LETTERS

Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea

Niels-Ulrik Frigaard[1]†, Asuncion Martinez[1], Tracy J. Mincer[1] & Edward F. DeLong[1]

Planktonic Bacteria, Archaea and Eukarya reside and compete in the ocean's photic zone under the pervasive influence of light. Bacteria in this environment were recently shown to contain photoproteins called proteorhodopsins, thought to contribute to cellular energy metabolism by catalysing light-driven proton translocation across the cell membrane[1–7]. So far, proteorhodopsin genes have been well documented only in proteobacteria and a few other bacterial groups. Here we report the presence and distribution of proteorhodopsin genes in Archaea affiliated with the order Thermoplasmatales, in the ocean's upper water column. The genomic context and phylogenetic relationships of the archaeal and proteobacterial proteorhodopsins indicate its probable lateral transfer between planktonic Bacteria and Archaea. About 10% of the euryarchaeotes in the photic zone contained the proteorhodopsin gene adjacent to their small-subunit ribosomal RNA. The archaeal proteorhodopsins were also found in other genomic regions, in the same or in different microbial lineages. Although euryarchaeotes were distributed throughout the water column, their proteorhodopsins were found only in the photic zone. The cosmopolitan phylogenetic distribution of proteorhodopsins reflects their significant light-dependent fitness contributions, which drive the photoprotein's lateral acquisition and retention, but constrain its dispersal to the photic zone.

Several different groups of Archaea are commonly found in marine plankton[8,9]. Pelagic crenarchaeotes comprise a large fraction of the prokaryotic plankton in coastal waters and the subphotic zone[8–14], and seem to be major participants in chemoautotrophic ammonia oxidation in the ocean[15–18]. Pelagic euryarchaeotes comprise at least three separate phylogenetic groups based on small-subunit (SSU) rRNA sequence analyses. The most abundant clade, so-called planktonic marine 'group II euryarchaeotes', is peripherally related to the Thermoplasmatales[9,14]. Group II euryarchaeotes are typically more abundant than crenarchaeotes in ocean surface waters[9–11], and blooms of these microorganisms have been observed in surface waters of Monterey Bay and the North Sea during summer[13,14]. A few genome fragments from these planktonic euryarchaeotes have now been characterized[19,20], but little else is known about their genetic makeup, physiology or ecology.

Prokaryotic rhodopsins, first discovered in extremely halophilic Archaea (haloarchaea), are membrane proteins that bind retinal and respond to light stimuli. The known functional repertoire of these photoproteins now includes energy-conserving transmembrane proton pumps (bacteriorhodopsins, proteorhodopsins and xanthorhodopsins), transmembrane chloride pumps (halorhodopsins) and light sensors (sensory rhodopsins)[1,21]. Among the haloarchaea, multiple rhodopsins with diversified functions can exist within a single cell. For example, the haloarchaeon *Haloarcula marismortui* genome encodes six rhodopsins: one proton-pumping bacteriorhodopsin, one chloride-pumping halorhodopsin, two sensory rhodopsins, and two opsins of unknown function[22] (Fig. 1a).

Prokaryotic rhodopsins were originally thought to exist exclusively in halophilic Archaea, but recent environmental genomic studies have revealed the existence, distribution and variability of a new class of such photoproteins, called proteorhodopsins, in members of the domain Bacteria[1–7,23] (Fig. 1a). Biochemical and biophysical characterization of heterologously expressed proteorhodopsins indicates their potential functional role in light-activated proton translocation across bacterial membranes[1,2,7]. Although planktonic Bacteria harbouring proteorhodopsins may not be photosynthetic *sensu stricto* with respect to carbon fixation, they probably gain a competitive advantage by using these photoproteins to generate a light-driven chemiosmotic potential. So far, proton-pumping proteorhodopsins have been clearly documented only in marine proteobacteria[1–3,6], including the recently cultivated and ubiquitous marine bacterium *Pelagibacter ubique*[4,5].

