LETTERS

Proteorhodopsin in the ubiquitous marine bacterium SAR11

Stephen J. Giovannoni[1], Lisa Bibbs[4], Jang-Cheon Cho[1]†, Martha D. Stapels[2]†, Russell Desiderio[3], Kevin L. Vergin[1], Michael S. Rappé[1]†, Samuel Laney[3], Lawrence J. Wilhelm[1], H. James Tripp[1], Eric J. Mathur[4] & Douglas F. Barofsky[2]

Proteorhodopsins are light-dependent proton pumps that are predicted to have an important role in the ecology of the oceans by supplying energy for microbial metabolism[1,2]. Proteorhodopsin genes were first discovered through the cloning and sequencing of large genomic DNA fragments from seawater[1]. They were later shown to be widely distributed, phylogenetically diverse, and active in the oceans[3–7]. Proteorhodopsin genes have not been found in cultured bacteria, and on the basis of environmental sequence data, it has not yet been possible to reconstruct the genomes of uncultured bacterial strains that have proteorhodopsin genes. Although the metabolic effect of proteorhodopsins is uncertain, they are thought to function in cells for which the primary mode of metabolism is the heterotrophic assimilation of dissolved organic carbon. Here we report that SAR11 strain HTCC1062 (*'Pelagibacter ubique'*)[8], the first cultivated member of the extraordinarily abundant SAR11 clade, expresses a proteorhodopsin gene when cultured in autoclaved seawater and in its natural environment, the ocean. The *Pelagibacter* proteorhodopsin functions as a light-dependent proton pump. The gene is expressed by cells grown in either diurnal light or in darkness, and there is no difference between the growth rates or cell yields of cultures grown in light or darkness.

The proteorhodopsin gene was discovered during annotation of the complete genome sequence of strain HTCC1062 (refs 9, 10). *Pelagibacter* strain HTCC1062 is a coastal, ocean surface ecotype. Previous studies have shown that the SAR11 clade occurs throughout the water column, and that it has differentiated into ecotypes that dominate in different oceanic regions and at different depths, including the deep ocean where no light penetrates[11].

We amplified and sequenced proteorhodopsin genes from each of ten *Pelagibacter* isolates from coastal Oregon. Amino acid substitutions were observed at two positions among the set of *Pelagibacter* proteorhodopsin genes (Supplementary Fig. S1b and Supplementary Table S1). In phylogenetic comparisons, the *Pelagibacter* proteorhodopsin genes fall within a deeply branching clade that includes genes recovered from disparate ocean water samples[5–7] by environmental DNA cloning (Fig. 1). The *Pelagibacter* proteorhodopsin genes are most closely related to genes recovered by the shotgun cloning and sequencing of DNA from the Sargasso Sea[7]. In amino acid alignments, the *Pelagibacter* proteorhodopsin genes differ from their closest homologues in public databases at 46 out of 255 positions.

The *Pelagibacter* proteorhodopsin genes have structural features typical of rhodopsin proton pumps, including Asp 102 and Glu 113 residues, which are appropriately positioned to act as proton acceptor and donor residues in the retinylidene Schiff-base transfer reactions that occur during the proteorhodopsin photocycle[12] (Supplementary Fig. S1b). Position 105, which is associated with spectral tuning, is occupied by a leucine residue; this is typical of shallow-water proteorhodopsins that absorb green light[13]. Proteorhodopsin genes recovered from the Sargasso Sea, although belonging to

