Received 21 July; accepted 8 December 2000.

1. Nield, J. *et al.* Three-dimensional structure of *Chlamydomonas reinhardtii* and *Synechococcus elongatus* photosystem II complexes allow for comparison of their oxygen-evolving complex organisation. *J. Biol. Chem.* **275,** 27940–27946 (2000).
2. Rhee, K.-H., Morris, E. P., Barber, J. & Kühlbrandt, W. Three-dimensional structure of photosystem II reaction centre at 8 Å resolution. *Nature* **396,** 283–286 (1998).
3. Zouni, A., Jordan, R., Schlodder, E., Fromme, P. & Witt, H. T. First photosystem II crystals capable of water oxidation. *Biochim. Biophys. Acta* **1457,** 103–105 (2000).
4. Witt, H. T. Primary reactions of oxygenic photosynthesis. *Ber. BunsenGes. Phys. Chem.* **100,** 1923–1942 (1996).
5. Zouni, A. *et al.* in *Photosynthesis: Mechanisms and Effects* (ed. Garab, G.) 925–928 (Kluwer Academic, Dordrecht, 1998).
6. Barry, B. A., Boerner, R. J. & de Paula, J. C. in *The Molecular Biology of Cyanobacteria* (ed. Bryant, D. A.) 217–257 (Kluwer Academic, Dordrecht, 1994).
7. Svensson, B. *et al.* A model for the photosystem II reaction center core including the structure of the primary donor P680. *Biochemistry* **35,** 14486–14502 (1996).
8. Xiong, J., Subramaniam, S. & Govindjee. A knowledge-based three dimensional model of the photosystem II reaction center of *Chlamydomonas reinhardtii*. *Photosynth. Res.* **56,** 229–254 (1998).
9. Michel, H. & Deisenhofer, J. Relevance of the photosynthetic reaction center from purple bacteria to the structure of photosystem II. *Biochemistry* **27,** 1–7 (1988).
10. Schubert, W.-D. *et al.* A common ancestor for oxygenic and anoxygenic photosynthetic systems: a comparison based on the structural model of photosystem I. *J. Mol. Biol.* **280,** 297–314 (1998).
11. Harrer, R., Bassi, R., Testi, M. G. & Schäfer, C. Nearest-neighbor analysis of a photosystem II complex from *Marchantia polymorpha* L. (liverwort), which contains reaction center and antenna proteins. *Eur. J. Biochem.* **255,** 196–205 (1998).
12. Ishikawa, Y. *et al.* Turnover of the aggregates and cross-linked products of the D1 protein generated by acceptor-side photoinhibition of photosystem II. *Biochim. Biophys. Acta* **1413,** 147–158 (1999).
13. Rhee, K.-H. *Three-dimensional Structure of Photosystem II Reaction Center by Electron Cryo-microscopy.* Thesis, Univ. Heidelberg (1998).
14. Tomo, T., Enami, I. & Satoh, K. Orientation and nearest neighbor analysis of psbI gene product in the photosystem II reaction center complex using bifunctional cross-linkers. *FEBS Lett.* **323,** 15–18 (1993).
15. Shi, L. X., Kim, S. J., Marchant, A., Robinson, C. & Schröder, W. P. Characterisation of the PsbX protein from photosystem II and light regulation of its gene expression in higher plants. *Plant. Mol. Biol.* **40,** 737–744 (1999).
16. Summer, E. J., Schmid, V. H., Bruns, B. U. & Schmidt, G. W. Requirement for the H phosphoprotein in photosystem II of *Chlamydomonas reinhardtii*. *Plant. Physiol.* **113,** 1359–1368 (1997).
17. Zheleva, D., Sharma, J., Panico, M., Morris, H. R. & Barber, J. Isolation and characterization of monomeric and dimeric CP47-reaction center photosystem II complexes. *J. Biol. Chem.* **273,** 16122–16127 (1998).
18. Mayers, S. R. *et al.* Further characterization of the psbH locus of *Synechocystis* sp. PCC 6803: inactivation of psbH impairs Q_A to Q_B electron transport in photosystem 2. *Biochemistry* **32,** 1454–1465 (1993).
19. Ahmed, A., Tajmir-Riahi, H. A. & Carpentier, R. A quantitative secondary structure analysis of the 33 kDa extrinsic polypeptide of photosystem II by FTIR spectroscopy. *FEBS Lett.* **363,** 65–68 (1995).
20. Ghanotakis, D. F., Tsiotis, G. & Bricker T. M. in *Concepts in Photobiology: Photosynthesis and Photomorphogenesis* (eds Singhal, G. S., Renger, G., Sopory, S. K., Irrgang, K.-D. & Govindjee) 264–291 (Narosa, New Delhi, 1999).
21. Zech, S. G. *et al.* Pulsed EPR measurement of the distance between $P680^{+-}$ and Q_A^{--} in photosystem II. *FEBS Lett.* **414,** 454–456 (1997).
22. Schelvis, J. P. M., van Noort, P. I., Aartsma, T. J. & van Gorkom, H. J. Energy transfer, charge separation and pigment arrangement in the reaction center of Photosystem II. *Biochim. Biophys. Acta* **1184,** 242–250 (1994).
23. Ruffle, S., Hutchison, R. & Sayre, R. T. in *Photosynthesis: Mechanisms and Effects* (ed. Garab, G.) 1013–1016 (Kluwer Academic, Dordrecht, 1998).
24. Buser, C. A., Diner, B. A. & Brudvig, G. W. Photooxidation of cytochrome b559 in oxygen-evolving photosystem II. *Biochemistry* **31,** 11449–11459 (1992).
25. Yachandra, V. K., Sauer, K. & Klein, M. P. Manganese cluster in photosynthesis: where plants oxidize water to dioxygen. *Chem. Rev.* **96,** 2927–2950 (1996).
26. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1996).
27. CCP4 Collaborative computational project number 4. The CCP4 suite: programmes for protein crystallography. *Acta Cryst. D* **50,** 760–763 (1994).
28. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47,** 110–119 (1991).
29. Esnouf, R. M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics* **15,** 132–134 (1997).
30. Merritt, E. A. & Bacon, D. J. Raster3D—photorealistic molecular graphics. *Methods Enzymol.* **277,** 505–524 (1997).

Acknowledgements

We thank C. Lüneberg, H. Schmidt and D. DiFiore for technical assistance, and W.-D. Schubert, R. Bittl, E. Schlodder, K. Irrgang and P. Jordan for discussions. Beamline assistance at DESY (Hamburg), ESRF (Grenoble) and Elettra (Trieste), and help of M. Burghammer with data collection at the manganese edge is gratefully acknowledged. We thank Egbert J. Boekema for providing us with electron micrographs of PSII. This work was supported by Deutsche Forschungsgemeinschaft, Sonderforschungsbereiche 312 and 498, BMBF (W.S.), and Fonds der Chemischen Industrie (W.S., H.-T.W., N.K. and P.F.).

Correspondence and requests for materials should be addressed to W.S. (e-mail: saenger@chemie.fu-berlin.de) or H.-T.W. (e-mail: witt@phosis1.chem.tu-berlin.de). Atomic coordinates are deposited in the Protein DataBank under accession number 1FE1.

corrections

Language trees support the express-train sequence of Austronesian expansion

Russell D. Gray & Fiona M. Jordan

Nature **405,** 1052–1055 (2000).

There was an error in the geographical character-state mapping in this paper. The authors inadvertently reported the values corresponding to mapping these characters onto the tree in an unordered manner. Correctly ordering the character-state changes according to the 'express-train' model does not change the main conclusion of the paper: the express-train model fits much better than would be expected owing to chance. The correct fit is 18 steps, and the fit of the randomly assigned character states now ranges from 95 to 122 (mean 108.4, s.d. 5.1). □

Warm-coding deficits and aberrant inflammatory pain in mice lacking $P2X_3$

Veronika Souslova, Paolo Cesare, Yanning Ding, Armen N. Akopian, Louise Stanfa, Rie Suzuki, Katherine Carpenter, Daniela Nebenius-Oosthuizen, Andrew J. H. Smith, Emma J. Kidd & John N. Wood

Nature **407,** 1015–1017 (2000)

The address of the author Emma J. Kidd was cited incorrectly. Her affiliation should have been given as the Glaxo Institute of Applied Pharmacology, Department of Pharmacology. University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, UK, which was where her work was carried out. The Welsh School of Pharmacy, Cardiff University, Cardiff CF1 3XF, UK, which was cited as her affiliation in this paper, is her present address. □

erratum

The protein–protein interaction map of *Helicobacter pylori*

Jean-Christophe Rain, Luc Selig, Hilde De Reuse, Véronique Battaglia, Céline Reverdy, Stéphane Simon, Gerlinde Lenzen, Fabien Petel, Jérôme Wojcik, Vincent Schächter, Y. Chemama, Agnès Labigne & Pierre Legrain

Nature **409,** 211–215 (2001).

The list of interactions between *Helicobacter pylori* proteins that are described and analysed in this paper are available as Supplementary Information on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*. Proteins are named according to the nomenclature of The Institute for Genomic Research microbial database. □

aluminum foil) and light bottles (28–33 μmol photons m^{-2} s^{-1}) and incubated them for 24 h at 3 °C. We isolated individual ciliates from each light and dark replicate, washed them three times with sterile media (10 ml) and transferred ten washed cells into a scintillation vial. During the 2–3 h isolation period we maintained the samples in the dark on ice. We prepared the samples for liquid scintillation counting as described[28]. We calculated rates of photosynthesis by subtracting average ^{14}C fixation in the dark from fixation in the light. For the determination of *M. rubrum* chlorophyll, we isolated and washed ten cells before transferring them into vials containing cold 90% acetone and incubating them at −20 °C overnight for extraction. We measured chlorophyll *a* using a 10-AU Turner fluorometer.

Received 15 February; accepted 2 May 2000.

1. Small, E. B. & Lynn, D. H. in *Illustrated Guide to the Protozog* (eds Lee, J. J., Hunter, S. H. & Bovee, E. C.) 393–575 (Allen Press, Kansas, 1985).
2. Krainer, K. -H. & Foissner, W. Revision of the genus *Askenasia* Blochmann, 1895, with proposal of two new species, and description of *Rhabdoaskenasia minima* N. G., N. Sp. (Ciliophora, Cyclotrichida). *J. Protozool.* **37,** 414–427 (1990).
3. Montagnes, D. J. S. & Lynn, D. H. The annual cycle of *Mesodinium rubrum* in the waters surrounding the Isles of Shoals, Gulf of Maine. *J. Plank. Res.* **11,** 193–201 (1989).
4. Lindholm, T. in *Advances in Aquatic Microbiology.* Vol. 3 (eds Jannasch, H. W. & Williams, P. J.) 1–48 (Academic, London 1985).
5. Crawford, D. W. *Mesodinium rubrum*: the phytoplankter that wasn't. *Mar. Ecol. Prog. Ser.* **58,** 161–174 (1989).
6. Oakley, B. R. & Taylor, F. J. R. Evidence for a new type of endosymbiotic organization in a population of the ciliate *Mesodinium rubrum* from British Columbia. *BioSystems* **10,** 361–369 (1978).
7. Hibberd, D. J. Observations on the ultrastructure of the cryptomonad endosymbiont of the red-water ciliate *Mesodinium rubrum*. *J. Mar. Biol. Assoc. UK* **57,** 45–61 (1977).
8. Lindholm, T., Lindroos, P. & Mörk, A.-C. Ultrastructure of the photosynthetic ciliate *Mesodinium rubrum*. *BioSystems* **21,** 141–149 (1988).
9. Taylor, F. J. R., Blackbourn, D. J. & Blackbourn, J. The red-water ciliate *Mesodinium rubrum* and its "incomplete symbionts": a review including new ultrastructural observations. *J. Fish. Res. Bd Can.* **28,** 391–407 (1971).
10. Barber, R. T., White, A. W. & Siegelman, H. W. Evidence for a cryptomonad symbiont in the ciliate, *Cyclotrichium meunieri*. *J. Phycol.* **5,** 86–88 (1969).
11. Laval-Peuto, M., Salvano, P., Gayol, P. & Gruet, C. Mixotrophy in marine planktonic ciliates: ultrastructure study of *Tontonia appendiculariformis* (Ciliophora, Oligotrichina). *Mar. Microb. Food Webs* **1,** 81–104 (1986).
12. Stoecker, D. K. & Silver, M. W. Replacement and aging of chloroplasts in *Strombidium capitatum* (Ciliophora:Oligotrichida). *Mar. Biol.* **107,** 491–502 (1990).
13. Lindholm, T. & Mörk, A.-C. Symbiotic algae and plastids in planktonic ciliates. *Mem. Soc. Fauna Flora Fennica* **65,** 17–22 (1989).
14. Hargraves, P. Narrow River phytoplankton. *Maritimes* **35,** 6–8 (1991).
15. Packard, T. T., Blasco, D. & Barber, R. T. in *Upwelling Ecosystems* (eds Boje, R. & Tomczak, M.) 73–89 (Springer, New York, 1978).
16. Crawford, D. W., Purdie, D. A., Lockwood, A. P. M. & Weissman, P. Recurrent red-tides in the Southampton Water estuary caused by the phototrophic ciliate *Mesodinium rubrum*. *Est. Coast. Shelf Sci.* **45,** 799–812 (1997).
17. Smith, W. O. & Barber, R. T. A carbon budget for the autrophic ciliate *Mesodinium rubrum*. *J. Phycol.* **15,** 27–33 (1979).
18. Wilkerson, F. P. & Grunseich, G. Formation of blooms by the symbiotic ciliate *Mesodinium rubrum:* the significance of nitrogen uptake. *J. Plank. Res.* **12,** 973–989 (1990).
19. Kifle, D. & Purdie, D. A. The seasonal abundance of the phototrophic ciliate *Mesodinium rubrum* in Southampton Water, England. *J. Plank. Res.* **15,** 823–833 (1993).
20. Stoecker, D. K., Taniguchi, A. & Michaels, A. E. Abundance of autotrophic, mixotrophic and heterotrophic planktonic ciliates in shelf and slope waters. *Mar. Ecol. Prog. Ser.* **50,** 241–254 (1989).
21. Stoecker, D. K., Putt, M., Davis, L. H. & Michaels, A. E. Photosynthesis in *Mesodinium rubrum*: species-specific measurements and comparison to community rates. *Mar. Ecol. Prog. Ser.* **73,** 245–252 (1991).
22. Sanders, R. W. Seasonal distributions of the photosynthesizing ciliates *Laboea strobila* and *Myrionecta rubra* (= *Mesodinium rubrum*) in an estuary of the Gulf of Maine. *Aquatic Microb. Ecol.* **9,** 237–242 (1995).
23. Crawford, D. W. & Lindholm, T. Some observations on vertical distribution and migration of the phototrophic ciliate *Mesodinium rubrum* (= *Myrionecta rubra*) in a stratified brackish inlet. *Aquatic Microb. Ecol.* **13,** 267–274 (1997).
24. Yih, W. & Shim, J. H. The planktonic phototrophic ciliate, *Mesodinium rubrum*, as a useful organism for marine biotechnological applications. *J. Mar. Biotechnol.* **5,** 82–85 (1997).
25. Stoecker, D. K. in *Protozoa and Their Role in Marine Processes* Vol. **G 25** (eds Reid, P. C. *et al.*) 161–179 (NATO ASI Series, Springer, Berlin–Heidelberg, 1991).
26. Stoecker, D. K., Michaels, A. E. & Davis, L. H. Large proportion of marine planktonic ciliates found to contain functional chloroplasts. *Nature* **326,** 790–792 (1987).
27. White, A. W., Sheath, R. G. & Hellebust, J. A. A red tide caused by the marine ciliate *Mesodinium rubrum* in Passamaquoddy Bay, including pigment and ultrastructure studies of the endosymbiont. *J. Fish. Res. Bd Can.* **34,** 413–416 (1977).
28. Stoecker, D. K., Silver, M. W., Michaels, A. E. & Davis, L. H. Obligate mixotrophy in *Laboea strobila*, a ciliate which retains chloroplasts. *Mar. Biol.* **99,** 415–423 (1988).