In a survey of the genomic diversity of marine picoplankton, we randomly sequenced both ends of large-insert DNA clones recovered from marine picoplankton in the North Pacific Subtropical Gyre[24]. A fosmid clone (HF70_39H11) was identified that encoded a proteorhodopsin protein adjacent to a euryarchaeal-like LysE homologue on one end, and a euryarchaeal-like lysyl-tRNA synthetase on the other (Fig. 2). The presence of a bacterial-like proteorhodopsin on a planktonic archaeal genome fragment was unexpected, because no proteorhodopsin-like genes have previously been reported in Archaea. Two fosmid clones containing the novel proteorhodopsin (HF70_39H11 and HF70_59C08) were fully sequenced and compared (see Methods). The two clones formed a pseudocontig (Fig. 2) sharing 16 kilobase pairs of sequence overlap having 97% DNA sequence identity and complete gene synteny, except for one hypothetical gene unique to each (Supplementary Tables S1 and S2). Clone HF70_59C08 also contained genes encoding 22 ribosomal proteins, a preprotein translocase SecY subunit, and the SSU rRNA. Phylogenetic analyses of all these genes clearly identified both fosmids as deriving from marine group II euryarchaeotes (Fig. 3 and Supplementary Fig. S1). Heterologous expression of the HF70_59C08 proteorhodopsin in *Escherichia coli* confirmed that the archaeal photoprotein binds retinal covalently, and preliminary experiments indicate its potential function as a light-driven transmembrane proton pump *in vivo* (A.M., N.U.F. and E.F.D., unpublished observations).

We screened the large-insert DNA clone libraries prepared from various depths[24] to determine the evolutionary relatedness and environmental distribution of planktonic euryarchaeotes and archaeal proteorhodopsins in the water column. A total of 9,216 fosmid clones from each depth (330 Mb of cloned DNA per depth) were hybridized with a planktonic euryarchaeal SSU rRNA gene probe and an archaeal proteorhodopsin gene probe (see Methods).

[1]Department of Civil and Environmental Engineering and Division of Biological Engineering, Massachusetts Institute of Technology, Building 48, 15 Vassar Street, Cambridge, Massachusetts 02139, USA. †Present address: Institute of Molecular Biology and Physiology, University of Copenhagen, Sølvgade 83H, 1307 Copenhagen K, Denmark.