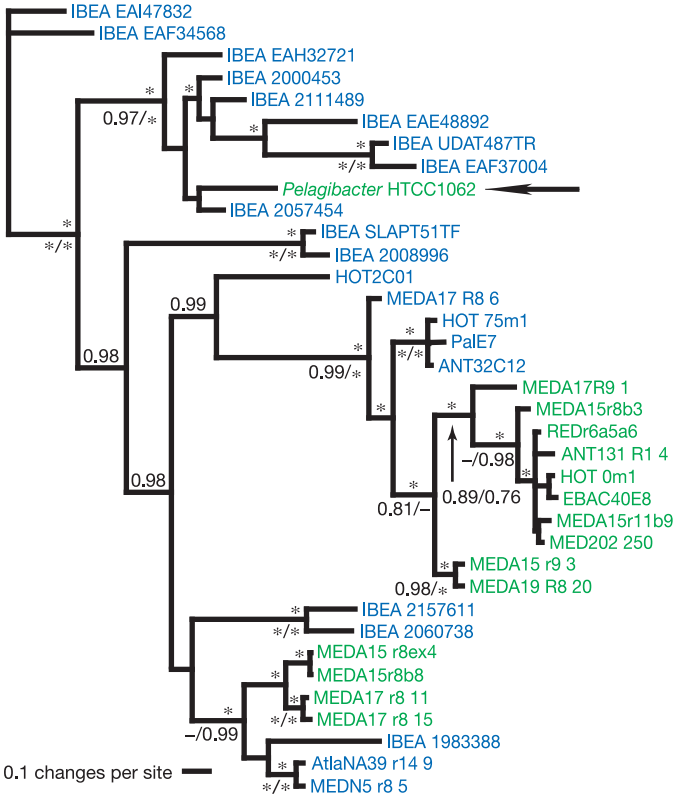

Figure 1 | Phylogenetic relationships between proteorhodopsin amino acid sequences. Shown is the relationship between the *Pelagibacter* strain HTCC1062 proteorhodopsin amino acid sequence and selected representatives of proteorhodopsin genes cloned from seawater DNA. The sequences most similar to *Pelagibacter* proteorhodopsin are included in the tree. Green text indicates proteorhodopsin genes encoding a leucine residue at position 105, blue text indicates a glutamine residue at this position. The tree was rooted with the sequences of rhodopsins from *Gloeobacter violaceus* and *Pyrocystis lunula*. Numbers above nodes are posterior probabilities; numbers below nodes are parsimony bootstrap values and neighbour-joining bootstrap values (separated by a slash). Asterisks indicate a value of 1.0.

[1]Department of Microbiology, [2]Department of Chemistry, and [3]College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA. [4]Diversa Corporation, 4955 Directors Place, San Diego, California 92121-1609, USA. †Present addresses: Department of Oceanography, 5N541, Inha University Younghyun-dong, Nam-Gu, Incheon 402-751, Korea (J.-C.C.); Waters Corporation, 34 Maple Street, Milford, Massachusetts 01757-3696, USA (M.D.S.); Hawaii Institute of Marine Biology, SOEST, University of Hawaii, PO Box 1346, Kaneohe, Hawaii 96744, USA (M.S.R.).

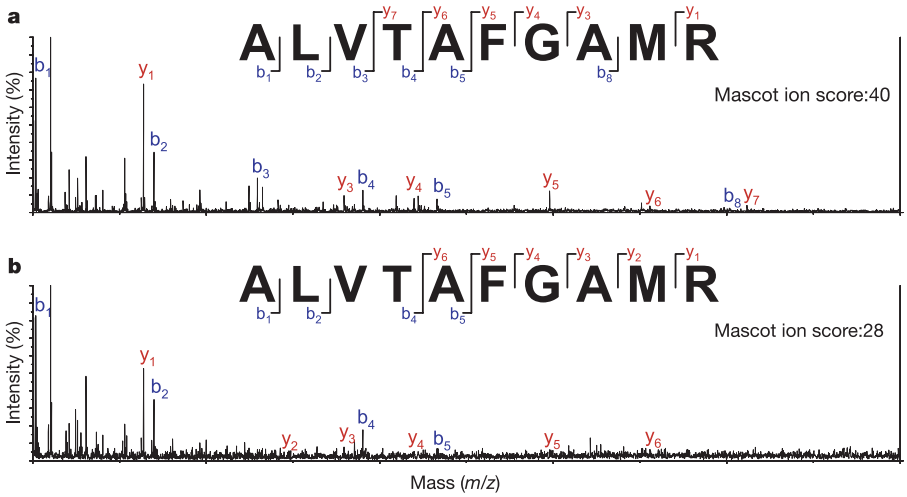

Figure 2 | ***Pelagibacter* proteorhodopsin peptide (PR3), detected by tandem MALDI mass spectrometry.** **a**, Cultured cells. **b**, Oregon coastal seawater.