Acknowledgements

We thank P. Del Giorgio for discussions; D. W. Coats, P. Krugens and G. Sellers for taxonomic assistance; and A. Li for isolating the cryptophyte. This work was partially supported by the US NSF Polar and Biological Oceanography (D.K.S.) and Maryland Sea Grant (K.S.).

Correspondence and requests for material should be addressed to D.E.G. (e-mail: gustafsn@hpl.umces.edu).

Language trees support the express-train sequence of Austronesian expansion

Russell D. Gray & Fiona M. Jordan

Department of Psychology, University of Auckland, Auckland 92019, New Zealand

Languages, like molecules, document evolutionary history. Darwin[1] observed that evolutionary change in languages greatly resembled the processes of biological evolution: inheritance from a common ancestor and convergent evolution operate in both. Despite many suggestions[2–4], few attempts have been made to apply the phylogenetic methods used in biology to linguistic data. Here we report a parsimony analysis of a large language data set. We use this analysis to test competing hypotheses—the "express-train"[5] and the "entangled-bank"[6,7] models—for the colonization of the Pacific by Austronesian-speaking peoples. The parsimony analysis of a matrix of 77 Austronesian languages with 5,185 lexical items produced a single most-parsimonious tree. The express-train model was converted into an ordered geographical character and mapped onto the language tree. We found that the topology of the language tree was highly compatible with the express-train model.

There are many parallels between the processes of biological and linguistic evolution and the methods used to analyse them[4]. Despite these parallels, however, historical linguists have not used the quantitative phylogenetic methods that have revolutionized evolutionary biology in the past 20 years[8]. So, although linguists routinely use the "comparative method"[9] to construct language family trees from discrete lexical, morphological and phonological data, they do not use an explicit optimality criterion to select the best tree, nor do they typically use an efficient computer algorithm to search for the best tree from the discrete data. This is surprising given that the task of finding the best tree is inherently a combinatorial optimization problem of considerable computational difficulty[10]. One potential problem with a quantitative phylogenetic approach to linguistic evolution arises from the more reticulate nature of cultural evolution. Some authors[11,12] have claimed that reticulate processes in linguistic evolution overshadow those of descent, leading them to reject the appropriateness of the family-tree model. We believe that this is an empirical claim, which can be evaluated using phylogenetic methods. If the data fit well on the tree and there is little systematic conflicting signal, then the family-tree model is supported. If the data fit poorly, then alternative phylogenetic methods that do not assume a tree model, such as spectral analysis or split decomposition, should be investigated. A critical part of phylogenetic inference involves testing for congruence between independent lines of evidence. Here we test a model of the colonization of the Pacific that is derived from predominantly archaeological data by quantitatively examining its fit with a parsimony tree of Austronesian languages.

Prehistoric human colonization in the Pacific happened in two phases. Initially, Pleistocene hunter–gatherer expansions from Island Southeast Asia through New Guinea reached the Bismarck archipelago by 33,000 BP and the Papuan-speaking descendants of these people are dispersed throughout New Guinea and parts of Island Melanesia[13]. The second colonization wave of Austronesian language speakers involved a diaspora of Neolithic farming peoples out of south China and Taiwan around 6,000 BP[13–15]. According to the 'express train to Polynesia' model, the Austronesian expansion from Taiwan was extremely rapid, taking roughly 2,100 years to reach the edges of western Polynesia—a distance of 10,000 kilometres.

Converging evidence from archaeology and molecular anthropology supports a rapid and relatively encapsulated dispersal of the Austronesian speakers throughout the Pacific[13,16–18] (Fig. 1); however, there is some dispute about the exact degree of interaction with earlier Melanesian settlers, the rate at which the migration occurred and the extent and location of any colonization pauses[19]. In broad terms, most Pacific scholars seem to favour the express-train model, but others have argued that the ancestral Polynesians derive from an older Melanesian "matrix"[7,20]. The latter authors stress that a phylogenetic, colonization-focused perspective obscures the high degree of prehistoric contact and inter-relationships amongst Pacific people; we use Terrell's phrase[6]—the entangled-bank model—to represent this. These two models are not mutually exclusive, but are best characterized as two ends of a continuum of modes of human prehistory, with a pure tree at one end and a maximally connected network at the other. The issues surrounding the settlement of the Pacific are thus a microcosm of the general debate about whether human cultural evolution can be appropriately represented as a tree.

We tested one aspect of the express-train model, the colonization sequence, in the way that biologists test hypotheses about the sequence of events in biological evolution. We constructed a tree and then mapped the trait onto the tree to see whether the inferred sequence of changes fits a particular scheme[21]. Figure 2 shows how a simple colonization scheme can be tested by mapping geography onto an independent tree. We grouped languages according to Diamond's archaeological/geographical stations[5,22]. Using character-state functions in the program MacClade[23], we assigned each station a character state from 0 to 9. The states were ordered in a character-state tree to fit the sequence proposed by the express-train model. For example, in Fig. 1 the Taiwanese languages were grouped as state 1, the Remote Oceanic languages as state 8; this means a change from state 1 to 8 would require five steps (according to the model presented in Fig. 1). By mapping these character states onto the most-parsimonious language tree (Fig. 3), we were able to evaluate the express-train model in a quantitative manner. If the language tree fits the express-train model well, then the character-state tree should fit well onto our obtained tree. The shortest possible tree length required to optimize the character-state tree onto the language tree was nine (that is, the number of character states minus one). When the character-state tree was mapped onto the optimal tree, we obtained a tree length of 13. To assess the statistical significance of the fit, we randomly shuffled the character states between the 77 languages 200 times[23]. This gave us a null distribution of tree lengths with a mean tree length of 48.9 steps (s.d. 1.98, range 43–53). This indicates that the express-train character-state tree fits the language tree with significantly fewer steps than would occur by chance. In fact, the obtained fit was very close to the shortest possible length (nine), indicating that the express-train model fits the language tree exceptionally well.

By definition, an entangled-bank model cannot be represented by a character-state tree; however, we can assess whether the language data support the entangled-bank model by examining the topology

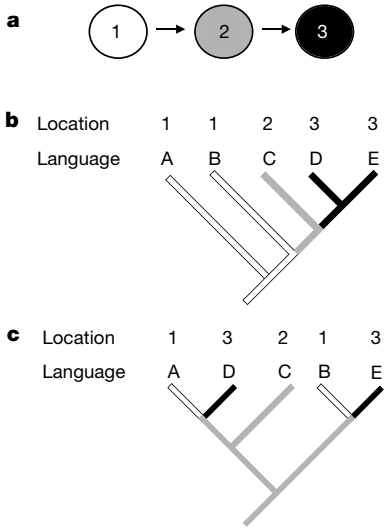

Figure 2 A phylogenetic approach to testing a colonization sequence. **a**, Model for the colonization of three areas, in which an ancestral population moves from area one to area two and then to area three. **b**, Tree that fits perfectly with the colonization model in **a** (fit = 2 steps). **c**, A tree that fits poorly with the colonization sequence (fit = 4 steps).

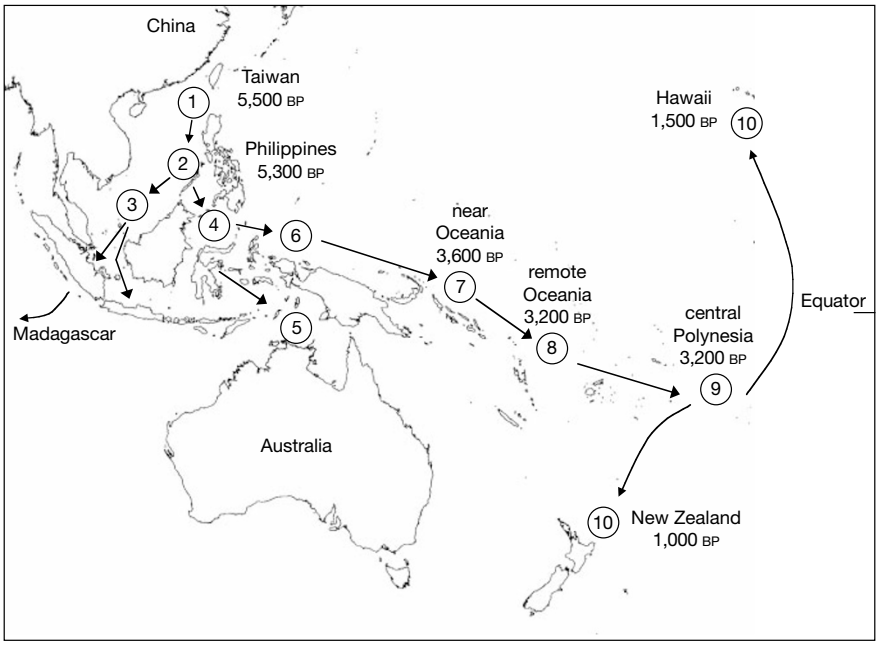

Figure 1 The express train to Polynesia model of the Austronesian colonization of the Pacific (adapted from refs 5 and 22). Approximate archaeological dates of settlement are indicated[13,22]. Each 'station' is a separate character state: 1, Taiwan; 2, Philippines, Chamorro, Palau; 3, Borneo, Indonesia, Malay; 4, Sulawesi; 5, central Malayo-Polynesian; 6, south Halmahera/west New Guinea; 7, near Oceania; 8, remote Oceania; 9, central Polynesia; 10, east Polynesia.

of the tree. While advocates of this model make no predictions about the likely shape of a language tree under an entangled-bank conception, they argue that large-scale migration patterns in languages are obscured by culture contact[7]. Consequently, they might predict a layered, 'candelabra-like' tree that emphasizes regional contact. In contrast, an (archaeologically) quick colonizing wave from Island Southeast Asia through the Pacific to Polynesia should produce a tree topology that is 'chain-like' (see Fig. 3). Proponents of the entangled-bank model argue that culture, language and biology 'combine and recombine' in such complex interactions that patterns of language relationships may tell us very little about the history of language speakers[7]. In this case, the tree should merely reflect geographical proximity. Our tree, however, shows several cases where the relationships fit the historical sequences implied by the express-train model but conflict with geographical proximity (see Fig. 3).

Although we reject the specific features of the entangled-bank model, we do not claim that Austronesian cultural history is totally tree-like. The consistency index (a measure of the fit of the lexical data on the tree) is only 0.25. This value is not substantially lower than would be expected for equivalently sized morphological and molecular data sets[24] in which hybridization is uncommon. Although it is probable that much of the poor fit in the lexical data is due to the loss of cultural or linguistic features[15,25], archaeological[26] and genetic[27] evidence do indicate that population interaction and 'borrowing' are likely to have occurred even between far-flung archipelagoes. A way of approaching the issue of borrowing is to examine languages whose placement conflicts with the colonization scheme. For example, Buli and Numfor are grouped inside the Oceanic language group on our tree, whereas the express-train model places these south Halmahera/west New Guinea languages outside the Oceanic group. Similarly, Chamorro and Palau—languages whose closest relationships are most likely with the Philippines[28]—are grouped with the Oceanic languages. In both these cases, borrowing is a likely cause of the incongruence between the express-train model and our tree. More detailed evidence for specific patterns of reticulation is evaluated elsewhere (F.M.J. and R.D.G, manuscript in preparation).