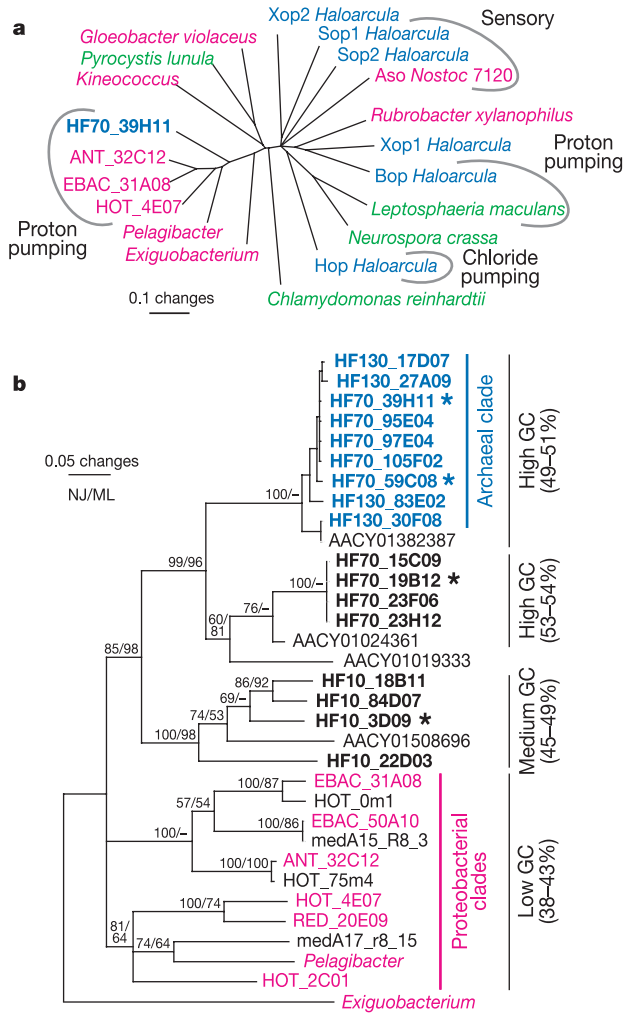

Figure 1 | Phylogenetic relationships of rhodopsins. **a**, Selected microbial rhodopsins representing their broad organismal and functional diversity. **b**, Proteorhodopsin-like rhodopsins (GC content refers to the nucleotide composition of the proteorhodopsin genes). Sequences obtained in this study are marked in bold. Fully sequenced fosmid clones are marked with an asterisk. Colours indicate organismal affiliation: blue, Archaea; magenta, Bacteria; green, Eukarya; black, uncertain. The trees are based on protein sequences and neighbour-joining distance calculations (see Methods). Bootstrap support (percentage) based on 1,000 replications is shown for neighbour-joining (NJ; first number) and maximum-likelihood (ML; second number) trees. Nodes with less than 50% support were collapsed.

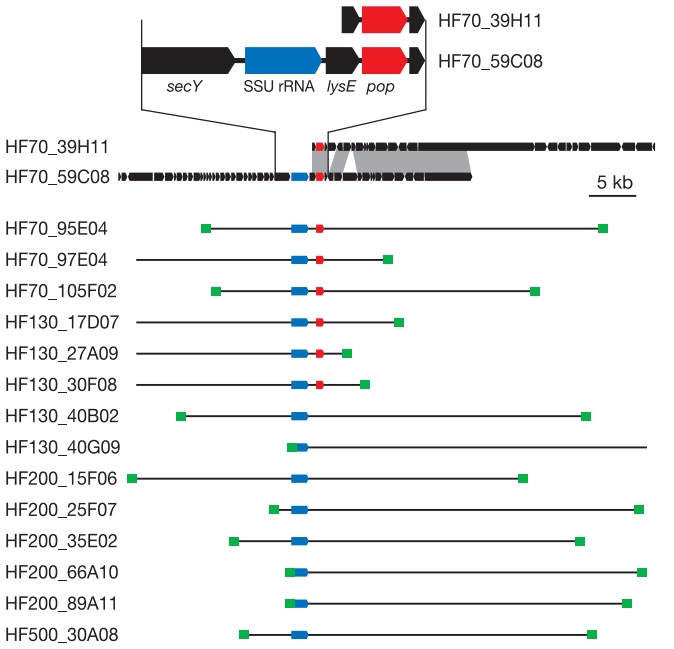

Figure 2 | Genetic maps and alignments of fosmid clones from the picoplankton genomic DNA libraries. Two fosmid clones containing archaeal SSU rRNA (HF70_39H11 and HF70_59C08) were fully sequenced. Proteorhodopsin, the SSU rRNA gene, and both clone termini were sequenced for the other fosmids shown. Colours indicate sequence type: blue, SSU rRNA gene; red, proteorhodopsin gene; green, clone terminus. Numbers directly following the HF clone designator indicate the depth of origin of the clone library: HF70, 70 m; HF130, 130 m; HF200, 200 m.

Archaeal SSU rRNA genes identified in fosmid clones from each depth were sequenced to determine their phylogenetic affiliation. About 96% of all archaeal rRNA genes recovered in the upper water column were affiliated with group II euryarchaeotes (Fig. 3). Archaeal proteorhodopsins were also PCR amplified and sequenced, and compared with previously characterized rhodopsins (Fig. 1b and Supplementary Fig. S2).

Nine of the sequenced proteorhodopsins, including those from the fully sequenced fosmid clones, clearly originated from group II euryarchaeotes on the basis of the genetic linkage of the proteorhodopsin gene with archaeal SSU rRNA (Fig. 3). The archaeal proteorhodopsins formed a phylogenetic clade separate from other known proteobacterial proteorhodopsins (Fig. 1b) and had a G+C content of 49–51%, significantly higher than that of most other previously reported proteorhodopsin genes. Two other fully sequenced fosmids contained archaeal-like proteorhodopsins that fell into either the other high G+C clade (HF70_19B12), or the medium G+C clade (HF10_3D09) (Fig. 1b and Supplementary Fig. S3). The gene content of these fosmids also suggested a euryarchaeal origin, and indicated that some archaeal-like proteorhodopsins are found in alternative genomic contexts unlinked to the SSU rRNA gene (Supplementary Fig. S3). Very few closely related homologues of the archaeal proteorhodopsins were found in the recently reported Sargasso Sea shotgun sequence data set[23], where 90% of all proteorhodopsin gene fragments screened had a G+C content of 35–45% (Supplementary Fig. S4), with most clustering in proteobacterial clades.

