* **308,** 681–685 (1983).
2. Karmali, M. A., Steele, B. T., Petric, M. & Lim, C. Sporadic cases of haemolytic–uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. *Lancet* **ii,** 619–620 (1983).
3. Mead, P. S. *et al.* Food-related illness and death in the United States. *Emerg. Infect. Dis.* **5,** 607–625 (1999).
4. Su, C. & Brandt, L. J. *Escherichia coli* O157:H7 infection in humans. *Ann. Intern. Med.* **123,** 698–714 (1995).
5. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453–1474 (1997).
6. Reid, S. D., Herbelin, C. J., Bumbaugh, A. C., Selander, R. K. & Whittam, T. S. Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* **406,** 64–67 (2000).
7. Blattner, F. R. *et al.* Comparative genome sequencing of *E. coli* O157:H7 versus *E. coli* K 12. *Microb. Compar. Genom.* **2,** 174 (1997).
8. Hacker, J. *et al.* Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal *Escherichia coli* isolates. *Microb. Pathog.* **8,** 213-225 (1990).
9. Perna, N. T. *et al.* Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect. Immun.* **66,** 3810–3817 (1998).
10. Plunkett, G. III, Rose, D. J., Durfee, T. J. & Blattner, F. R. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. *J. Bacteriol.* **181,** 1767–1778 (1999).
11. Campbell, A. & Botstein, D. in *Lambda II* (eds Hendrix, R. W., Roberts, J. W., Stahl, F. W. & Weisberg, R. A.) 365–380 (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983).
12. O'Brien, A. D. *et al.* Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* **226,** 694–696 (1984).
13. Walterspiel, J. N., Ashkenazi, S., Morrow, A. L. & Cleary, T. G. Effect of subinhibitory concentrations of antibiotics on extracellular Shiga-like toxin I. *Infection* **20,** 25–29 (1992).
14. Neely, M. N. & Friedman, D. I. Functional and genetic analysis of regulatory regions of coliphage H-19B: location of shiga-like toxin and lysis genes suggest a role for phage functions in toxin release. *Mol. Microbiol.* **28,** 1255–1267 (1998).
15. Grif, K., Dierich, M. P., Karch, H. & Allerberger, F. Strain-specific differences in the amount of Shiga toxin released from enterohemorrhagic *Escherichia coli* O157 following exposure to subinhibitory concentrations of antimicrobial agents. *Eur. J. Clin. Microbiol. Infect. Dis.* **17,** 761–766 (1998).
16. McDonough, M. A. & Butterton, J. R. Spontaneous tandem amplification and deletion of the shiga toxin operon in *Shigella dysenteriae* 1. *Mol. Microbiol.* **34,** 1058–1069 (1999).
17. Lawrence, J. G. & Ochman, H. Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44,** 383–397 (1997).
18. Francino, M. P., Chao, L., Riley, M. A. & Ochman, H. Asymmetries generated by transcription-coupled repair in enterobacterial genes. *Science* **272,** 107–109 (1996).
19. Blaisdell, J. O., Hatahet, Z. & Wallace, S. S. A novel role for *Escherichia coli* endonuclease VIII in prevention of spontaneous G→T transversions. *J. Bacteriol.* **181,** 6396–6402 (1999).
20. Le Page, F. *et al.* Transcription-coupled repair of 8-oxoguanine: requirement for XPG, TFIIH, and CSB and implications for Cockayne syndrome. *Cell* **101,** 159–171 (2000).
21. Boyd, E. F., Li, J., Ochman, H. & Selander, R. K. Comparative genetics of the *inv–spa* invasion gene complex of *Salmonella enterica*. *J. Bacteriol.* **179,** 1985–1991 (1997).
22. Sharp, P. M., Kelleher, J. E., Daniel, A. S., Cowan, G. M. & Murray, N. E. Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. *Proc. Natl Acad. Sci. USA* **89,** 9836–9840 (1992).
23. Alm, R. A. & Trust, T. J. Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *J. Mol. Med.* **77,** 834–846 (1999).
24. Read, T. D. *et al.* Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* **28,** 1397–1406 (2000).
25. Ohnishi, M. *et al.* Chromosome of the enterohemorrhagic *Escherichia coli* O157:H7: comparative analysis with K-12 MG1655 revealed the acquisition of a large amount of foreign DNAs. *DNA Res.* **6,** 361–368 (1999).
26. Mahillon, J. *et al.* Subdivision of *Escherichia coli* K-12 genome for sequencing: manipulation and DNA sequence of transposable elements introducing unique restriction sites. *Gene* **223,** 47–54 (1998).
27. Lin, J. *et al.* Whole-genome shotgun optical mapping of *Deinococcus radiodurans*. *Science* **285,** 1558–1562 (1999).
28. Lukashin, A. V. & Borodovsky, M. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26,** 1107–1115 (1998).
29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
30. Delcher, A. L. *et al.* Alignment of whole genomes. *Nucleic Acids Res.* **27,** 2369–2376 (1999).