The patterns apparent in linguistic relationships are integrally tied to the movements, contacts and activities of language speakers. Our preliminary investigations have shown that a phylogenetic approach to languages offers the ability to test hypotheses about human prehistory. In biology, phylogenetic methods have become invaluable tools for investigating patterns and processes in evolution. In the future, phylogenetic methods may provide a common methodology and analytic framework to integrate data from ethnography, archaeology, linguistics and genetics. This is an important step towards a unified approach to biological and cultural evolution. □

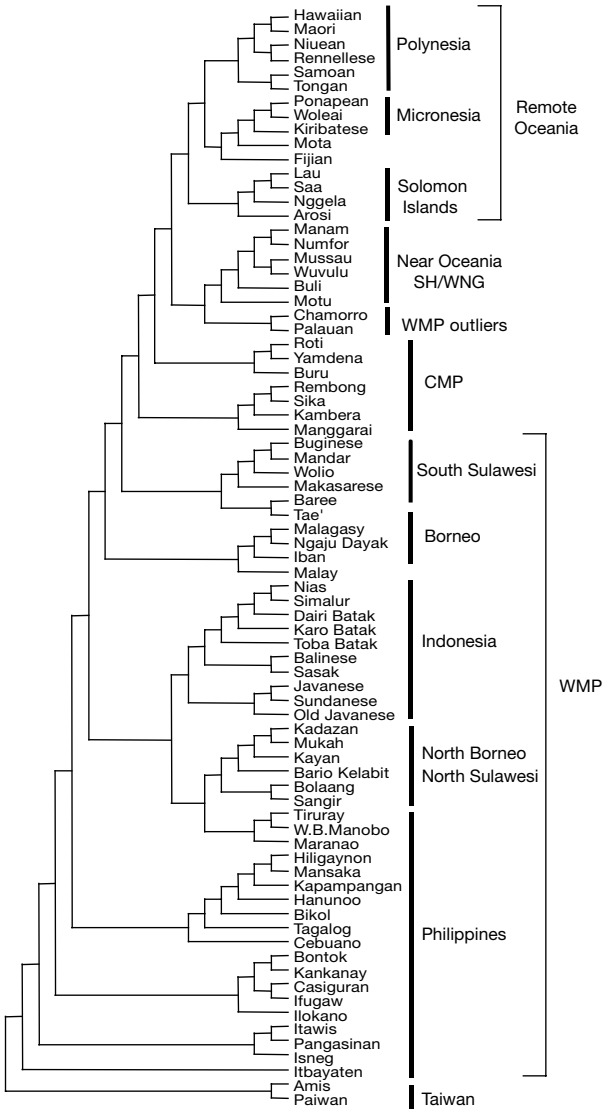

Figure 3 Phylogenetic tree of 77 Austronesian languages. WMP, Western Malayo-Polynesian; CMP, Central Malayo-Polynesian; SH/WNG, South Halmahera/West New Guinea. The topology of the tree shows considerable agreement with traditional linguistic groupings[14,15,28]; these groupings reflect historical relationships, not just geographical proximity. For instance, Malagasy (spoken on Madagascar) is grouped with Ngaju Dayak from western Borneo. Tae' (from Central Sulawesi) groups within the south Sulawesi languages, whereas the north Sulawesi languages (for example, Bolaang Mongondow (Bolaang)) are more closely related to languages of north Borneo than are other Sulawesi languages.

Methods

Data were taken from Blust's Austronesian Comparative Dictionary (R. Blust, personal communication). This is a continuing project to compile comparative lexical data from the largest language family in the world. Currently, the dictionary is 25% complete and comprises 5,185 lexical items across 191 languages. Each lexical item has a set of cognate terms listed with the languages in which they appear. To ensure that there was sufficient information in the data set for phylogenetic analysis, we cut the number of languages from 191 to 68 by using a criterion of 150 or more appearances in a cognate set. An additional nine languages were then added to provide a balanced representation of the principal Austronesian language subgroups, giving us 77 languages in total. The presence of a language in a cognate set was coded as '1' in a matrix of 77 languages × 5,185 lexical items. If a language was not present in a particular cognate set, that language was coded as '0' for that item in the matrix. Linguistic[15,28], archaeological[13] and genetic[16,18] evidence agrees that Taiwan is the most likely Austronesian homeland, and so the two Taiwanese languages (Amis and Paiwan) were used to root the tree. We used PAUP* 4.0d65 (ref. 29) to find the set of most-parsimonious trees. To maximize the chance of finding optimal trees, 1,200 random addition sequences and tree bisection–reconnection branch swapping were used. Characters were typed as easy loss (5:1 ratio) on the assumption that independent losses of lexical items were more likely than independent gains. Similar assumptions about character coding have been used for complex behavioural characters[30], and linguistic features (such as phonemes) have been shown to be lost in a west-to-east direction across the Pacific[25]. Other easy loss codings and equally weighted parsimony produced similar results (R.D.G. and F.M.J., manuscript in preparation). The search found one shortest tree of 52,129 steps with a consistency index of 0.25. The linguistic data set contained significant phylogenetic signal (treelength skewness index $g_1 = -0.505$ calculated from 100,000 random trees).

Received 6 January; accepted 28 April 2000.

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Oxford Univ. Press, Oxford, 1871).
2. Kirch, P. V. & Green, R. C. History, phylogeny, and evolution in Polynesia. *Curr. Anthropol.* **28,** 431–456 (1987).
3. Mace, R. & Pagel, M. The comparative method in anthropology. *Curr. Anthropol.* **35,** 549–564 (1994).
4. Ruvolo, M. in *Biological Metaphor and Cladistic Classification: An Interdisciplinary Perspective* (eds Hoenigswald, H. M. & Wiener, L. F.) 193–216 (Univ. Pennsylvania Press, Pennsylvania, 1987).
5. Diamond, J. M. Express train to Polynesia. *Nature* **336,** 307–308 (1988).
6. Terrell, J. History as a family tree, history as an entangled bank: Constructing images and interpretations of prehistory in the South Pacific. *Antiquity* **62,** 642–657 (1988).
7. Terrell, J., Hunt, T. L. & Gosden, C. The dimensions of social life in the Pacific. *Curr. Anthropol.* **38,** 155–195 (1997).
8. Hillis, D. M., Huelsenbeck, J. P. & Cunningham, C. W. Application and accuracy of molecular phylogenies. *Science* **264,** 671–676 (1994).
9. Crowley, T. *An Introduction to Historical Linguistics* 3rd edn (Oxford Univ. Press, Auckland, 1997).

10. Warnow, T. Mathematical approaches to comparative linguistics. *Proc. Natl Acad. Sci. USA* **94,** 6585–6590 (1997).
11. Bateman, R. *et al.* Speaking of forked tongues. *Curr. Anthropol.* **31,** 1–24 (1990).
12. Moore, J. H. Putting anthropology back together again: The ethnogenetic critique of cladistic theory. *Am. Anthropol.* **96,** 925–948 (1994).
13. Bellwood, P. *The Prehistory of the Indo-Malaysian Archipelago* 2nd edn (Univ. Hawaii Press, Honolulu, 1997).
14. Bellwood, P. The Austronesian dispersal and the origin of languages. *Sci. Am.* **265,** 88–93 (1991).
15. Blust, R. The prehistory of the Austronesian speaking peoples: a view from language. *J. World Prehist.* **9,** 453–510 (1995).
16. Melton, T. *et al.* Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am. J. Hum. Genet.* **57,** 403–414 (1995).
17. Green, R. C. Integrating historical linguistics with archaeology: Insights from research in Remote Oceania. *Bull. Indo-Pacific Prehist. Ass.* **18,** 3–16 (1999).
18. Lum, J. K. & Cann, R. L. mtDNA and language support a common origin of Micronesians and Polynesians in Island Southest Asia. *Am. J. Phys. Anthropol.* **105,** 109–119 (1998).
19. Kirch, P. V. *The Lapita Peoples* (Blackwell, Cambridge, 1997).
20. Terrell, J. *Prehistory in the Pacific Islands: A Study of Variation in Language, Customs, and Human Biology* (Cambridge Univ. Press, Cambridge, 1986).
21. Martin, A. Hammerhead shark origins. *Nature* **364,** 494 (1993).
22. Diamond, J. M. *Guns, Germs and Steel* (Jonathan Cape, London, 1997).
23. Maddison, W. P. & Maddison, D. R. *MacClade: Analysis of Phylogeny and Character Variation Version 3.05.* (Sinauer Associates, Massachusetts, 1992).
24. Sanderson, M. J. & Donoghue, M. J. Patterns of variation in levels of homoplasy. *Evolution* **43,** 1781–1795 (1989).
25. Blust, R. in *Currents in Pacific Linguistics: Papers on Austronesian Languages and Ethnolinguistics in Honour of George W. Grace* (ed. Blust, R.) 27–42 (Australian National Univ., Canberra, 1991).
26. Weisler, M. I. Hard evidence for prehistoric interaction in Polynesia. *Curr. Anthropol.* **39,** 521–531 (1998).
27. Matisoo-Smith, E. *et al.* Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. *Proc. Natl Acad. Sci. USA* **95,** 15145–15150 (1998).
28. Ross, M. D. in *Comparative Austronesian Dictionary: An Introduction to Austronesian Studies* (ed. Tryon, D. T.) 27–42 (ANU, Canberra, 1994).
29. Swofford, D. L. *Phylogenetic Analysis Using Parsimony (PAUP*) Version 4.0d65.* (Sinauer Associates, Massachusetts, 1999).
30. Cunningham, C. W., Omland, K. E. & Oakley, T. H. Reconstructing ancestral character states: a critical reappraisal. *Trends Ecol. Evol.* **13,** 361–366 (1998).

Acknowledgements

We are grateful to R. Blust for making the Austronesian Comparative Dictionary available to us. We thank M. Corballis, R. Green and A. Rodrigo for comments on the manuscript; R. Clark for advice; and H. Tse for programming assistance.

Correspondence and requests for materials should be addressed to R. D. G. (e-mail: rd.gray@auckland.ac.nz).

Mosaic evolution of brain structure in mammals

Robert A. Barton* & Paul H. Harvey†

** Evolutionary Anthropology Research Group, Department of Anthropology, University of Durham, Durham DH1 3HN, UK*
† Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

The mammalian brain comprises a number of functionally distinct systems. It might therefore be expected that natural selection on particular behavioural capacities would have caused size changes selectively, in the systems mediating those capacities[1–3]. It has been claimed, however, that developmental constraints limited such mosaic evolution, causing co-ordinated size change among individual brain components[3]. Here we analyse comparative data to demonstrate that mosaic change has been an important factor in brain structure evolution. First, the neocortex shows about a fivefold difference in volume between primates and insectivores even after accounting for its scaling relationship with the rest of the brain. Second, brain structures with major anatomical and functional links evolved together independently of evolutionary change in other structures. This is true at the level of both basic brain subdivisions and more fine-grained functional systems. Hence, brain evolution in these groups involved complex relationships among individual brain components.

Studies of mammalian brain evolution have highlighted the neocortex as a structure associated with intelligence and flexible behaviour, which varies enormously in size between species[4–6]. Large-brained mammals, such as primates, tend to have a neocortex that is disproportionately expanded relative to other structures[3]. The extent to which this size variation can be explained by allometric scaling relative to the rest of the brain, as opposed to size changes independent of other brain structures, remains unclear however[3,7]. Figure 1 indicates clearly that neocortex size varies even after accounting for its scaling relationship with the size of the rest of the brain. The three parallel lines with different intercepts indicate taxonomic differences (grade shifts) in relative neocortex size between primates and insectivores, and, within the primates, between strepsirhine and haplorhine sub-orders. Independent contrasts analysis confirms the presence of significant grade shifts in relative neocortex size. First, the slopes are statistically indistinguishable (haplorhine versus strepsirhine primates: $t = 1.6$, degrees of freedom, d.f. $= 37$, $P = 0.13$; primates versus insectivores: $t = 0.6$, d.f. $= 71$, $P = 0.54$). Second, the absolute values of the contrasts between orders and sub-orders are unusually large and beyond the range of all other contrasts in each data set (haplorhine versus strepsirhine residual = 2.8 standard deviations greater than the mean; primate versus insectivore residual = 5.6 standard deviations greater than the mean). On the basis of separate regression equations for insectivores and primates (averaging between strepsirhines and haplorhines), a primate with a non-neocortical brain size of 1,000 mm^3 would have a neocortex nearly five times larger than would an insectivore with the same non-neocortical brain size (881 mm^3 versus 187 mm^3). In some specific cases, we observe an even greater difference in relative size. For example, the common tenrec *Tenrec ecaudatus*, an insectivore, has a non-neocortical brain volume somewhat greater than that of the

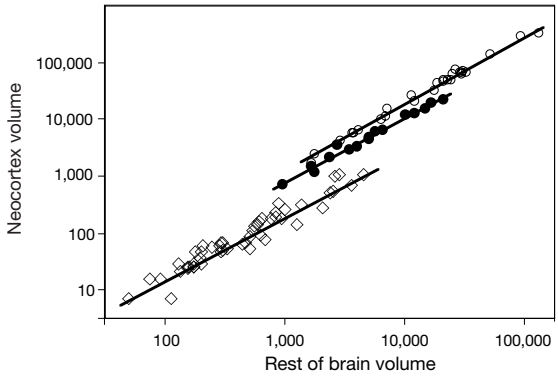

Figure 1 Taxonomic differences in relative neocortex size among primates (strepsirhines and haplorhines) and insectivores. Brain part volumes are in cubic millimetres. Open circles, harplorhine primates; closed circles, strepsirhine primates; diamonds, insectivores. Slopes (and 95% confidence intervals) for insectivores, strepsirhines and haplorhines respectively are 1.11 (1.03–1.20), 1.13 (1.04–1.22) and 1.20 (1.14–1.26).