Fosmid clones encoding both euryarchaeal SSU rRNA and proteorhodopsin (Fig. 3) all shared near-identical SSU rRNAs (minimum 99.7% nucleotide sequence identity) and proteorhodopsins (minimum 95% amino acid sequence identity). In addition, terminal sequences of euryarchaeal rRNA-containing fosmids shared near-complete synteny with the two fully sequenced clones (HF70_39H11 and HF70_59C08). The coherent phylogenetic clustering of the linked SSU rRNA and proteorhodopsins on these genome fragments (Fig. 2), and the absence of their linkage in closely related, co-occurring strains (Figs 2 and 3), indicate that this arrangement might have been acquired only recently. Because all available data suggest that planktonic euryarchaeote genomes contain only a single SSU rRNA, this gene linkage seems to occur in about 10% of all the euryarchaeotes detected in the photic zone (8 of a total of 82 clones; Fig. 3). The remaining archaeal-like proteorhodopsin genes (unlinked to the SSU rRNA) seem to reside in different regions of the euryarchaeal genome, as indicated by photoproteins existing in different genetic contexts on euryarchaeal fosmids (HF70_19B12 and HF10_3D09; Fig. 1b) that lack the SSU rRNA gene (Supplementary Tables S3 and S4, and Supplementary Fig. S3).

Substantial numbers of euryarchaeal SSU rRNA genes and archaeal-type proteorhodopsin genes were observed in the photic zone, in roughly equal proportions (Fig. 4). In contrast, while euryarchaeal SSU rRNA genes were well represented in the subphotic zone (200 m and below), archaeal-type proteorhodopsins were virtually absent (Fig. 4). Several archaeal clones lacked