the same proteorhodopsin clade, had glutamine residues at the corresponding position, which is typical of proteorhodopsins that absorb blue light.

The genomic region around the HTCC1062 proteorhodopsin gene shows substantial synteny with proteorhodopsin-containing contigs recovered from the Sargasso Sea by environmental DNA sequencing (Supplementary Fig. S1a), suggesting that many of these proteorhodopsin genes might originate from SAR11 strains adapted to life in ocean gyres. Among the Sargasso Sea contigs for which genes downstream of proteorhodopsin were identified, in 49% of the cases the gene encoded a ferridoxin, as observed in the HTCC1062 strain[10]. No genes related to carbon fixation were found in the HTCC1062 genome[10]. This is consistent with the function of proteorhodopsin proteins, which can supply a transmembrane electrochemical potential but not the reduced nucleotide cofactors needed for carbon dioxide reduction.

We used mass spectrometry to study expression of the proteorhodopsin protein. We used matrix-assisted laser desorption/ionization (MALDI) mass spectrometry, owing to the presence of mostly arginine-containing peptides in the tryptic digest of proteorhodopsin, which are favoured by the MALDI ionization process[14]. From membrane-fraction digests of HTCC1062 samples grown in sterilized natural seawater, we identified three proteorhodopsin canonical tryptic peptides with high confidence (Fig. 2a and Supplementary Table S1). The peptides were observed in cultures grown both in diurnal light and in the dark. Notably, two of these proteorhodopsin peptides were also identified in the highly complex protein mixture derived from cells that were collected by filtration from coastal Oregon seawater (Supplementary Table S1 and Fig. 2b). Our mass spectroscopy observations provide strong evidence that native populations of *Pelagibacter* express the proteorhodopsin gene. Other reports have successfully applied mass spectrometry to study the proteome state of natural microbial communities that consist of relatively few species[15]. Our successful use of mass spectrometry to detect proteorhodopsin in complex plankton communities can be attributed to the very high abundance of *Pelagibacter* cells in seawater[8] and to the use of protocols designed to enrich *Pelagibacter* proteorhodopsin protein, by size-fractionating both cells and proteins (see the Supplementary Methods).

We used physical methods to show that the HTCC1062 proteorhodopsin protein has the kinetic characteristics of an ion pump, and also confirmed that it is expressed by cells growing in the light or in darkness. Consistent with predictions from the amino acid sequence, cloned HTCC1062 proteorhodopsin expressed in *Escherichia coli* had an absorption spectrum with a maximum at 530 nm, which is typical of proteorhodopsin genes cloned from ocean surface samples, where the light field is dominated by green wavelengths[1] (Fig. 3a). Flash-induced absorbance transients from cultures of HTCC1062 grown in the light and in the dark are shown in Fig. 3b, c. The absorption of transient species produced by an excitation flash was monitored as a function of time. The 488-nm absorption transients fit exponential decay models with time constants of 34 ms for dark-grown cells and 13 ms for light-grown cells. Rapid photocycle rates of this order are characteristic of ion pumps rather than sensory rhodopsins, which have slower turnover times. The HTCC1062 proteorhodopsin expressed in *E. coli* showed photocycle rates similar to cells grown in darkness. Suspensions of *E. coli* cells expressing proteorhodopsin produced transient pH decreases when exposed to light, providing further support for the conclusion that this proteorhodopsin is a light-driven proton pump (Supplementary Fig. S2).