Table 1 Regression statistics for the scaling of neocortical white and grey matter volume on volume of the rest of the brain

	White matter volume			Grey matter volume		
	Slope	Confidence intervals	r^2	Slope	Confidence intervals	r^2
Insectivores	1.32	1.23–1.41	0.95	1.09	0.94–1.18	0.94
Strepsirhines	1.48	1.32–1.65	0.99	1.06	0.98–1.14	0.99
Haplorhines	1.53	1.37–1.67	0.98	1.12	1.07–1.18	0.99
New World Monkeys	1.40	1.20–1.59	0.98	1.08	0.96–1.21	0.98
Old World monkeys	1.42	0.13–2.71	0.92	0.97	0.45–1.49	0.97

heteromeric $P2X_{2/3}$ receptors appear most important. Second, our formalin test data are consistent with a role for ATP activation of $P2X_3$ in mediating some nociceptive responses to tissue damage. Finally, we show that $P2X_3$ is critical in regulating micturition reflex excitability. One explanation for these data is that ATP, released in response to stretch during distension and filling of the urinary bladder, excites primary afferent voiding circuitry through direct interaction with $P2X_3$ receptors. ATP is released from rabbit urothelium in response to stretch[7], and $P2X_3$ is clearly present on nerve fibres innervating urinary bladder[23] (Fig. 4). Electrophysiological evidence also indicates that α,β-Me-ATP directly activates and desensitizes mechanosensitive pelvic afferents arising from rat urinary bladder[9]. Thus, loss of $P2X_3$ might impair sensory neuron activity during bladder filling, raising the volume threshold for activation of the micturition reflex. As loss of compliance and lowered volume thresholds are a component of many bladder storage disorders (for example, overactive bladder)[24], selective modulation of $P2X_3$ may provide new therapies. The potential for similar $P2X_3$ roles in mechanosensation in other hollow organs (for example, GI tract and lung)[25] needs to be explored. □

Methods

Physiological studies

F_2 and F_3 mice were used for *in vitro* and *in vivo* studies, respectively. All experiments were performed blind. Dissociation of neurons and whole-cell patch-clamp recording was carried out as described previously[26]. Agonists were applied rapidly by microperfusion from a 4-barrel manifold controlled by computer-driven solenoid valves. Exchange of solution around the cell was complete in less than 100 ms. Time between applications was 2 min (nodose) and 3.5 min (DRG), allowing sufficient time to achieve reproducible responses. The minimum detectable response was 20 pA. Traces were acquired using FETCHEX (pCLAMP V.6.04 software, Axon Instruments), and plotted using ORIGIN V.4.1 (Microcal). Pain-related responses to injection of ATP into the hindpaw were measured essentially as described for rat[17]. The hindpaw lifting time was measured for a total of 4 min following injection of ATP. Thermal sensitivity was assessed using a radiant heat stimulus and tail immersion in a 52 °C water bath. Mechanical sensitivity was assessed using a set of calibrated von Frey filaments. For the formalin test, the hindpaw lifting and licking time was measured for a total of 30 min. Conscious mouse cystometry was performed essentially as described for rat[27]. Recovery following catheter implantation was for 7 days, and intravesical saline infusion was at a rate of 50 μl min^{-1}. For transurethral cystometry, bladder reflexes were assessed in urethane-anesthetized mice essentially as described for rat[28]. Each cystometrogram consisted of intravesical distension to a total volume of 0.3 ml, at a rate of 20 μl min^{-1}. Contractions greater than 10 cm of H_20 were taken as micturition contractions.

Generation of $P2X_3$ receptor-deficient mice and immunohistochemistry methods are described in Supplementary Information.

Received 19 May; accepted 17 August 2000.

1. Burnstock, G. P2X receptors in sensory neurons. *Br. J. Anaesth.* **84,** 476–488 (2000).
2. Chen, C. C. *et al.* A P2X purinoceptor expressed by a subset of sensory neurons. *Nature* **377,** 428–431 (1995).
3. Lewis, C. *et al.* Coexpression of $P2X_2$ and $P2X_3$ receptor subunits can account for ATP-gated currents in sensory neurons. *Nature* **377,** 432–435 (1995).
4. Bradbury, E. J., Burnstock, G. & McMahon, S. B. The expression of $P2X_3$ purinoreceptors in sensory neurons: Effects of axotomy and glial-derived neurotrophic factor. *Mol. Cell. Neurosci.* **12,** 256–268 (1998).
5. Ralevic, V. & Burnstock, G. Receptors for purines and pyrimidines. *Pharmacol. Rev.* **50,** 413–492 (1998).
6. Burnstock, G., McMahon, S. B., Humphrey, P. P. A. & Hamilton, S. G. in *Proceedings of the 9th World Congress on Pain* (eds Devor, M., Rowbotham, M. & Wiesenfeld-Hallin, Z.) 63–76 (IASP Press, Seattle, 2000).
7. Ferguson, D. R., Kennedy, I. & Burton, T. J. ATP is released from rabbit urinary bladder epithelial cells by hydrostatic pressure changes—a possible sensory mechanism? *J. Physiol. (Lond.)* **505,** 503–511 (1999).
8. Tsuda, M., Ueno, S. & Inoue, K. Evidence for the involvement of spinal endogenous ATP and P2X receptors in nociceptive responses caused by formalin and capsaicin in mice. *Br. J. Pharmacol.* **128,** 1497–1504 (1999).
9. Namasivayam, S., Eardley, I. & Morrison, J. F. B. Purinergic sensory neurotransmission in the urinary bladder: an *in vitro* study in the rat. *Br. J. Urol. Int.* **84,** 854–860 (1999).
10. Vulchanova, L. *et al.* $P2X_3$ is expressed by DRG neurons that terminate in inner lamina II. *Eur. J. Neurosci.* **10,** 3470–3478 (1998).
11. Robertson, S. J., Rae, M. G., Rowan, E. G. & Kennedy, C. Characterization of a P2X-purinoceptor in cultured neurones of the rat dorsal root ganglia. *Br. J. Pharmacol.* **118,** 951–956 (1996).
12. Rae, M. G., Rowan, E. G. & Kennedy, C. Pharmacological properties of $P2X_3$-receptors present in neurones of the rat dorsal root ganglia. *Br. J. Pharmacol.* **124,** 176–180 (1998).
13. Burgard, E. C. *et al.* P2X receptor-mediated ionic currents in dorsal root ganglion neurons. *J. Neurophysiol.* **82,** 1590–1598 (1999).
14. Ueno, S., Tsuda, M., Iwanaga, T. & Inoue, K. Cell type-specific ATP-activated responses in rat dorsal root ganglion neurons. *Br. J. Pharmacol.* **126,** 429–436 (1999).
15. Grubb, B. D. & Evans, R. J. Characterization of cultured dorsal root ganglion neuron P2X receptors. *Eur. J. Neurosci.* **11,** 149–154 (1999).
16. Thomas, S., Virginio, C., North, R. A. & Surprenant, A. The antagonist trinitrophenyl-ATP reveals co-existence of distinct P2X receptor channels in rat nodose neurones. *J. Physiol. (Lond.)* **509,** 411–417 (1998).
17. Hamilton, S. G., Wade, A. & McMahon, S. B. The effects of inflammation and inflammatory mediators on nociceptive behaviour induced by ATP analogues in the rat. *Br. J. Pharmacol.* **126,** 326–332 (1999).
18. Bleehan, T. & Keele, C. A. Observations on the algogenic actions of adenosine compounds on the human skin blister base preparation. *Pain* **3,** 367–377 (1977).
19. Hamilton, S. G., Warburton, J., Bhattacharjee, A., Ward, J. & McMahon, S. B. ATP in human skin elicits a dose related pain response which is potentiated under conditions of hyperalgesia. *Brain* **123,** 1238–1246 (2000).
20. Bland-Ward, P. A. & Humphrey, P. P. A. Acute nociception mediated by hindpaw P2X receptor activation in the rat. *Br. J. Pharmacol.* **122,** 365–371 (1997).
21. McMahon, S. B. in *Progress in Brain Research* Vol. 67 (eds Cervero, F. & Morrison, J. F. B.) 245–255 (Elsevier, Amsterdam, 1986).
22. de Groat, W. C. *et al.* Developmental and injury induced plasticity in the micturition reflex pathway. *Behav. Brain Res.* **92,** 127–140 (1998).
23. Elneil, S., Skepper, J. N., Kidd, E. J., Williamson, J. G. & Ferguson, D. R. The distribution of $P2X_1$ and $P2X_3$ receptors in the rat and human urinary bladder. *Neurourol. Urodyn.* **18,** 339–340 (1999).
24. Fitzgerald, J. M. & Krane, R. J. *The Bladder* (Churchill Livingstone; Edinburgh, London, Melbourne, New York and Tokyo; 1995).
25. Burnstock, G. Release of vasoactive substances from endothelial cells by shear stress and purinergic mechanosensory transduction. *J. Anat.* **194,** 335–342 (1999).
26. Zhong, Y., Dunn, P. M., Xiang, Z., Bo, X. & Burnstock, G. Pharmacological and molecular characterization of P2X receptors in rat pelvic ganglion neurons. *Br. J. Pharmacol.* **125,** 771–781 (1998).
27. Ishizuka, O., Mattiasson, A. & Andersson, K.-E. Role of spinal and peripheral alpha2 adrenoceptors in micturition in normal conscious rats. *J. Urol.* **156,** 1853–1857 (1996).
28. Dmitrieva, N., Shelton, D., Rice, A. S. & McMahon, S. B. The role of nerve growth factor in a model of visceral inflammation. *Neuroscience* **78,** 449–459 (1997).

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

Acknowledgements

We thank J. Muraski for microinjection and colony management; J. Thompson, S. Bingham and J. Sutton for behavioural tests; M. Bardini for immunohistochemistry; and K. Gregrow for necropsy and pathology.

Correspondence and requests for material should be addressed to D.A.C. (e-mail: debra.cockayne@roche.com).

Warm-coding deficits and aberrant inflammatory pain in mice lacking $P2X_3$ receptors

Veronika Souslova*, Paolo Cesare*, Yanning Ding*, Armen N. Akopian*, Louise Stanfa†, Rie Suzuki†, Katherine Carpenter†, Anthony Dickenson†, Susan Boyce‡, Ray Hill‡, Daniela Nebenius-Oosthuizen§, Andrew J.H. Smith§, Emma J. Kidd‖ & John N. Wood*

** Department of Biology and † Department of Pharmacology, University College London, London WC1E 6BT, UK*
‡ Merck Sharp and Dohme Research Labs, Terlings Park, Essex CM20 2QR, UK
§ Centre for Genome Research, Edinburgh University, Edinburgh EH9 3JQ, UK
‖ Welsh School of Pharmacy, Cardiff University, Cardiff CF1 3XF, UK

ATP activates damage-sensing neurons (nociceptors) and can evoke a sensation of pain[1]. The ATP receptor $P2X_3$ is selectively expressed by nociceptors[2,3] and is one of seven ATP-gated, cation-selective ion channels[4–6]. Here we demonstrate that ablation of the $P2X_3$ gene results in the loss of rapidly desensitizing ATP-gated cation currents in dorsal root ganglion neurons, and that the responses of nodose ganglion neurons to ATP show altered kinetics and pharmacology resulting from the loss of expression of $P2X_{2/3}$ heteromultimers. Null mutants have normal sensorimotor function. Behavioural responses to noxious mechanical and thermal stimuli are also normal, although formalin-induced

pain behaviour is reduced. In contrast, deletion of the $P2X_3$ receptor causes enhanced thermal hyperalgesia in chronic inflammation. Notably, although dorsal-horn neuronal responses to mechanical and noxious heat application are normal, $P2X_3$-null mice are unable to code the intensity of non-noxious 'warming' stimuli.

To examine the physiological role of the $P2X_3$ receptor in the absence of selective antagonists, null-mutant mice were generated on an inbred (C57Bl6) and outbred (MF1) genetic background[7,8] (see Supplementary Information). In most wild-type DRG neurons (87.5%, $n = 32$; Fig. 1a), ATP elicited a large inward membrane current (amplitude: 797 ± 122 pA, mean ± s.e.m., $n = 28$) that desensitized rapidly (72 ± 6% desensitization in 1 s). In $P2X_3^{-/-}$ neurons, ATP evoked an inward current in a smaller proportion of cells (56.7%, $n = 37$), the mean current was smaller (151 ± 52 pA; $n = 21$) and less desensitized (11 ± 2% desensitized after 1 s). In wild-type DRG neurons, 88.5% ($n = 26$) of the cells activated by ATP responded to a subsequent application of α,β-Me-ATP with a large (464 ± 85 pA; $n = 23$) and rapidly desensitizing current (87 ± 4%). In $P2X_3^{-/-}$ DRG, of the 20 neurons responding to ATP only 1