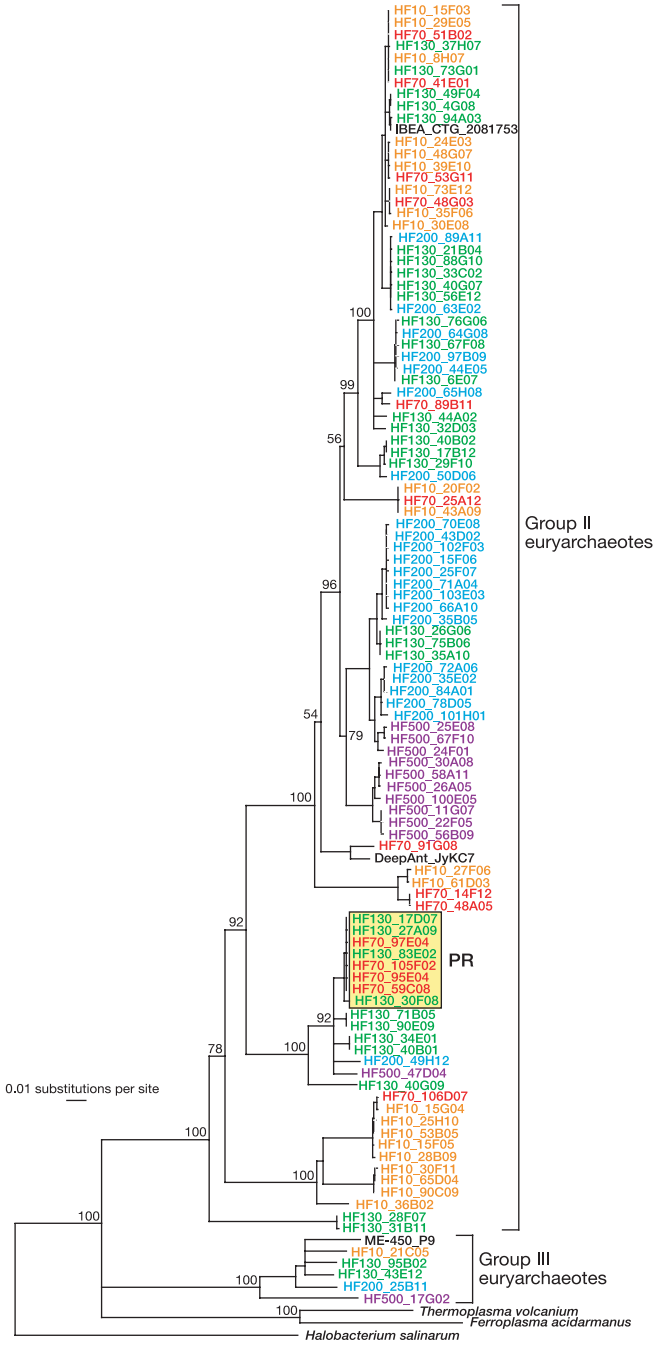

Figure 3 | Phylogenetic tree of euryarchaeal SSU rRNA sequences from the picoplankton genomic DNA libraries. Numbers directly following the HF clone designator indicate the depth of origin of each clone. Origins are also colour-coded: orange, 10 m; red, 70 m; green, 130 m; blue, 200 m; purple, 500 m; black, other origin. Clones containing both SSU rRNA and proteorhodopsin genes are indicated by PR. The tree is based on neighbour-joining distance calculations with Jukes–Cantor correction (bootstrap support (percentage) based on 1,000 replications is shown). *Methanosarcina mazei* was used as an outgroup (not shown). Nodes with less than 50% support were collapsed.

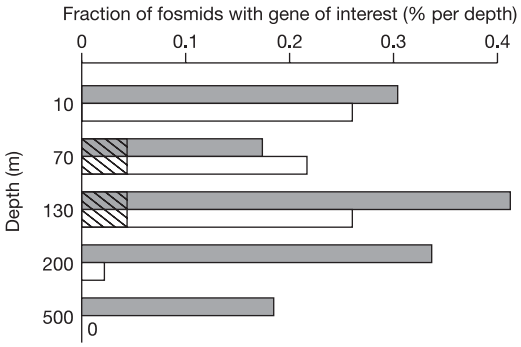

Figure 4 | Fosmid distribution in the picoplankton genomic DNA libraries. Bars indicate the fraction of fosmid clones at each depth interval that contain a euryarchaeal SSU rRNA gene (grey), an archaeal-like proteorhodopsin gene (white), or both a euryarchaeal SSU rRNA gene and an archaeal-like proteorhodopsin gene (hatched). $n = 9{,}216$ clones assayed at each depth.

proteorhodopsin, but otherwise appeared syntenic with proteorhodopsin-containing clones based on SSU rRNA gene presence and aligned terminal sequences (Fig. 2). This group shares a minimum of 88% SSU rRNA sequence identity with the proteorhodopsin-containing types, indicating substantial genetic heterogeneity within group II euryarchaeotes. These phylotypes, derived primarily from the subphotic zone, represent lineages of group II euryarchaeotes related to the proteorhodopsin-containing genotypes but apparently lacking proteorhodopsin anywhere in their genomes (Fig. 4). Planktonic euryarchaeotes in the subphotic zone would presumably gain no competitive advantage from proteorhodopsin, so it is not surprising that they either have not acquired the photoprotein gene, or have not retained it.

Microbial rhodopsin-based photosystems are relatively simple and are encoded by very few genes. The retinal chromophore can be generated in one enzymatic step by the oxidative cleavage of a carotenoid[7,25,26]. Assuming that carotenoid biosynthesis is present, just one gene enabling this enzymatic step, and another encoding an opsin, are sufficient to generate a functional rhodopsin. The genetic simplicity of these photosystems, their ability to assemble and function properly in the membranes of divergent microbial groups, and their potential to contribute to cellular energy metabolism, all are indicative of their likely predisposition for genetic mobility. Apparently, lateral gene dispersal mechanisms, coupled with strong selection for proteorhodopsin in the light, have contributed to the distribution of these photoproteins among various members of all three of life's domains. Proteorhodospins therefore represent a category of 'cosmopolitan genes'[27] whose broad phylogenetic distribution is driven in part by lateral gene transfer, which further influences the recipient lineage's evolution and speciation[28,29]. The spatial distribution of such promiscuous genes, including those encoding proteorhodopsin, appears more reflective of their functional properties in relation to the environment than of the specific organisms that harbour them.