Acknowledgements

We thank T. Forsythe, M. Goeden, H. Kijenski, B. Leininger, J. McHugh, B. Peterson, G. Peyrot, D. Sands, P. Soni, E. Travanty and other members of the University of Wisconsin genomics team for their expert technical assistance. This work was funded by grants from the NIH (NIAID and NCHGR), the University of Wisconsin Graduate School and the RMHC to F.R.B., the NIH (NCHGR, NIAID) to D.C.S., HHMI/OTKA to G.P., an Alfred P. Sloan/DOE Fellowship to B.M., a CDC/APHL Fellowship to P.S.E., and an Alfred P. Sloan/NSF Fellowship to N.T.P. Sixteen University of Wisconsin undergraduates participated in this work and particular thanks are due to A. Byrnes for web-site development to complement this project, and to A. Darling for programming.

Correspondence and requests for materials should be addressed to N.T.P. (e-mail: perna@ahabs.wisc.edu). The GenBank accession number for the annotated sequence is AE00517H.

Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF

Vishwanath R. Iyer*†‡, Christine E. Horak§‡, Charles S. Scafe‖¶, David Botstein‖, Michael Snyder§ & Patrick O. Brown*#

** Department of Biochemistry and # Howard Hughes Medical Institute, Stanford University Medical Center, Stanford, California 94305, USA*
‖ Department of Genetics, Stanford University Medical Center, Stanford, California 94305, USA
§ Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA
† These authors contributed equally to this work

Proteins interact with genomic DNA to bring the genome to life; and these interactions also define many functional features of the genome. SBF and MBF are sequence-specific transcription factors that activate gene expression during the G1/S transition of the cell cycle in yeast[1,2]. SBF is a heterodimer of Swi4 and Swi6, and MBF is a heterodimer of Mbp1 and Swi6 (refs 1, 3). The related Swi4 and Mbp1 proteins are the DNA-binding components of the respective factors, and Swi6 may have a regulatory function[4,5]. A small number of SBF and MBF target genes have been identified[3,6–10]. Here we define the genomic binding sites of the SBF and MBF transcription factors *in vivo*, by using DNA microarrays. In addition to the previously characterized targets, we have identified about 200 new putative targets. Our results support the hypothesis that SBF activated genes are predominantly involved in budding, and in membrane and cell-wall biosynthesis, whereas DNA replication and repair are the dominant functions among MBF activated genes[6,11]. The functional specialization of these factors may provide a mechanism for independent regulation of distinct molecular processes that normally occur in synchrony during the mitotic cell cycle.

To identify the targets of SBF and MBF, we combined chromatin immunoprecipitation and microarray hybridization (Fig. 1). Proteins were crosslinked with formaldehyde to their target sites *in vivo*. DNA that was specifically crosslinked to either of the transcription factors was purified by immunoprecipitation using an antibody against either the native protein or an epitope tag that was fused to the protein. Polymerase chain reaction (PCR) analysis of immunoprecipitated DNA confirmed the specific association of Swi4, Swi6 and Mbp1 with several known target promoters, and other target promoters that are identified here (see Supplementary Information). After reversal of the crosslinks, immunoprecipitated DNA was amplified and fluorescently labelled with the Cy5 fluoro-

‡ Present address: Institute of Molecular and Cellular Biology, University of Texas at Austin, Austin, Texas 78712, USA.
¶ Present address: Applied Biosystems, Foster City, California 94404, USA.