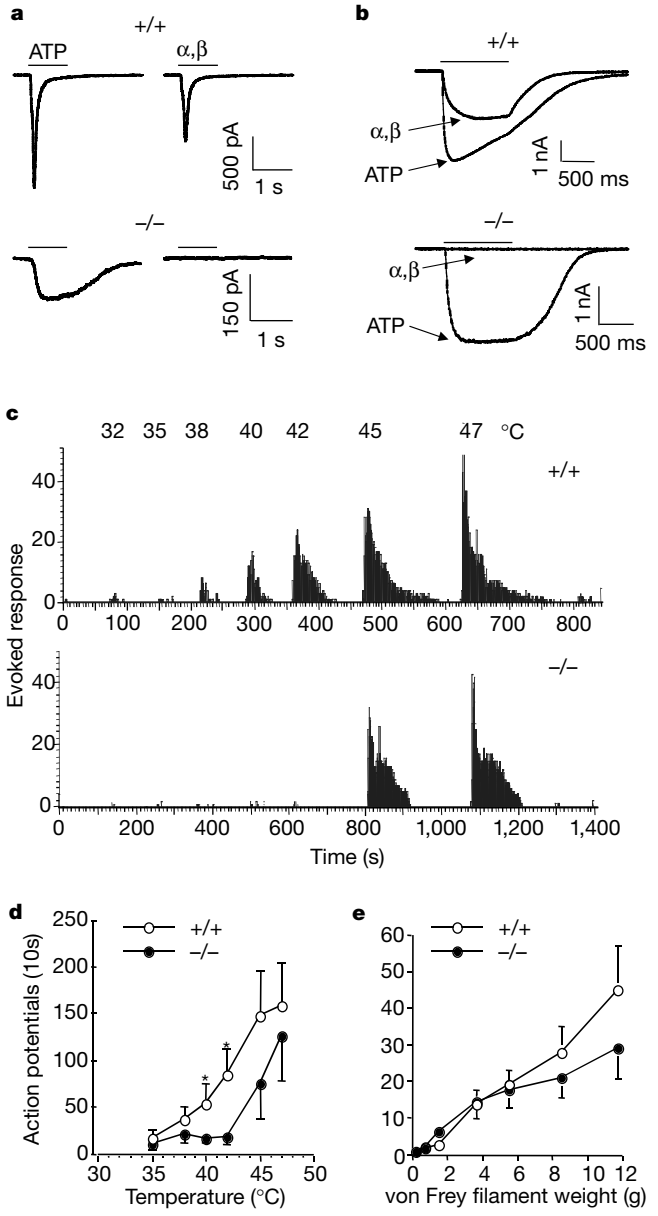

Figure 1 Physiological effects of deleting the $P2X_3$ receptor. **a**, **b**, The responses to a 1-s application of 10 μM ATP in wild-type (+/+) and $P2X_3$-null (–/–) neurons. Cells responding to ATP were subsequently treated with 10 μM α,β-Me-ATP (α,β) **a**, DRG neurons; **b**, nodose ganglion neurons. **c**–**e**, Responses of dorsal horn neurons to peripheral stimuli in wild-type and $P2X_3^{-/-}$ mice. **c**, Individual examples of neuronal responses to graded thermal stimuli recorded in wild-type and $P2X_3^{-/-}$ mice. **d**, Thermal stimulus-response curves recorded in wild-type ($n = 12$) and $P2X_3^{-/-}$ ($n = 17$) mice. Responses are presented as mean number of action potentials ± s.e.m. Asterisk, $P \leq 0.05$. **e**, Responses to mechanical stimuli elicited with von Frey filaments show no significant differences between wild-type and $P2X_3^{-/-}$ mice. Electrical thresholds and responses to suprathreshold electrical stimulation of A and C fibre primary afferents are shown in the Supplementary Information.

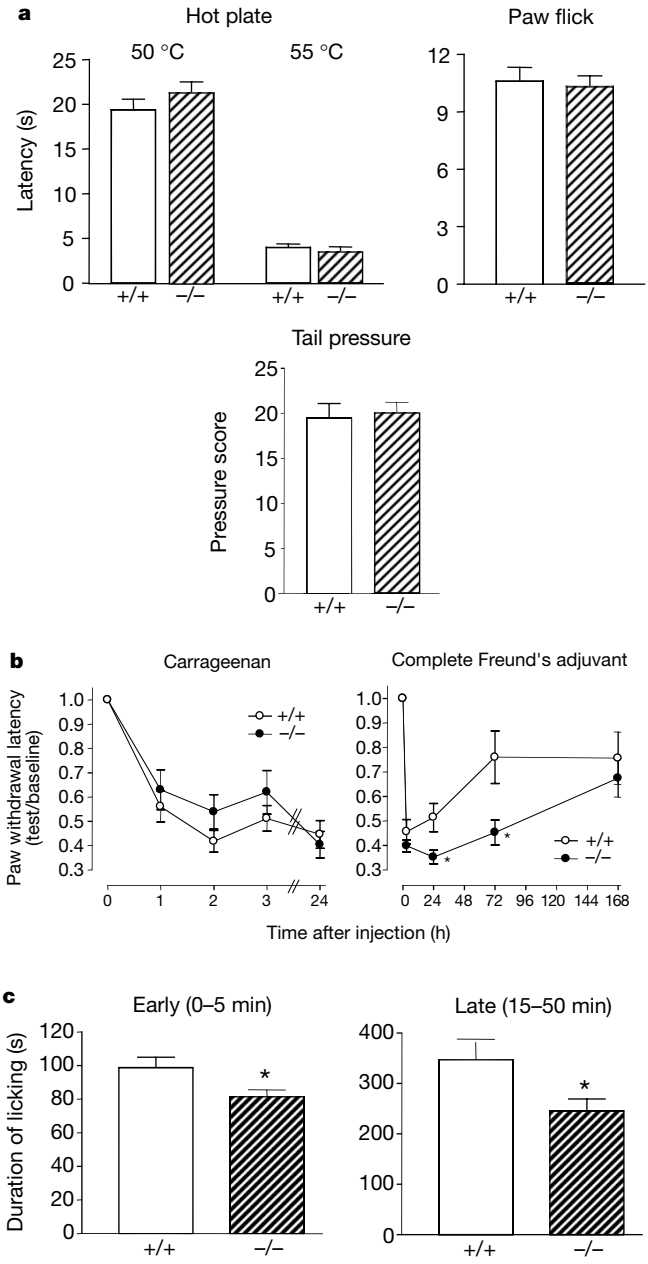

Figure 2 Behavioural effects of deleting the $P2X_3$ receptor. Sensorimotor function was unaffected (see Supplementary Information). Responses to acute noxious stimuli were normal (**a**, $n \geq 10$). Carrageenan-induced thermal hyperalgesia developed to similar levels in wild-type and $P2X_3^{-/-}$ mice, but longer term thermal hyperalgesia induced by complete Freund's adjuvant was potentiated ($P < 0.05$, $n = 12$) in null mutants on two genetic backgrounds (**b**) and formalin-induced pain behaviour was attenuated in both phases (**c**, $P < 0.05$, $n = 10$). Asterisk, $P < 0.05$.

responded to α,β-Me-ATP, with a small (22 pA) and non-desensitizing current. This indicates that homomeric $P2X_3$ receptors are the major P2X subtype in mouse DRG neurons.

Patch-clamp recordings of nodose ganglion neurons in normal and $P2X_3$-null mice revealed very large ATP-evoked non-desensitizing currents. In nodose ganglion neurons (Fig. 1b), peak current had mean values of 3,365 ± 428 pA in wild-type (n = 27) and 2,250 ± 684 (n = 23) in $P2X_3^{-/-}$ mice. In wild-type mice, desensitization was 25.9 ± 4.4% (n = 27), but $P2X_3^{-/-}$ neurons desensitized more slowly (7.6 ± 4.7%; n = 23). In the wild type, α,β-Me-ATP application induced a mean current of 1,705 ± 255 pA (n = 27). In the null mutant, α,β-Me-ATP application failed to evoke any response (n = 23). Almost half of the currents evoked by ATP are due to $P2X_{2/3}$ heteromultimers; $P2X_2$ receptors, as well as other P2X subtypes, must account for the rest of the current. The lack of $P2X_1$ receptors is confirmed by the fact that in $P2X_3^{-/-}$ nodose ganglion neurons, α,β-Me-ATP was inactive.

In vivo recordings of dorsal horn neurons were used to examine the consequences of ablating $P2X_3$ on peripheral input into the spinal cord[9]. There was no significant change in the mechanically evoked responses in $P2X_3$-null mice, although the responses to the greatest pressure (8.5 and 11.7 g) were slightly reduced. Both threshold and suprathreshold responses of dorsal horn neurons evoked after electrical activation of A and C fibre primary afferents were unaltered in the $P2X_3$-null mice (see Supplementary Information). This indicates that centrally located $P2X_3$ receptors are not significantly involved in spinal nociceptive processing under acute conditions, consistent with the lack of effect of spinally administered P2X antagonists on nociceptive neuronal responses and behaviour measured in rats[10–12].

Differences were seen between the $P2X_3$-null and wild-type mice in their response to thermal stimuli, although no change was observed in the neuronal responses to mechanical stimuli (Fig. 1). Application of a range of thermal stimuli to the hindpaw of wild-type mice, evoked a graded neuronal response in the dorsal horn (Fig. 1c–d). In contrast, in $P2X_3$-null mice very little neuronal activity was evoked until noxious temperatures were reached[13]. This is in agreement with a 'normal' behavioural response to temperatures within the noxious range, consistent with behavioural studies showing no effect of pyridoxalphosphate-6-azophenyl-2′,4′-disulphonic acid (PPADS) in acute tests of thermal nociception[12], but indicates a deficit in the ability of the null-mutant mice to code the intensity of non-noxious 'warming' stimuli.

Behavioural tests showed that responses to acute mechanical and thermal stimuli were normal, precluding a major role for $P2X_3$ in transducing acute noxious stimuli (Fig. 2, and see Supplementary Information). However formalin-induced pain behaviour was significantly reduced in both phases, consistent with data obtained using PPADS in mice[14]. Most strikingly, the development of thermal hyperalgesia in response to chronic inflammation induced with complete Freund's adjuvant but not short-term inflammatory stimuli such as carrageenan (Fig. 2b) or capsaicin (data not shown) was markedly potentiated.

These data indicate that $P2X_3$ receptors within the dorsal horn have a significant regulatory role in persistent inflammatory pain. In spinal cord slices, ATP has been shown to potentiate glutamate-induced and synaptically evoked currents, as well as causing delayed depression of synaptic currents, probably owing to the actions of adenosine[11,12,15,16]. The unexpected hyperalgesia present in $P2X_3$ receptor knockouts may therefore be due to the actions of ATP on other less rapidly desensitizing P2 receptors within the dorsal horn, or may result from the loss of a regulatory function of $P2X_3$ receptors on other modulatory systems. Thus, there is no major role for the $P2X_3$ receptor in noxious mechanosensation and acute pain responses[17,18]. However, these studies demonstrate a role for ATP acting through $P2X_3$ in inflammatory pain processing, as well as an unsuspected role for $P2X_3$ receptors in the coding of innocuous warmth. □

Methods

Gene targeting

A targeting construct in which exons 2–7 of the $P2X_3$ gene were deleted was constructed and knockout mice were generated and analysed by standard methods (see Supplementary Information)[7,8,19].

Electrophysiology

DRG or nodose ganglia from adult mice (6–9 weeks old) were removed and neurons were cultured[20]. Recordings were made 1–3 days after plating, using patch-clamp electrodes of tip diameter 0.7 μm and an Axopatch 200B amplifier (Axon Instruments) as described[20]. Drugs were used at 10 μM ATP (magnesium salt), 10 μM α,β-Me-ATP (lithium salt). Cells between 15- and 35-μm diameter were randomly chosen for experiments[20].

Dorsal horn recordings

Extracellular recordings were made *in vivo* as described[10,21]. The neuronal responses evoked by transcutaneous electrical (train of 16 stimuli, 0.5 Hz, 2-ms wide pulse) and thermal (water jet applied for 10 s, 32–47 °C) stimuli applied to the peripheral receptive field located on the ipsilateral hind paw were characterized. Mechanical stimulation (von Frey filaments 0.166 g to 11.7 g) in the innocuous and noxious ranges was also applied to peripheral receptive fields.

Behavioural studies

Mice were examined for spinal reflexes and motor skills, as well as responses to acute and inflammatory pain, as described in detail[21].

Received 19 May; accepted 17 August 2000.

1. Bleehen T. & Keele, C. A. Observations on the algogenic actions of adenosine compounds on the human blister base preparation. *Pain* **3,** 367–377 (1977).
2. Chen, C. *et al.* A P2X purinoceptor expressed by a subset of sensory neurons. *Nature* **377,** 428–431 (1995).
3. Lewis, C. *et al.* Co-expression of P2X2 and P2X3 receptor subunits can account for ATP-gated currents in sensory neurons. *Nature* **377,** 432–435 (1995).
4. MacKenzie, A. B., Surprenant, A. & North, R. A. Functional and molecular diversity of purinergic ion channel receptors. *Ann. NY Acad. Sci.* **30,** 716–729 (1999).
5. Grubb, B. D. & Evans, R. J. Characterization of cultured dorsal root ganglion neuron P2X receptors. *Eur. J. Neurosci.* **11,** 149–154 (1999).
6. Ueno, S., Tsuda, M., Iwanaga, T. & Inoue, K. Cell type-specific ATP-activated responses in rat dorsal root ganglion neurons. *Br. J. Pharmcol.* **126,** 429–436 (1999).
7. Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51,** 503–512 (1987).
8. Mogil, J. S. *et al.* Heritability of nociception I: responses of 11 inbred mouse strains on 12 measures of nociception. *Pain* **80,** 67–82 (1999).
9. Chapman, V., Ng, J. & Dickenson, A. H. A novel spinal action of mexiletine in spinal somatosensory transmission of nerve injured rats. *Pain* **77,** 289–296 (1998).
10. Stanfa, L. C., Kontinen, V. K. & Dickenson, A. H. Effects of spinally administered P2X receptor agonists and antagonists on the responses of dorsal horn neurones recorded in normal, carrageenan-inflamed and neuropathic rats. *Br. J. Pharmacol.* **129,** 351–359 (2000).
11. Li, J. & Perl, E. R. ATP modulation of synaptic transmission in the spinal substantia gelatinosa. *J. Neurosci.* **15,** 3357–3365 (1995).
12. Li, P., Calejesan, A. A. & Zhua, M. ATP P2X receptors and sensory synaptic transmission between primary afferent fibres and spinal dorsal horn neurons in rats. *J. Neurophysiol.* **80,** 3356–3360 (1998).
13. Cesare, P. & McNaughton, P. Peripheral pain mechanisms. *Curr. Opin. Neurobiol.* **7,** 493–499 (1997).
14. Tsuda, M., Ueno, S. & Inoue, K. Evidence for the involvement of spinal endogenous ATP and P2X receptors in nociceptive responses caused by formalin and capsaicin in mice. *Br. J. Pharmacol.* **128,** 1497–1504 (1999).
15. Gu, J. G., Bardoni, R., Magherini, P. C. & MacDermott, A. B. Effects of the P2-purinoceptor antagonists suramin and PPADS on glutamatergic synaptic transmission in rat dorsal horn neurons of the spinal cord. *Neurosci. Lett.* **253,** 167–170 (1998).
16. Gu, J. G. & MacDermott, A. B. Activation of ATP P2X receptors elicits glutamate release from sensory neuron synapses. *Nature* **389,** 749–753 (1997).
17. Bland-Ward, P. A. & Humphrey, P. P. A. Acute nociception mediated by hindpaw P2X receptor activation in the rat. *Br. J. Pharmacol.* **122,** 365–371 (1997).
18. Nakamura, F. & Strittmatter, S. M. P2Y1 purinergic receptors in sensory neurons: contribution to touch-induced impulse generation. *Proc. Natl Acad. Sc.i USA* **17,** 10465–10470 (1996).
19. Kidd, E. J., Miller, K. J., Sansum, A. J. & Humphrey, P. P. Evidence for P2X3 receptors in the developing rat brain. *Neuroscience* **87,** 533–539 (1998).
20. Cesare, P., Souslova, V., Akopian, A., Rufian, O. & Wood, J. N. Fast inactivatating responses to ATP are lost in dorsal root ganglion neurons isolated from P2X3 knock-out mice. *Proc. Physiol. Soc.* **523P,** (2000).
21. Akopian, A. N. *et al.* The tetrodotoxin-resistant sodium channel SNS plays a specialised role in pain pathways. *Nature Neurosci.* **2,** 541–548 (1999).