METHODS

Seawater samples collected at 22° 45′ N, 158° 00′ W (100 km north of Oahu, Hawaii) in October 2002 were passed through a 1.6-μm prefilter and retained on a 0.22-μm filter[24]. Large-insert genomic DNA fosmid clone libraries were prepared by using the vector pCC1FOS (Epicentre) as described previously[30]. Clones were spotted on Hybond-N^+ nylon filters (Amersham Biosciences) by Amplicon Express (Pullman, Washington). Hybridization was performed at 60 °C with polymerase chain reaction (PCR)-generated DNA probes with the use of AlkPhos Direct Labelling and ECF Chemifluorescent Detection kits (Amersham Biosciences). PCR amplification of proteorhodopsin genes was done either with primer pair ArPRfor1 (5′-GACTATGTGGGTATTTCC-3′) and ArPRrev1 (5′-GCCGAATGCGGTCTTATTGACCAAATC-3′) or with primer pair o-PR2 (5′-WWNMGNTAYGTNGAYTGG-3′) and o-PR3 (5′-GGRTA DATNGCCCANCC-3′)[7]. PCR amplification of SSU rRNA genes was performed with primers Ar20F (5′-TTCCGGTTGATCCYGCCR-3′) and U1390R (5′-GACGGGCGGTGTGTRC-3′). PCR products were sequenced directly with an ABI PRISM BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). PCR products made with degenerate primers were cloned with the TOPO TA Cloning for Sequencing kit (Invitrogen) before being sequenced. Fosmid clones were fully sequenced with the TOPO Shotgun Subcloning kit (Invitrogen)

combined with sequence assembly with Sequencher v. 4.5 (Gene Codes Corporation) and annotation with FGENESB (Softberry) and Artemis v. 6 (The Wellcome Trust Sanger Institute). Sequences were aligned with the use of ClustalX v. 1.83 (ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/) and SeaView (http://pbil.univ-lyon1.fr/software/seaview.html), and phylogenetic analyses were performed with PAUP 4.0 (Sinauer Associates) and Tree-Puzzle v. 5.2 (http://www.tree-puzzle.de/).

Received 8 August; accepted 15 November 2005.