We estimated the number of proteorhodopsin molecules per *Pelagibacter* cell by assuming that the ratio between the 633-nm absorbance change 5 ms after the flash, and the absorption of proteorhodopsin at its visible wavelength maximum, is the same in our experiments as cited in the literature[2]. We also assumed an extinction coefficient of 50,000 $M^{-1}\,cm^{-1}$ for proteorhodopsin at this wavelength. We calculate that *Pelagibacter* cells have on the order of 10,000 proteorhodopsin molecules. Considering the unusually small physical size of these bacteria, this value agrees well with the previously reported estimate of 25,000 proteorhodopsin molecules per bacterioplankton cell from seawater samples[2].

Pelagibacter cells grew equally well in seawater in the presence and absence of retinal, a cofactor that is required for synthesis of the proteorhodopsin holoenzyme (Fig. 4). An operon was found in the HTCC1062 genome that probably encodes the enzymatic pathway for β-carotene and retinal biosynthesis. In *Halobacterium*, the *crtB*, *crtI* and *crtY* genes catalyse β-carotene synthesis from GGPP (geranylgeranyl pyrophosphate). The *blh* gene is thought to mediate cleavage of β-carotene to retinal[16]. We found a *crtIBY* operon in the HTCC1062 genome that was immediately followed by an open reading frame (ORF) with weak but discernable homology (PSI-BLAST expect $= 8 \times 10^{-6}$) to a *brp* homologue from *Haloarcula marismortui*. This leads us to speculate that *Pelagibacter* has the *blh* gene and is able to produce retinal from β-carotene. The operon also included the *lytB* gene, which encodes a protein that catalyses the conversion of 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate into IPP (isopentenyl-diphosphate) and DMAPP (dimethylallyl-diphosphate), which are precursors for geranylgeranyl pyrophosphate synthesis.

The proteorhodopsin photosystem is encoded by a single gene. Despite its simplicity, proteorhodopsin occupies space in the cell membrane, thereby reducing the surface area available for nutrient transporters in an ecosystem in which nanomolar substrate

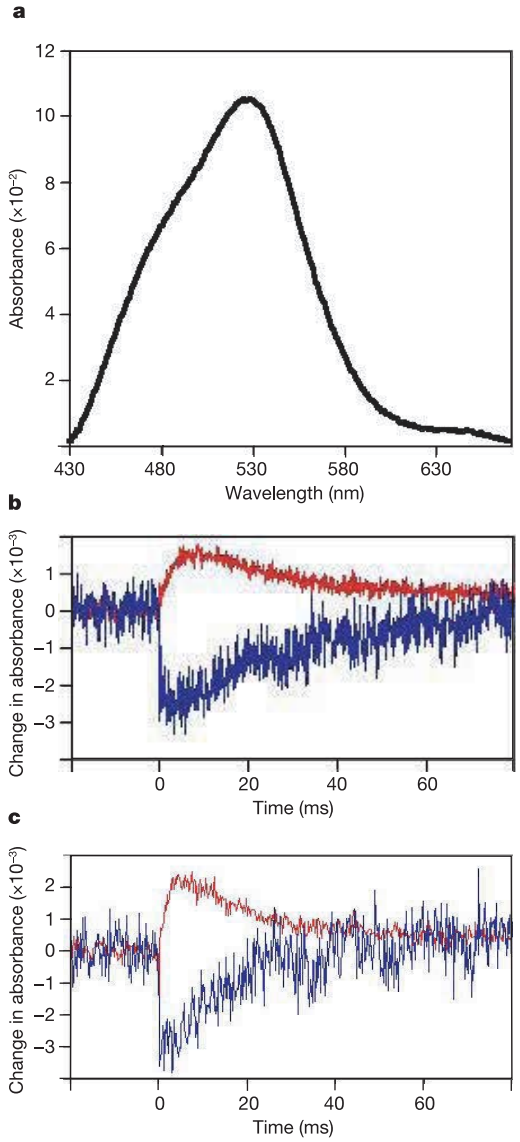

Figure 3 | Spectroscopy. a, Absorption spectrum from HTCC1062 proteorhodopsin cloned into an *E. coli* expression vector, showing a characteristic absorption maximum at 530 nm. See main text for experimental details. **b**, **c**, Flash-induced absorption changes in HTCC1062 cells grown on a diurnal light cycle (**b**) or in the dark (**c**), showing short-timescale transients typical of rhodopsin proton pumps. Red, 633-nm absorbance; blue, 488-nm absorbance.