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 thank the Wellcome Trust, EC Biomed and the MRC for support and H. Thompson for help with genotyping mice.

Correspondence and requests for materials should be addressed to J.N.W. (e-mail: J.Wood@ucl.ac.uk).

indirect immunofluorescence using an anti-cytochrome *c* antibody (65972A, PharMingen, 1:1,000), or by cell fractionation into cytosol and membrane components and immunoblotting (cytochrome *c* antibody 7H8.2C12, PharMingen, 1:1,000 dilution).

Received 29 June; accepted 10 November 2000.

1. Chin, L., Merlino, G. & DePinho, R. A. Malignant melanoma: modern black plague and genetic black box. *Genes Dev.* **12,** 3467–3481 (1998).
2. Serrone, L. & Hersey, P. The chemoresistance of human malignant melanoma: an update. *Melanoma Res.* **9,** 51–58 (1999).
3. Soengas, M. S. *et al.* Apaf-1 and caspase-9 in p53-dependent apoptosis and tumor inhibition. *Science* **284,** 156–159 (1999).
4. Schmitt, C. A. & Lowe, S. W. Apoptosis and therapy. *J. Pathol.* **187,** 127–137 (1999).
5. Lowe, S. W., Ruley, H. E., Jacks, T. & Housman, D. E. p53-dependent apoptosis modulates the cytotoxicity of anticancer agents. *Cell* **74,** 957–967 (1993).
6. Hakem, R. *et al.* Differential requirement for caspase 9 in apoptotic pathways *in vivo*. *Cell* **94,** 339–352 (1998).
7. Fearnhead, H. O. *et al.* Oncogene-dependent apoptosis is mediated by caspase-9. *Proc. Natl Acad. Sci. USA* **95,** 13664–13669 (1998).
8. Baylin, S. B. & Herman, J. G. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet.* **16,** 168–174 (2000).
9. Robertson, K. D. & Jones, P. A. DNA methylation: past, present and future directions. *Carcinogenesis* **21,** 461–467 (2000).
10. Bird, A. P. & Wolffe, A. P. Methylation-induced repression—belts, braces, and chromatin. *Cell* **99,** 451–454 (1999).
11. Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405,** 486–489 (2000).
12. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405,** 482–485 (2000).
13. Lowe, S. W. & Ruley, H. E. Stabilization of the p53 tumor suppressor is induced by adenovirus 5 E1A and accompanies apoptosis. *Genes Dev.* **7,** 535–545 (1993).
14. Green, D. R. & Reed, J. C. Mitochondria and apoptosis. *Science* **281,** 1309–1312 (1998).
15. Rodriguez, J. & Lazebnik, Y. Caspase-9 and APAF-1 form an active holoenzyme. *Genes Dev.* **13,** 3179–3184 (1999).
16. Zou, H., Li, Y., Liu, X. & Wang, X. An APAF-1 cytochrome c multimeric complex is a functional apoptosome that activates procaspase-9. *J. Biol. Chem.* **274,** 11549–11556 (1999).
17. Hu, Y., Benedict, M. A., Ding, L. & Nunez, G. Role of cytochrome c and dATP/ATP hydrolysis in Apaf-1-mediated caspase- 9 activation and apoptosis. *EMBO J.* **18,** 3586–3595 (1999).
18. Teitz, T. *et al.* Caspase 8 is deleted or silenced preferentially in childhood neuroblastomas with amplification of MYCN. *Nature Med.* **6,** 529–535 (2000).
19. Yamamoto, H., Gil, J., Schwartz, S. Jr & Perucho, M. Frameshift mutations in Fas, Apaf-1, and Bcl-10 in gastro-intestinal cancer of the microsatellite mutator phenotype. *Cell Death Differ.* **7,** 238–239 (2000).
20. Schmitt, C. A., McCurrach, M. E., de Stanchina, E., Wallace-Brodeur, R. R. & Lowe, S. W. INK4a/ARF mutations accelerate lymphomagenesis and promote chemoresistance by disabling p53. *Genes Dev.* **13,** 2670–2677 (1999).
21. Chin, L. *et al.* Cooperative effects of INK4a and ras in melanoma susceptibility *in vivo*. *Genes Dev.* **11,** 2822–2834 (1997).
22. Walker, G. J. *et al.* Virtually 100% of melanoma cell lines harbor alterations at the DNA level within CDKN2A, CDKN2B, or one of their downstream targets. *Genes Chromosom. Cancer* **22,** 157–163 (1998).
23. Grossman, D., McNiff, J. M., Li, F. & Altieri, D. C. Expression and targeting of the apoptosis inhibitor, survivin, in human melanoma. *J. Invest. Dermatol.* **113,** 1076–1081 (1999).
24. Irmler, M. *et al.* Inhibition of death receptor signals by cellular FLIP. *Nature* **388,** 190–195 (1997).
25. Selzer, E. *et al.* Expression of Bcl-2 family members in human melanocytes, in melanoma metastases and in melanoma cell lines. *Melanoma Res.* **8,** 197–203 (1998).
26. Mora, J. *et al.* Clinical categories of neuroblastoma are associated with different patterns of loss of heterozygosity on chromosome Arm 1p. *J. Mol. Diag.* **2,** 1–9 (2000).
27. Woo, M. *et al.* Essential contribution of caspase 3/CPP32 to apoptosis and its associated nuclear changes. *Genes Dev.* **12,** 806–819 (1998).
28. Marcellus, R. C., Teodoro, J. G., Charbonneau, R., Shore, G. C. & Branton, P. E. Expression of p53 in Saos-2 osteosarcoma cells induces apoptosis which can be inhibited by Bcl-2 or the adenovirus E1B-55 kDa protein. *Cell Growth Differ.* **7,** 1643–1650 (1996).
29. Capodieci, P., Magi-Galluzzi, C., Moreira, G. Jr Zeheb, R. & Loda, M. Automated *in situ* hybridization: diagnostic and research applications. *Diagn. Mol. Pathol.* **7,** 69–75 (1998).

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 thank B. Stillman and N. K. Cheung for support; M. McCurrach, W. Jiang, L. Chen and M. Dudas for technical assistance; M. Zhang, R. Davuluri and I. Ioshikhes for analysis of Apaf-1 CpG islands and assembly of Apaf-1 genomic sequence; G. Núñez for Apaf-1_{XL} pcDNA construct; E. Querido, P. E. Branton and F. L. Graham for Ad-p53 and Ad-LacZ; and A. Houghton and R. Camalier for melanoma cell lines. We also thank G. Hannon, C. Schmitt, C. Hazan, K. Pohar, T. Tiongson and the Cold Spring Harbor Genome Center for help and advice. Y.L. is a Pew Scholar and S.W.L. is a Rita Allen Scholar. This work was supported by a long-term Postdoctoral Fellowship from the Human Frontiers in Science Program and a Special Fellowship from the Leukemia and Lymphoma Society (M.S.S), a K08 grant (D.P.), an ASCO Young Investigator Award (J.M.), grants from the Seraph Foundation (YL), and from the NIH and the NCI (C.C.C. and S.W.L.).

Correspondence and requests for materials should be addressed to S.W.L. (e-mail: lowe@cshl.org).

The protein–protein interaction map of *Helicobacter pylori*

Jean-Christophe Rain*, Luc Selig*, Hilde De Reuse†, Véronique Battaglia*, Céline Reverdy*, Stéphane Simon*, Gerlinde Lenzen*, Fabien Petel*, Jérôme Wojcik*, Vincent Schächter*, Y. Chemama*, Agnès Labigne† & Pierre Legrain*

** Hybrigenics SA, 180 avenue Daumesnil, Paris 75012, France*
† Unité de Pathogénie Bactérienne des Muqueuses, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

With the availability of complete DNA sequences for many prokaryotic and eukaryotic genomes, and soon for the human genome itself, it is important to develop reliable proteome-wide approaches for a better understanding of protein function[1]. As elementary constituents of cellular protein complexes and pathways, protein–protein interactions are key determinants of protein function. Here we have built a large-scale protein–protein interaction map of the human gastric pathogen *Helicobacter pylori*. We have used a high-throughput strategy of the yeast two-hybrid assay to screen 261 *H. pylori* proteins against a highly complex library of genome-encoded polypeptides[2]. Over 1,200 interactions were identified between *H. pylori* proteins, connecting 46.6% of the proteome. The determination of a reliability score for every single protein–protein interaction and the identification of the actual interacting domains permitted the assignment of unannotated proteins to biological pathways.

During the past few years, interaction maps have been proposed for viral[3–5] and eukaryotic (*Saccharomyces cerevisiae*[2,6,7] and *Caenorhabditis elegans*[8], for example) genomes. Here we describe the first procaryotic interaction map, for which the strategy used (Fig. 1) is also a variation of the two-hybrid assay[2]. It differs considerably by both the type of results generated and the throughput of the experimental procedures. We constructed a library of random genomic fragments of the *H. pylori* strain 26695 that had been previously sequenced[9]. A high complexity library was first obtained in *Escherichia coli* (over ten million clones). Ninety-seven per cent of the plasmids contained a single genomic insert (mean size 1,000 ± 550 nucleotides). This library was then introduced into yeast by transformation. Two million independent yeast colonies were collected, pooled and stored at −80 °C as equivalent aliquot fractions of the same library.

In parallel, a large set of bait plasmids was constructed in a bait vector designed to decrease the level of transcriptional autoactivation[2]. Bait constructs were specifically adapted for interaction screens, yielding 'validated baits'; for example, hydrophobic putative trans-membrane domains were discarded to avoid any non-nuclear localization of the bait protein in the yeast cell. In some cases, specific open reading frame (ORF) domains were selected for the bait design. For every single bait construct, we performed a preliminary small-scale screening experiment that determines the selective pressure (that is, modifying the selective medium, see Methods) to be applied to obtain a small number of independent positive clones per million interactions tested (usually less than 10). All positive colonies were then picked, and prey fragments were individually identified by sequence analysis and comparison with the genomic database through a dedicated integrated laboratory production management system (Fig. 1).

Protein–protein interaction maps are built on experimental data that ideally yield a heuristic value for each connection. Our procedure involves several steps of processing of raw two-hybrid results (Fig. 1). First, positive prey fragments are clustered into families of overlapping fragments. The common sequence shared by

these fragments is referred to as the selected interacting domain (SID). Second, SIDs that do not code for part of an *H. pylori* ORF are discarded. Third, for every remaining SID, a PIM biological score (PBS) is computed. The PBS is based on a statistical model of the competition for bait-binding between fragments. It is computed like a classical expected value (*E* value), and ranges from 0 (specific interaction) to 1 (probable artefact). For practical use, the scores were divided into four categories, from A (score very close to 0) to D (close to 1). A fifth category, E, was added to distinguish interactions involving only highly connected prey domains (SIDs found as prey with frequency greater than a fixed threshold). These are most probably two-hybrid artefacts. Although they may have some biological significance, they add little specific information to the interaction map. It should be emphasized that domains, rather than whole proteins, are tagged. For example, the carboxy-terminal region of protein HP0705/UvrA was found as a prey in 16 different interactions that were scored in the E category, but the same protein was selected through another domain as a specific interacting prey. Because global connectivity is taken into account, the PBS is computed incrementally over the whole PIM and its discriminatory power increases as screening results accumulate.