1. Béjà, O. *et al.* Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science* **289,** 1902–1906 (2000).
2. Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411,** 786–789 (2001).
3. de la Torre, J. R. *et al.* Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc. Natl Acad. Sci. USA* **100,** 12830–12835 (2003).
4. Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309,** 1242–1245 (2005).
5. Giovannoni, S. J. *et al.* Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438,** 82–85 (2005).
6. Sabehi, G., Béjà, O., Suzuki, M. T., Preston, C. M. & DeLong, E. F. Different SAR86 subgroups harbour divergent proteorhodopsins. *Environ. Microbiol.* **6,** 903–910 (2004).
7. Sabehi, G. *et al.* New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol.* **3,** 1–9 (2005).
8. Fuhrman, J. A., McCallum, K. & Davis, A. A. Novel major archaebacterial group from marine plankton. *Nature* **356,** 148–149 (1992).
9. DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl Acad. Sci. USA* **89,** 5685–5689 (1992).
10. Massana, R., Murray, A. E., Preston, C. M. & DeLong, E. F. Vertical distribution and phylogenetic characterization of marine planktonic *Archaea* in the Santa Barbara Channel. *Appl. Environ. Microbiol.* **63,** 50–56 (1997).
11. Massana, R., DeLong, E. F. & Pedrós-Alió, C. A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl. Environ. Microbiol.* **66,** 1777–1787 (2000).
12. Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409,** 507–510 (2001).
13. Pernthaler, A., Preston, C. M., Pernthaler, J., DeLong, E. F. & Amann, R. Comparison of fluorescently labeled oligonucleotide and polynucleotide probes for the detection of pelagic marine Bacteria and Archaea. *Appl. Environ. Microbiol.* **68,** 661–667 (2002).
14. DeLong, E. F. Oceans of *Archaea*. *ASM News* **69,** 503–511 (2003).
15. Pearson, A., McNichol, A. P., Benitez-Nelson, B. C., Hayes, J. M. & Eglinton, T. I. Origins of lipid biomarkers in Santa Monica Basin surface sediment: A case study using compound-specific $\Delta^{14}C$ analysis. *Geochim. Cosmochim. Acta* **65,** 3123–3137 (2001).
16. Wuchter, C., Schouten, S., Boschker, H. T. & Sinninghe Damsté, J. S. Bicarbonate uptake by marine Crenarchaeota. *FEMS Microbiol. Lett.* **219,** 203–207 (2003).
17. Herndl, G. J. *et al.* Contribution of *Archaea* to total prokaryotic production in the deep Atlantic Ocean. *Appl. Environ. Microbiol.* **71,** 2303–2309 (2005).
18. Könneke, M. *et al.* Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437,** 543–546 (2005).
19. Béjà, O. *et al.* Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* **2,** 516–529 (2000).
20. Moreira, D., Rodríguez-Valera, F. & López-García, P. Analysis of a genome fragment of a deep-sea uncultivated Group II euryarchaeote containing 16S rDNA, a spectinomycin-like operon and several energy metabolism genes. *Environ. Microbiol.* **6,** 959–969 (2004).
21. Balashov, S. P. *et al.* Xanthorhodopsin: A proton pump with a light-harvesting carotenoid antenna. *Science* **309,** 2061–2064 (2005).
22. Baliga, N. S. *et al.* Genome sequence of *Haloarcula marismortui*: A halophilic archaeon from the Dead Sea. *Genome Res.* **14,** 2221–2234 (2004).
23. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304,** 66–74 (2004).
24. DeLong, E. F. *et al.* Community genomics among stratified microbial assemblages in the ocean's interior. *Science* (in the press).
25. Giuliano, G., Al-Babili, S. & von Lintig, J. Carotenoid oxygenases: Cleave it or leave it. *Trends Plant Sci.* **8,** 145–149 (2003).
26. Ruch, S., Beyer, P., Ernst, H. & Al-Babili, S. Retinal biosynthesis in Eubacteria: *in vitro* characterization of a novel carotenoid oxygenase from *Synechocystis* sp. PCC 6803. *Mol. Microbiol.* **55,** 1015–1024 (2005).
27. Woese, C. R. A new biology for a new century. *Microbiol. Mol. Biol. Rev.* **68,** 173–186 (2004).
28. Ochman, H., Lerat, E. & Daubin, V. Examining bacterial species under the specter of gene transfer and exchange. *Proc. Natl Acad. Sci. USA* **102** (Suppl. 1), 6595–6599 (2005).
29. Ochman, H., Lerat, E., Daubin, V. & Moran, N. A. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* **3,** e130 (2005).
30. Suzuki, M. T. *et al.* Phylogenetic screening of ribosomal RNA gene-containing clones in bacterial artificial chromosome (BAC) libraries from different depths in Monterey Bay. *Microb. Ecol.* **48,** 473–488 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Preston and L. Christianson for assistance in the initial sample collection and fosmid library construction; P. Richardson, K. Barry, S. Pitluck and the Joint Genome Institute production team for their help in sequencing; and D. Karl at the University of Hawaii, and the Hawaii Ocean Time Series team, who made sample collection possible. This research was supported by a grant to E.F.D. from The Gordon and Betty Moore Foundation under the Marine Microbiology Initiative program, and a National Science Foundation Microbial Observatory grant to E.F.D. Terminus sequencing of large-insert clones was performed by the Joint Genome Institute (Walnut Creek, California) under the auspices of the US Department of Energy's Office of Science, Biological, and Environmental Research Microbial Genomes Program.

Author Contributions E.F.D. made the initial bioinformatic observation of linkage between proteorhodopsin and archaeal genes. N.-U.F. and E.F.D. developed the concepts of the paper together. N.-U.F. performed the experiments, except the proteorhodopsin expression experiments, which were performed by A.M. T.J.M. participated in obtaining and analysing the SSU rRNA sequences. N.-U.F. wrote the first draft of the paper, which was completed by N.-U.F. and E.F.D. together.

Author Information The sequences reported here have been deposited in GenBank under accession numbers DQ257435, DQ257434, DQ156349, DQ156348 (fosmids HF10_3D09, HF70_19B12, HF70_39H11 and HF70_59C08), DQ156379–DQ156483 (SSU rRNA), DQ156350–DQ156363, DU708536–DU708556 (fosmid termini) and DQ156364–DQ156378 (proteorhodopsin sequences). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.F.D. (delong@mit.edu).