concentrations prevail and nutrient competition is a dominant factor. From the estimated number of proteorhodopsin molecules per *Pelagibacter* cell (10,000), and the measured size of the *Pelagibacter* cytoplasm, we calculate that proteorhodopsin occupies approximately 20% of the inner membrane surface area. Thus, the expression of proteorhodopsin may involve metabolic trade-offs that become fully manifest only when specific conditions that occur periodically in nature are modelled and reproduced.

Notably, we found no consistent differences in growth rate or maximum cell yield in HTCC1062 cultures grown in a diurnal light regime or in complete darkness on natural seawater (Fig. 4). This observation suggests that proteorhodopsins have a subtle role related to the ecological complexities of the ocean surface environment. We speculate that the benefits of proteorhodopsin may be most evident when organic carbon limitation decreases the ability of cells to generate a proton motive force by respiration. This hypothesis cannot be tested at present because the factors that limit the growth

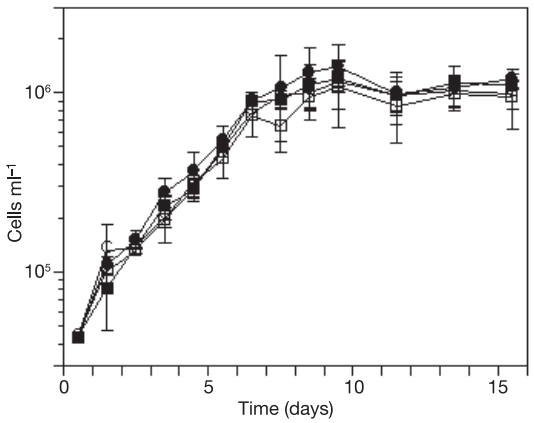

Figure 4 | Growth characteristics of HTCC1062. Bacteria were grown in seawater supplemented with N and P (LNHM) with no added organic carbon, on a diurnal light cycle (open symbols) or in darkness (closed symbols) under high-range light intensity (circles, 680 μmol m^{-2} s^{-1}) or middle-range light intensity (squares, 250 μmol m^{-2} s^{-1}). Error bars show standard deviation for triplicate experiments. No difference was observed for replicates with and without added retinal (data not shown).

of *Pelagibacter* in seawater remain unidentified. Alternatively, proteorhodopsin may confer a slight advantage in growth rate or biomass yield that has not been resolved experimentally. In some experiments, the cells grew slightly better in the light, but the results varied—possibly as a consequence of variation in the composition of the natural seawater used in different experiments.

Owing to the vast expanse of the ocean habitat, its age and the extraordinary population sizes of bacterioplankton[17], theory predicts that genetic traits conveying subtle differences in fitness should become fixed by selection, yielding strains that are highly adapted to their environment[18]. Proteorhodopsin appears to be one of a suite of characteristics that enable bacterioplankton cells to thrive in the competitive, low-nutrient conditions that prevail at sea. Fully understanding the role of this unique photochemistry may ultimately require the delicate reproduction of conditions that occur periodically in the life history of these cells.