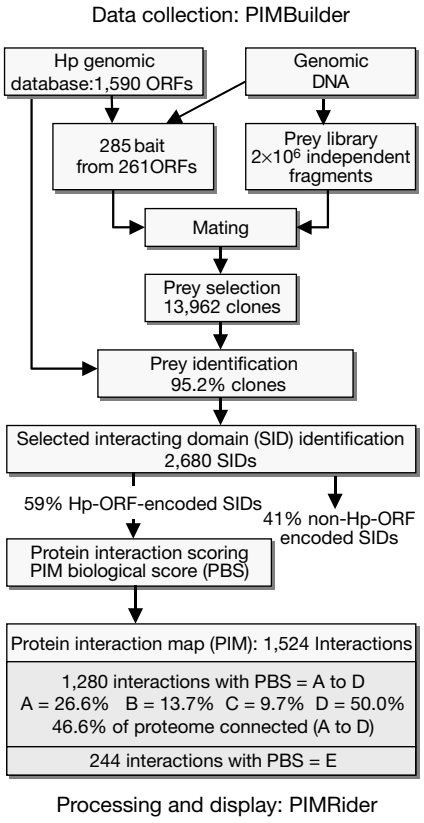

Figure 1 Outline of the strategy for building an *H. pylori* (Hp) proteome-wide interaction map. A production database (the PIMBuilder) was built that contains information related to the genomic sequence of *H. pylori*, which codes for 1,590 putative proteins or ORFs. It was populated with raw data from screening experiments. The PIMBuilder tracks all biotechnological or bioinformatics operations performed during the production processes, stores information about all biological objects produced during experiments, and interfaces with robots and bioinformatics modules. It also implements the procedure used to construct PIMs from raw experimental data. After identification of almost all positive clones, overlapping prey fragments were clustered into families to define SIDs. Those families that had no biological coding capability (antisense or intergenic region, out of frame fragments occurring in a single frame) were discarded. A PIM biological score (PBS; see Methods) was then calculated for *H. pylori* ORF-encoded SIDs. Interactions were grouped into categories A to D (from high to low heuristic values). The global connectivity of the PIM was also analysed to detect highly connected prey polypeptides. Those interactions were grouped in the E category. Processing of data and visualization of interactions were performed by an in-house bioinformatic platform (PIMRider).

We carried out 285 screens on 261 *H. pylori* bait ORFs chosen as follows: first, a core set of 50 proteins known to be involved in complexes and/or in pathogenicity was used to validate the approach; and second, 211 baits were then picked randomly, with a slight bias toward regions of the proteome that were still unexplored. Positive colonies (13,962) were selected (Table 1), and more than 95% of them were identified by sequencing the prey insert. From these prey fragments, 2,680 independent SIDs were defined, out of which 1,100 fell into non *H. pylori* ORF coding regions and were discarded (Fig. 1). Thirty-one SIDs were classified in the E category. The remaining SIDs define 1,280 interactions including 62 homo-oligomeric interactions. In total, 46.6% *H. pylori* proteins of the proteome were connected, which corresponds to an average connectivity of 3.36 partners per connected protein (without counting homodimeric connections). Only 14 screens out of 285 yielded no positive clones or interactions with nonsignificant score values, illustrating the fact that our technology reduces the rate of false negatives.

Given that little information is available on protein interactions in *H. pylori*, we used a data set of known interactions in *E. coli* to validate further our experimental strategy, and to evaluate the correlation between the PBS and the actual biological significance of interactions. For each *H. pylori* protein present in the interaction map, significant *E. coli* orthologues (FastA score < 0.01) were selected and their annotations in the SwissProt database (release 38)[10] were verified manually for known interactions shown by various biochemical means. The resulting *E. coli* interaction list was compared with interactions found in *H. pylori* (Fig. 2). Among these *E. coli* interacting pairs, 53% of the homodimers and 67% of the heterodimers were found. Among heterodimers, five out of six negative pairs were tested only in one direction, suggesting that performing the reciprocal screens would decrease this number. Most interactions that were described for orthologous proteins in *E. coli* fell into the high-scoring interaction category (A) according to our PBS calculation (7/10 homodimers and 9/12 heterodimers) confirming the heuristic value of the classification. The interaction map was also analysed according to the classification of *H. pylori* proteins into 14 functional categories previously proposed[9]. For 10 categories out of 14, more intra-category protein–protein interactions were observed than expected from a random theoretical distribution (Z-scores ranging from 2 to 50; in all cases, $P < 0.05$), suggesting the existence of a significant correlation between functional grouping and detection of interactions.

To display and analyse the interaction data, we developed a software platform composed of a database, a web-based graphical interface layer and various query and analysis tools (the PIMRider, Fig. 3; see also http://pim.hybrigenics.com). Starting from a gene name (or an ORF name), the PIMRider draws an automatic layout of the neighbourhood of this protein in the protein interaction map (Fig. 3a). Paths connecting two proteins can also be queried.

Table 1 General features of the *H. pylori* interaction screens

Screening data	Total or mean value	Comments
Number of screens	285	261 different *H. pylori* ORFs
Combination of bait/prey polypeptides assayed	5.6×10^9	
Number of prey polypeptides assayed per bait	2×10^7	
Number of selected colonies per bait (1st reporter)	916	From 0 to 15,000
Number of selected clones per bait (2 reporters)	49	From 0 to 139
Number of selected colonies per million pairs tested	3.8	From 0 to 25
Number of selected clones	13,962	
Number of sequenced clones	13,296	95.2 %

Connections are displayed with their PBS scores, and can be filtered according to score categories. A graphical summary of information describing all interaction domains within a given protein can be displayed (Fig. 3b). Raw data on every interaction can be retrieved. Finally, the PIMRider supplies a description of each gene with functional and genomic information, and includes links to significant bibliographic references and to relevant external databases, such as PyloriGene (http://genolist.pasteur.fr/PyloriGene/). PyloriGene is a manually annotated database of the two *H. pylori* sequenced genomes[9,11], that integrates all publicly available information on genes and proteins and has been elaborated with a structure similar to that of the *Bacillus subtilis* SubtiList database[12].

Exploring the protein–protein interaction map reveals biological pathways and allows prediction of protein function. A first example concerns chemotaxis (see Fig. 3a). The *H. pylori* genome reveals three homologues of *E. coli* proteins that are involved in the chemotactic pathway (CheA, CheW and CheY) and proteins such as TlpA similar to chemotaxis receptors (MCPs). CheA was found to interact with CheY and CheW, and distinct interacting domains were precisely identified (Fig. 3b). The domain of CheA which binds to CheY precisely overlaps with the interacting domain assigned by a structural study in *E. coli*[13]. The TlpA-binding site for CheW was localized in a domain known, in *E. coli*, to be methylated and implicated in the transduction of the chemotactic signal.

The urease complex was also examined. Urease activity is essential for *H. pylori* pathogenicity and its synthesis requires two structural subunits, UreA and UreB, and the product of four accessory genes: *ureE, ureF, ureG, ureH*[14]. Complexes between accessory proteins and their role in nickel incorporation at the urease active site have been described for orthologues, but little information is available for *H. pylori* (for review, see ref. 15). The protein interaction map revealed the connection between UreA and UreB, and one of the two expected homo-oligomeric interactions of structural subunits (UreA); the UreB homodimer could not be detected. A connection between accessory proteins and the structural subunits occurs via UreH and UreA, which is consistent with the presumed chaperone role of UreH[15]. A new structural link was found between UreG and UreE. The UreF and UreH proteins were connected, but no connection between UreG and UreF or UreH was detected. In addition to the accessory proteins, the urease operon codes for an inner-membrane protein, UreI, essential for resistance to acidity[16] and recently described as a H^+-gated urea channel[17]. The third cytoplasmic domain of this protein reveals a potential interaction with the ExbD protein which is involved in transmission of PMF (proton motor force) energy to outer-membrane receptors.

Combination of genomic and proteomic data also permits function prediction. The *H. pylori* proteome contains a homologue of the *E. coli* HolB protein. In *E. coli,* this protein interacts with HolA to form part of the DNA polymerase core[18]. We found one high-scoring interaction between *H. pylori* HolB and an uncharacterized polypeptide, HP1247. A pairwise alignment between *E. coli* HolA and HP1247 highlighted structural homology (Fig. 4) not found by previous sequence analysis (Fig. 2). We thus assign the HolA function to the *H. pylori* HP1247 protein. The organization of bacterial genomes into operons suggests a functional relationship between the corresponding gene products that can be directly compared with our protein interaction map. Indeed, we found interactions between proteins that were likely to be expressed from a single operon (Table 2). Among these, we detected interactions between proteins known to interact in *H. pylori* (ScoA–ScoB) or in other organisms (RpsR–RpsF, MoaE–MoaD, FtsA–FtsZ), and between polypeptides involved in the same enzymatic activity and not yet described as interacting (HypE–HypF).

Selected interacting domains can also be analysed in terms of protein structure. The prokaryotic RNA polymerase, composed of a core enzyme ($\alpha_2\beta\beta'$) associated with a σ-factor, is one of the best studied multisubunit enzymes. In *H. pylori*, the β- and β′-subunits

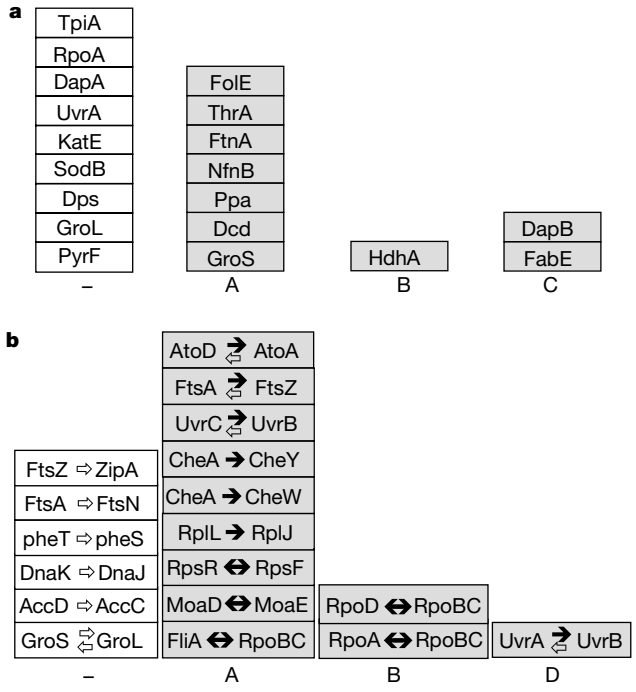

Figure 2 Sets of *E. coli* interaction data for which *H. pylori* orthologous proteins were identified and assayed in interaction screens. **a**, Homodimers; **b**, heterodimers. Names of *E. coli* proteins are boxed. Unidentified interactions between *H. pylori* orthologues are scored as '–' and shown in white boxes. Identified *H. pylori* interactions are indicated in grey boxes with their PBS score category (A, B, C or D). When several orthologues were identified, only the best scoring homologue was considered. For heterodimers, arrows indicate the direction of the screen (bait to prey) that was performed. The black (or white) colour of the arrow indicates a positive (or negative) result in the interaction screen.

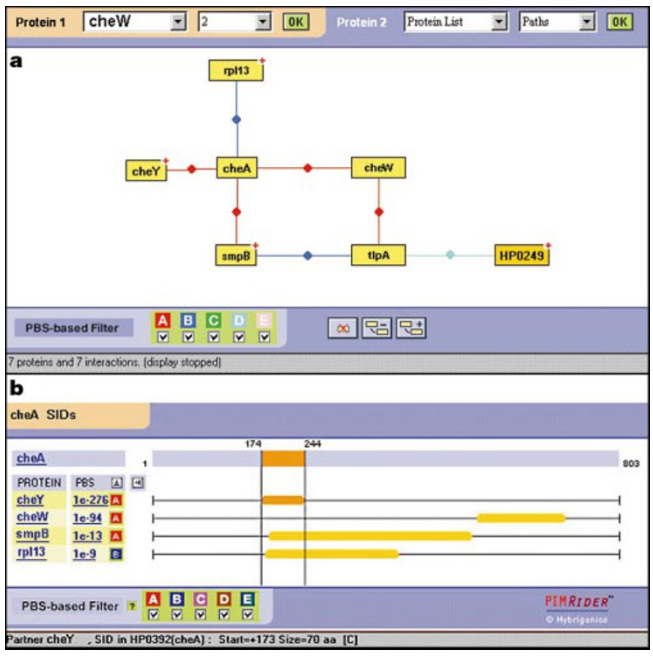

Figure 3 PIMRider screen shots. **a**, The PIMViewer displays a portion of the protein interaction map around the CheA protein. Links between proteins identify connections with their colour-coded PBS score; **b**, the MultiSID Viewer exhibits the various interacting domains in the CheA protein (for details see http://pim.hybrigenics.com).

usually found in other bacteria are fused into a single polypeptide (RpoB). One of the two alternative σ-factors present in *H. pylori*, (HP1032), is similar to the *E. coli* FliA protein (sigma 28) necessary for transcription of genes involved in flagellar biosynthesis[19]. We identified a precise region of RpoB that interacts with the *H. pylori* FliA. This selected interacting domain (residues 841–959) maps exactly to a structural domain called the flexible flap[20]. The RpoB-interacting domain of FliA falls in the regions 3.2–4.2 of this σ-factor (residues 175–255). Biochemical studies suggest an interaction between the flap domain of RpoB and region 4 of sigma factors[20]. Our experiments thus characterize this interaction and support the role of the flap domain and σ-factors in the transition from an open complex to a processive elongation complex[21].

Our work provides a way to characterize proteome-wide protein interaction maps. Our results identify complexes that have been shown or postulated to exist in other organisms, such as in *E. coli*. They also complement sequence information about homologous proteins and operon prediction from the location of genes on the chromosome. Finally, identified interacting domains could be mapped on three-dimensional structures of proteins. As a whole they lead to the assignment of a functional role for many as yet uncharacterized proteins and provide tools, such as interacting domains, for further biological experiments. Our technology was designed to specifically address the main known causes for false-negative and false-positive results in two-hybrid assays. It is now clear that interactions are not necessarily detected positive by two-hybrid assays when used in reciprocal directions (Fig. 2; see also refs 4, 8, 22). Parallel screening against highly complex libraries of fragments increases the number of 'two-hybrid-capable' candidates and reduces the rate of false negatives that arise with the classical two-hybrid matrix approach (that is, pair wise testing of a collection of proteins[6,7]). Concerning the false positives, the specific design of selection procedures that permit a strong selectivity for all baits and the statistical analysis made possible by the experimental procedure (that is, reproducible exhaustive screening of fragment libraries) allows us to detect and tag nonspecific partners through a global scoring scheme.