METHODS

Cultivation. Cells were cultivated as described in ref. 9, on LNHM (low nutrient heterotrophic medium) medium with the addition of 1 μM retinal. Seawater for media was collected with Niskin bottles from station NH5, five miles offshore of Newport, Oregon. For routine cultivation, cells were grown under a diurnal light cycle: cool-white light was supplied at 24 μmol m^{-2} s^{-1} in a 14-h light:10-h dark cycle. Larger banks of cool-white lights were used to increase intensities for the experiments shown in Fig. 4. The light treatments and dark controls for Fig. 4 were incubated in duplicate in the same water baths, and the medium for all treatments was made from the same seawater sample.

Sample collection. Plankton cells were collected at station NH5 from 40 l of seawater (at 10-m depth), pre-filtered with an 0.8-μm filter (Supor) and concentrated by tangential flow filtration with a Millipore Pellicon II Mini system equipped with a 30-kDa regenerated cellulose filter. Concentrated cells were pelleted by centrifuging in a Beckman J2-21 centrifuge using a JA-20 rotor at 48,400*g* and a temperature of 4 °C. Cells were resuspended in a minimal volume of seawater and stored at −20 °C until analysis.

Gene cloning. The proteorhodopsin gene was amplified from isolate HTCC1062 by polymerase chain reaction using primers 5′-ACCATGGGTAAAAAA CTAAAATTGTTTGC-3′ and 5′-CTTAGCTCTACCAGGTTGAGA-3′. Cloning and expression were performed as described in ref. 1.

Phylogenetic analysis. Sequences were aligned using Clustal W. Bayesian analyses were conducted using MRBAYES 3.0 (ref. 19); searches were conducted for a total of 1,000,000 generations with phylogenetic trees sampled every 100 generations. Monte Carlo Markov chains were calculated using an integration of ten fixed amino-acid rate matrices. Out of 10,000 resulting trees, the initial 1,000 generations were identified as preceding the convergence of likelihoods and excluded from post-run analyses (burn-in). A majority-rule consensus tree with

averaged branch lengths and posterior probabilities was calculated from the remaining 9,000 trees (Fig. 1). Maximum parsimony and neighbour-joining analyses were conducted using PAUP* 4.0b10, with the following settings: 100 replicates of random sequence addition, TBR branch swapping, and MulTrees in effect. The relative support for the resulting trees was determined by 1,000 bootstrap replications with the same search options as previously described.

Mass spectrometry. Mass spectrometry (MS) methods are described elsewhere[20]. Briefly, cultured cell pellets were washed and lysed, and membrane material was collected by centrifugation and then dissolved with the detergent dodecyl maltoside. Proteins were then digested in-solution using trypsin. Total protein from lysates of environmental cell pellets was separated by one-dimensional SDS–PAGE. Gel lanes were cut into six pieces, and in-gel tryptic digestions were performed. Chromatography was performed using a Waters CapLC system with a 0.32-mm (inner diameter) symmetry column packed with 5-μm C18 particles, and the eluate was spotted onto MALDI plates using a MALDIprep spotter (Waters Corp). MALDI-MS and tandem MS were performed on an Applied Biosystems 4700 proteomics analyser. Peptides were identified using the program Mascot[21].

Spectroscopy. Attenuance spectra of proteorhodopsin were acquired from standard 1-cm cuvettes on a Cary 300 spectrophotometer (Varian Instruments), using intact *E. coli* cells expressing the cloned proteorhodopsin protein. The absorption spectrum depicted in Fig. 3a was obtained by subtracting a cubic baseline, determined by fitting the attenuance values from 350–450 nm and from 625–800 nm, which represents scattering of the cell suspension and absorption of *E. coli* exclusive of proteorhodopsin. Transient absorption spectra (Fig. 3b, c) of HTCC1062 cells in 1-cm pathlength fluorometer sub-microcuvettes (16.160F-Q-10/Z20, Starna Cells) were acquired on a laboratory-constructed flash-photolysis apparatus. Each flash-induced transient absorption trace was obtained by averaging 32 acquisitions, one every 25 s. The excitation flash (10 mJ $pulse^{-1}$) at 532 nm was supplied by a Nd:YAG laser (Molectron MY-34-10, Coherent). The probe beams were supplied by a helium-neon laser (633 nm; Hughes 5000, Melles Griot) and an argon laser (488 nm; Omnichrome 532, Melles Griot). These wavelengths were chosen because for probe wavelengths shorter than the transient isosbestic point (at about 550 nm), the transient absorption is expected to be negative, whereas the converse is expected for wavelengths longer than the isosbestic wavelength[2] (as observed in Fig. 3b, c). Polarizations of the laser beams were set to 'magic angle' orientations (valid because the transient absorption is small).