Ultimate validation of biological significance should come from additional biological information, but comparison of our results with previously described interactions of *H. pylori* proteins and also of orthologous *E. coli* proteins supports the reliability of the approach. Protein interaction maps can be built at the scale of a proteome. Our technology is applicable to higher eukaryotes, for which highly complex random-primed complementary DNA libraries are screened for interacting domains. The identification of interacting domains is a direct consequence of the library screening approach and presents key advantages such as mapping of new functional domains or correlation between sequence similarity and functional homology. These interacting domains also constitute a first step towards the construction of dominant-negative mutants or the development of an assay for interaction modulation, applicable to new drug design. □

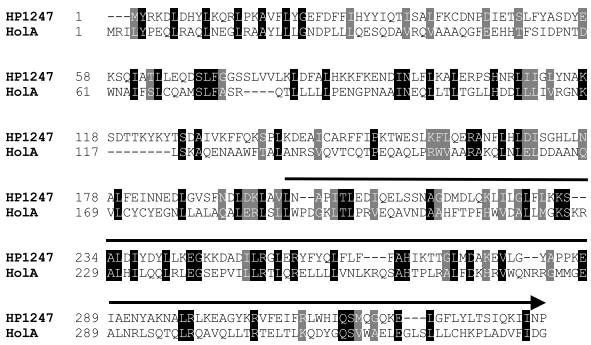

Figure 4 Alignment between *H. pylori* HP1247 protein and *E. coli* HolA. The two sequences were aligned using the FastA algorithm. Identical (black) and similar (grey) residues are outlined. The position of HP1247 C-terminal domain interacting with HolB is indicated by a line.

Table 2 Interacting proteins encoded by physically related genes

Protein 1		Protein 2	
ORF	Comments	ORF	Comments
HP0047†	HypE*, hydrogenase functionally related protein	HP0048	HypF*, hydrogenase putative transcriptional regulator
HP0061	–	HP0066	ATP-binding motif
HP0064	–	HP0063	–
HP0311	–	HP0312	ATP-binding motif
HP0338	–	HP0337	–
HP0691	ScoA, 3-oxoadipate coA-transferase subunit A	HP0692	ScoB, 3-oxoadipate coA-transferase subunit B
HP0697	–	HP0696	Putative hydantoin utilization*
HP0800	MoaE*, molybdopterin converting factor, subunit 2	HP0801	MoaD*, molybdopterin converting factor, subunit 1
HP0868	–	HP0869	HypA*, hydrogenase functionally related protein
HP0874	–	HP0875	KatA, catalase
HP0978	FtsA*, cell-division protein	HP0979	FtsZ*, cell-divison protein
HP1244	RpsR, ribosomal protein	HP1246	RpsF, ribosomal protein

*Function assigned by homology.
†The nomenclature used is that of PyloriGene database.

Methods

Bait cloning

Baits were constructed by PCR amplification and cloning in the pB6 plasmid derived from the original pAS2ΔΔ (ref. 2). Design of the primers was automatically proposed by a software and validated for each ORF. PCR fragments were cloned by classical enzymatic methods in a 96-well-plate format. All bait constructs used for interaction screens were fully sequenced and compared to the genome sequence; any mutant clone was discarded.

Library construction

We extracted genomic DNA from *H. pylori* 26695 as described[23], and nebulized and blunted it with a cocktail of mung bean nuclease, T4 and Klenow polymerase (NEB). Adapters containing *Sfi*1 sites were ligated to blunt DNA. The adapted DNA was cloned into pP6 plasmid derived from the original pACT2 and transformed in *E. coli* (DH10B; LifeTechnologies). Sequence analysis was performed on one hundred randomly chosen clones to establish the general characteristics of the library.

Screening procedure

The principle of the technique has been described[2]. Briefly, the screening conditions were adapted for each bait during a test screen, before performing the full-size screening experiment. The selectivity of the HIS3 reporter was modulated with 3-aminotriazole (3AT). For about 15% of the screens, diploid cells were plated on selective medium containing 3AT. In 44% of the screens, the second reporter (*lacZ*) was used directly on plates for the selection of clones. In other cases, the *lacZ* reporter was used as a second round of selection only for the selected clones. In all cases, LacZ activity was measured in a quantitative luminometric assay (Tropix).

Identification of interacting fragments

The prey fragments of the positive clones were amplified by PCR, analysed on agarose gel, partially sequenced at their 5′ junction on a PE 3700 Sequencer and mapped on the genomic sequence. Many clones were sequenced at their 3′ junction to map precisely the SID. All the steps after picking of positive clones were performed in bar-coded 96-well plates and automated with Beckman Biomek 2000 and Multimek automats. At the end of each screening experiment, the identity of the bait plasmid was controlled on a few positive clones.

PIM biological score

The PIM biological score (PBS) computation relies on two different levels of analysis: first, a local (that is, taking into account only the results of one screen) score is computed for each screen; and second, the global score is computed from the local scores by integrating results from all screens performed within the same genomic library. Local scores are thus computed only once, while global scores are recomputed each time new screens are performed. For each screen, fragments are clustered by overlap to delimit SIDs. Fragments that have no or very improbable coding capability (antisense, intergenic region, and out-of-frame fusion fragments selected in a single frame) are then eliminated from the set of prey fragments identified from positive clones. Assuming that prey fragments compete for the bait with 'equal chances', the probability p for a given fragment to be selected in an experiment is proportional to its expected number of occurrences within the library. p is

computed as a function of the fragment length and position, and of the length and position distributions of fragments in the prey library (these distributions are calibrated using data from random sequencing).

The local score is the probability for a given SID to be obtained under the equal chance hypothesis, that is, as a result of random noise. It is deduced by combining probabilities p (using a binomial law) from each of the independent fragment defining it. A (global) PBS is computed for each protein interaction after pooling results from all screens. On the basis of an independence hypothesis, scores from different screens are combined together when the same protein domain pair is involved. The resulting PBS thus represents the probability that the protein–protein interaction is due to noise. Scores are real numbers ranging from 0 to 1, but are grouped in four categories (A, B, C and D) for practical purposes. Finally, the global connectivity of the interaction map is analysed to tag separately (category E) SIDs found as prey with frequency greater than a fixed threshold: the PBS of each protein–protein interaction involving highly connected SIDs is set to 1. Both the intercategory thresholds and the high-connectivity threshold were defined manually, taking into account the nature of the studied organism, the relevant library and the current coverage of the proteome ($A < 1^{e-10} < B < 1^{e-5} < C < 1^{e-2.5} < D$; the E category corresponds to prey SIDs selected with more than 4 baits and was arbitrarily attributed a PBS value of 1).

Bioinformatics

Several algorithms and software were implemented in the production database to facilitate experimental steps, such as a 'bait program' that designed automatically oligonucleotides for PCR amplification and sequencing of bait constructs, a 'prey program' that determined the position of each fragment in the genome and its coding capacity (such as intergene, antisense, nucleotide position in an ORF, coding frame). The interactions were then analysed through a web-based software platform, the PIMRider developed at Hybrigenics and accessible through the web interface (http://pim.hybrigenics.com). Academic users will be granted a free licence. Other users will have to purchase a commercial licence. The *H. pylori* PIMRider platform is linked to the PyloriGene database.

Received 8 August; accepted 10 October 2000.

1. Fields, S. The future is function. *Nature Genet.* **15,** 325–327 (1997).
2. Fromont-Racine, M., Rain, J. C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16,** 277–282 (1997).
3. Bartel, P. L., Roecklein, J. A., SenGupta, D. & Fields, S. A protein linkage map of *Escherichia coli* bacteriophage T7. *Nature Genet.* **12,** 72–77 (1996).
4. Flajolet, M. *et al.* A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene* **242,** 369–379 (2000).
5. McCraith, S., Holtzman, T., Moss, B. & Fields, S. Genome-wide analysis of vaccinia virus protein–protein interactions. *Proc. Natl Acad. Sci. USA* **97,** 4879–4884 (2000).
6. Ito, T. *et al.* Toward a protein–protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl Acad. Sci. USA* **97,** 1143–1147 (2000).
7. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403,** 623–627 (2000).
8. Walhout, A. J. M. *et al.* Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287,** 116–122 (2000).
9. Tomb, J. F. *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388,** 539–547 (1997).
10. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28,** 45–48 (2000).
11. Alm, R. A. *et al.* Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397,** 176–180 (1999).
12. Moszer, I. The complete genome of *Bacillus subtilis*: from sequence annotation to data management and analysis. *FEBS Lett.* **430,** 28–36 (1998).
13. Welch, M., Chinardet, N., Mourey, L., Birck, C. & Samama, J. P. Structure of the CheY-binding domain of histidine kinase CheA in complex with CheY. *Nature Struct. Biol.* **5,** 25–29 (1998).
14. Cussac, V., Ferrero, R. L. & Labigne, A. Expression of *Helicobacter pylori* urease genes in *Escherichia coli* grown under nitrogen-limiting conditions. *J. Bacteriol.* **174,** 2466–2473 (1992).
15. Mobley, H. L., Island, M. D. & Hausinger, R. P. Molecular biology of microbial ureases. *Microbiol. Rev.* **59,** 451–480 (1995).
16. Skouloubris, S., Thiberge, J. M., Labigne, A. & De Reuse, H. The *Helicobacter pylori* UreI protein is not involved in urease activity but is essential for bacterial survival *in vivo*. *Infect. Immun.* **66,** 4517–4521 (1998).
17. Weeks, D. L., Eskandari, S., Scott, D. R. & Sachs, G. A H^+-gated urea channel: the link between *Helicobacter pylori* urease and gastritic colonization. *Science* **287,** 482–485 (2000).
18. Dong, Z., Onrust, R., Skangalis, M. & O'Donnell, M. DNA polymerase III accessory proteins. I. holA and holB encoding delta and delta′. *J. Biol. Chem.* **268,** 11758–11765 (1993).
19. Liu, X. & Matsumura, P. An alternative sigma factor controls transcription of flagellar class-III operons in *Escherichia coli*: gene sequence, overproduction, purification and characterization. *Gene* **164,** 81–84 (1995).
20. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 A resolution. *Cell* **98,** 811–824 (1999).
21. Mooney, R. A. & Landick, R. RNA polymerase unveiled. *Cell* **98,** 687–690 (1999).
22. Vidal, M. & Legrain, P. Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res.* **27,** 919–929 (1999).
23. Ferrero, R. L., Cussac, V., Courcoux, P. & Labigne, A. Construction of isogenic urease-negative mutants of *Helicobacter pylori* by allelic exchange. *J. Bacteriol.* **174,** 4212–4217 (1992).

Acknowledgements

We thank M. Fromont-Racine, P. Glaser, A. Jacquier, A. Brunet and L. Decourty for their help at the launch of this project; M. Fejes, G. Conan and P. Desmoucelle for technical assistance; G. Boissy and J.-L. Divol for their help in software development; F. Colland for his contribution to the mapping of FliA interacting domain on the 3D structure of the core RNA polymerase; and S. Whiteside for a thorough and critical reading of the manuscript. We are very grateful to R. Benarous, J. Camonis, L. Daviet, M. Rosbash, A.D. Strosberg and S. Whiteside for many stimulating discussions. This work was supported by an interest-free loan from the ANVAR. P.L. is on leave from the CNRS.

Correspondence and requests for materials should be addressed to P.L. (e-mail: plegrain@hybrigenics.fr).

Crystal structures of SarA, a pleiotropic regulator of virulence genes in *S. aureus*

Maria A. Schumacher*†, Barry K. Hurlburt‡ & Richard G. Brennan*

** Department of Biochemistry and Molecular Biology and † Vollum Institute, Oregon Health Sciences University, Portland, Oregon 97201-3098, USA*
‡ Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, USA

Staphylococcus aureus **is a major human pathogen, the potency of which can be attributed to the regulated expression of an impressive array of virulence determinants. A key pleiotropic transcriptional regulator of these virulence factors is SarA, which is encoded by the *sar* (staphylococcal accessory regulator) locus[1–3]. SarA was characterized initially as an activator of a second virulence regulatory locus, *agr*, through its interaction with a series of heptad repeats (AGTTAAG) within the *agr* promoter[4]. Subsequent DNA-binding studies have revealed that SarA binds readily to multiple AT-rich sequences of variable lengths[4–11]. Here we describe the crystal structure of SarA and a SarA–DNA complex at resolutions of 2.50 Å and 2.95 Å, respectively. SarA has a fold consisting of a four-helix core region and 'inducible**

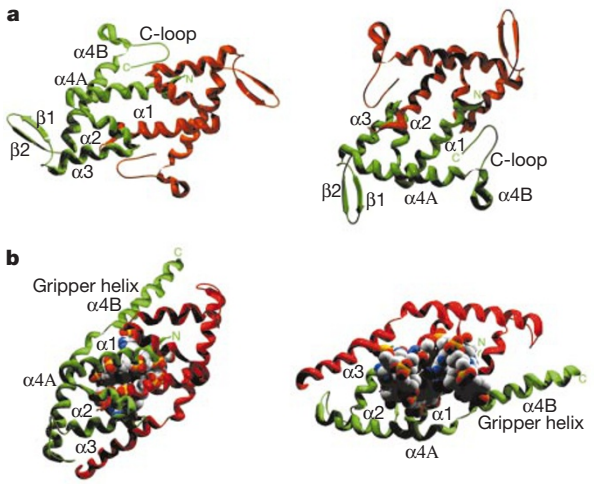

Figure 1 The SarA dimer and SarA–DNA complex. **a**, Views of SarA looking directly into the DNA-binding pocket or from the 'back' side. The secondary structures, β-hairpin and C-terminal loop of one monomer are labelled. Each monomer is red or green. **b**, The SarA–DNA complex in the identical orientation of the corresponding apo SarA directly above. The DNA duplex is shown as CPK atoms, with carbon, nitrogen, oxygen and phosphates coloured white, blue, red and yellow, respectively. The α4B gripper helix of one monomer is also labelled. The narrow and deep minor groove and major groove are seen on the left and right respectively.