Received 29 November 2004; accepted 11 July 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. 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).
4. Man-Aharonovich, D. *et al.* Characterization of RS29, a blue-green proteorhodopsin variant from the Red Sea. *Photochem. Photobiol. Sci.* **3,** 459–462 (2004).
5. 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).
6. Sabehi, G. *et al.* Novel Proteorhodopsin variants from the Mediterranean and Red Seas. *Environ. Microbiol.* **5,** 842–849 (2003).
7. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304,** 66–74 (2004).
8. Morris, R. M. *et al.* SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420,** 806–810 (2002).
9. Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418,** 630–633 (2002).
10. Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309,** 1242–1245 (2005).
11. Field, K. G. *et al.* Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl. Environ. Microbiol.* **61,** 63–70 (1997).
12. Wang, W. W., Sineshchekov, O. A., Spudich, E. N. & Spudich, J. L. Spectroscopic and photochemical characterization of a deep ocean proteorhodopsin. *J. Biol. Chem.* **278,** 33985–33991 (2003).
13. Man, D. *et al.* Diversification and spectral tuning in marine proteorhodopsins. *EMBO J.* **22,** 1725–1731 (2003).
14. Stapels, M. D. & Barofsky, D. F. Complementary use of MALDI and ESI for the HPLC-MS/MS analysis of DNA-binding proteins. *Anal. Chem.* **76,** 5423–5430 (2004).
15. Ram, R. J. *et al.* Community proteomics of a natural microbial biofilm. *Science* **308,** 1915–1920 (2005).
16. Peck, R. F., Johnson, E. A. & Krebs, M. P. Identification of a lycopene β-cyclase required for bacteriorhodopsin biogenesis in the archaeon *Halobacterium salinarum*. *J. Bacteriol.* **184,** 2889–2897 (2002).
17. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: The unseen majority. *Proc. Natl Acad. Sci. USA* **95,** 6578–6583 (1998).
18. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
19. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17,** 754–755 (2001).
20. Stapels, M. D., Cho, J. C., Giovannoni, S. J. & Barofsky, D. F. Proteomic analysis of novel marine bacteria using MALDI and ESI mass spectrometry. *J. Biomol. Tech.* **15,** 191–198 (2004).
21. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20,** 3551–3567 (1999).

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

Acknowledgements We thank J. Nibler and the chemistry department at Oregon State University for the use of the Nd:YAG laser and student laser laboratory facilities; W. Hetherington, R. Letelier, B. Geller and O. Béjà for helpful discussions; and E. L. Barofsky for her assistance with MALDI mass spectrometry. This research was supported by the National Science Foundation, Diversa Corporation and the National Institute of Environmental Health Sciences.

Author Contributions S.J.G. led the genome sequencing project, provided the bioinformatics analyses and was the primary writer. L.B. and E.J.M. led the DNA sequencing team at Diversa Corporation. J.-C.C., L.J.W. and H.J.T. provided the growth data. M.D.S. and D.F.B. provided the mass spectrometry analysis. R.D. and S.L. performed the light spectroscopy experiments. K.L.V. cloned the proteorhodopsin gene and showed that it was a light-dependent proton pump, with the assistance of R.D. M.S.R. isolated the *Pelagibacter*.

Author Information The HTCC1062 proteorhodopsin gene sequence has been deposited in GenBank under accession number CP000084. 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 S.J.G. (steve.giovannoni@oregonstate.edu).